

The Upside of Acknowledging Prejudiced Behavior

Aaron J. Moss

Rachel D. Budd

M. Annelise Blanchard

Laurie T. O'Brien

Tulane University

Author Note

Aaron Moss is now at CloudResearch, Queens, NY. Rachel Budd is now at the U.S. Department of Justice, Washington, D.C. M. Annelise Blanchard is now at the Department of Psychology, Katholieke Universiteit Leuven.

Correspondence concerning this article should be addressed to Aaron Moss. Email: amos1@tulane.edu

Abstract

What do people think of those who respond to confrontation by acknowledging personally prejudiced behavior? In six experiments ($N = 3,344$), people judged a man who made a prejudiced comment and responded to confrontation by acknowledging, denying, or, in some cases, saying nothing about his prejudice. Participants consistently evaluated someone who acknowledged prejudice as warmer, more moral, and ironically, less prejudiced than someone who denied. People also perceived acknowledging as more appropriate and less typical than denying regardless of whether the prejudice was racism or sexism. Moreover, men and women, White, Black, and Asian people alike evaluated acknowledgements more positively than denials. Evidence from multiple experiments suggests that people form more positive impressions of those who acknowledge than deny because acknowledgment signals more of a learning orientation to prejudice and intergroup relations. Although people frequently respond to confrontation by denying prejudiced behavior, there appears to be an upside to acknowledging.

Abstract WC: 150

Key words: acknowledging prejudice, intergroup interactions, confronting prejudice, performance and learning-oriented behaviors

The Upside of Acknowledging Prejudiced Behavior

A growing social movement encourages people to speak up and confront others who express prejudice (e.g., Ashburn-Nardo et al., 2008; Mallett & Monteith, 2019; Paluck, 2011). Part of the excitement behind this approach to reducing prejudice stems from research that shows people who are confronted tend to monitor their future behavior and express less prejudice than those who are not confronted (Chaney & Sanchez, 2018; Czopp et al., 2006; Mallett & Wagner, 2011; Parker et al., 2018). Perhaps unsurprisingly, people who are confronted for making prejudiced comments tend to be viewed negatively and, in many instances, are viewed even more negatively than people who make the same biased comments but are not confronted (e.g., see Czopp, 2019). This raises the question: how should people respond when confronted for prejudice?

In the present research, we examined how a perpetrator's response to confrontation shapes other people's impressions of the perpetrator's character. The current experiments integrate research examining how confrontation affects evaluations of a perpetrator's character (e.g., Boysen, 2013; Dickter et al., 2011; Martinez et al., 2017; Rasinki & Czopp, 2010; Vaccarino & Kawakami, 2020) with research examining how motivational mindsets affect intergroup relations (e.g., Carr et al., 2012; Murphy et al. 2011; Neel & Shapiro, 2012; Rattan & Dweck, 2010; Rattan & Dweck, 2018; Simon et al., 2018). We propose that a perpetrator's response to confrontation signals something meaningful about their motivational approach to prejudice and intergroup relations. While people who are confronted for prejudice may tend to feel defensive and deny being prejudiced, we argue that acknowledging prejudiced behavior communicates that one is motivated to learn, which in turn shapes other people's evaluations of the perpetrator's character.

Confronting Prejudice

Even though research on confrontation has been growing (Mallet & Monteith, 2019), one relatively unexplored question is how confrontation affects interpersonal evaluations of a perpetrator. The available evidence indicates that people who are confronted are viewed negatively (e.g., Boysen, 2013; Dickter et al., 2012; Martinez et al., 2017; Rasinki & Czopp, 2010; Vaccarino & Kawakami, 2020). For example, in one study, a White man who was confronted by another White person for making racist remarks about Black people was seen as less respectable and less moral than a White man who made racist remarks but was not confronted (Dickter et al., 2012). Similarly, in another study, a man who was confronted by another man for making sexist remarks was liked less than a man who made sexist remarks but was not confronted (Boysen, 2013). In other words, people who are confronted for expressing prejudice are evaluated even more negatively than people who simply express prejudice but are not confronted.

If people evaluate those who are confronted for prejudice negatively, how should someone who is confronted respond? Is it best to deny engaging in prejudiced behavior, to acknowledge one's prejudice, or do something else? Regardless of what is ideal, research reveals that people engage in a variety of responses when actually confronted for prejudiced behavior (e.g., Chaney & Sanchez, 2018; Czopp & Monteith, 2003; Czopp et al., 2006; Winslow, 2004). In one influential paper, for example, Czopp and colleagues (2006) elicited racially prejudiced statements from participants, had a confederate confront participants, and then examined participants' written responses to the confrontation. Although many people denied that race influenced their behavior or that they were a racist person, a meaningful minority of people responded by either recognizing their bias or apologizing. Other studies using different methods

or examining different types of prejudice (e.g., racism vs. sexism) have yielded somewhat different conclusions about how often people engage in various responses. Nevertheless, the existing studies demonstrate variability in how people respond to confrontation (e.g., Czopp & Monteith, 2003; Winslow, 2004; Chaney & Sanchez, 2018).

In the present research, we focused on two potential responses to confrontation: acknowledgements and denials. At a minimum, we argue, an acknowledgment of prejudice must include a verbal response that recognizes one's behavior was influenced by prejudice. However, acknowledgements may also frequently include language that validates the confronter's perspective. In comparison, a denial of prejudice is a verbal response that rejects the confronter's accusation of prejudice. Focusing on how acknowledgements and denials shape people's evaluations of a confronted person's character has at least two benefits. First, past research suggests that acknowledgements and denials are both relatively common responses to confrontation (e.g., Czopp et al., 2006), meaning that our study can explore how people are perceived in many instances. Second, we believe that acknowledgement and denial represent two distinct responses to confrontation that may signal something important about a perpetrator's motivational mindset.

Motivational Mindsets

Prejudice often creates contradictions. For instance, existing theories of racial and gender prejudice emphasize the conflict created by prejudices that people internalize at a young age and the egalitarian social norms that often prohibit the blatant expression of these prejudices (Crandall & Eshleman, 2003). Due to social norms, White people and men must frequently regulate their thoughts, feelings, and behaviors in situations where their racism and sexism are activated but expressing these prejudices is proscribed (e.g., Murphy et al., 2011; Plant &

Devine, 1998; Trawalter et al., 2009). To regulate prejudice, people often adopt different types of goals (Murphy et al., 2011).

The goals that people from high status groups adopt when trying to regulate their prejudice can be influenced by their mindset, or implicit theory of prejudice (Carr et al., 2012; Murphy et al., 2011; Neel & Shapiro, 2012). While some people have growth mindsets and view prejudice as a malleable trait that is changeable, other people have fixed mindsets and view prejudice as a stable trait that is fixed (Carr et al., 2012; Neel & Shapiro, 2012; Rattan & Dweck, 2010; Rattan & Dweck, 2018). When people view a trait as relatively malleable, they are more likely to adopt learning goals whereas when they view a trait as relatively fixed they tend to adopt performance goals (Dweck and Leggett, 1988). Applied to the regulation of prejudice, people with a growth mindset may adopt learning goals and focus on reducing their prejudice while developing competence in intergroup interactions. Often, this means people may aim to connect with others in interracial interactions and to learn from the situation (Carr et al., 2012; Murphy et al., 2011; Neel & Shapiro, 2012). By contrast, people with a fixed mindset may adopt performance-oriented strategies to help them deny, hide, or avoid the appearance of prejudice. (Neel & Shapiro, 2012, pg. 103).

Drawing on this theory, we conceptualized acknowledging and denying prejudice in response to confrontation as examples of learning-oriented and performance-oriented behaviors that, respectively, reflect different motivational mindsets (e.g., Dweck & Leggett, 1988; Murphy et al., 2011). A person with a growth mindset who is confronted for prejudice may adopt a learning orientation and view acknowledging as an opportunity to learn from the interaction (e.g., Neel & Shapiro, 2012). Such a person may view an expression of prejudice as a mistake that can be corrected in the future (e.g., Rattan & Dweck, 2010; Rattan & Dweck, 2018). By

comparison, a person with a fixed mindset may adopt a performance orientation and view denying as a behavior that may convince others they are unprejudiced.

In the present research, we sought to understand how a perpetrator's response to confrontation shapes other people's evaluations of the perpetrator's character. Past research suggests that other people interpret and respond to learning-oriented behaviors more positively than performance-oriented behaviors (Carr et al., 2012; Neel & Shapiro, 2012; Simon et al., 2019). Acknowledging that one has expressed prejudice may signal to others that one is willing to take other people's perspectives, learn from past mistakes, and change future behavior. In comparison, denying that one has expressed prejudice may signal to others that one is unwilling to change or to see the situation from other people's perspectives. On this basis, we expected that, following confrontation for expressing prejudice, someone who acknowledges their prejudice would be evaluated more positively than someone who denied their prejudice.¹

Overview

We conducted six experiments to examine how people from different groups perceived someone who responded to confrontation by either acknowledging, denying, or, in some cases, saying nothing about his prejudiced behavior. Across experiments, we measured what people thought of the perpetrator's response to being confronted, how people evaluated the perpetrator's character, and whether people viewed the perpetrator as motivated by a performance or learning orientation to intergroup interactions. We considered evaluations of character especially important and sought to measure character by assessing people's perceptions of warmth, competence (Fiske et al., 2007), and two constructs directly relevant to our research: prejudice and morality. Research shows that warmth is often more important in character evaluations than

¹ Initially, we expected that only members of minority groups, and not majority groups, would perceive acknowledging more favorably than denying. However, after two pilot studies showed that White and Black people alike perceived acknowledging more favorably than denying, we revised our predictions for the current research. See supplemental online materials (SOM).

competence (Willis & Todorov, 2006) because warmth corresponds to other-oriented outcomes (e.g., friendliness, trustworthiness). Because prejudice and morality also appear similarly other-oriented, we expected acknowledging to affect people's evaluations of warmth, prejudice, and morality more than competence.

In Experiment 1, we examined perceptions of a man who responded to confrontation by acknowledging or denying sexism and tested the hypothesis that a man who acknowledged would be evaluated more favorably than a man who denied. In the remaining experiments (2a, 2b, 3, 4, and 5), we examined perceptions of a White man who responded to confrontation by acknowledging or denying racism. We sought to understand not only if acknowledging is perceived more positively than denying but also why. Our hypothesis was that acknowledging prejudice is perceived favorably because it signals a stronger learning-orientation than denying does. In Experiment 2a, we examined how acknowledging anti-Black prejudice affects perceptions of learning and performance orientation, while in Experiment 2b we tested whether acknowledging anti-Black prejudice has an indirect effect on perceptions of character that operates through perceptions of learning orientation. In Experiment 3, we expanded our investigation to an instance of anti-Asian prejudice and tested whether members of both the racial majority and minority viewed acknowledging as a sign of learning orientation.

Finally, in Experiments 4 and 5 we returned to the study of anti-Black prejudice and tested whether members of both the racial majority and minority groups viewed acknowledging as a sign of learning orientation. Most importantly, Experiments 4 and 5 compared different types of acknowledgements to examine the role of perceived learning orientation in restoring positive evaluations of character. Experiment 4 compared an other-focused acknowledgement that included a recognition of biased behavior and a validation of the confronter's perspective

with a self-focused acknowledgement that included a recognition of biased behavior but refuted the confronter's perspective. Experiment 5 compared an other-focused acknowledgement to a mere acknowledgment that only contained a recognition of biased behavior. Thus, the goal of the final two experiments was to isolate the necessary and sufficient conditions for an acknowledgment of prejudice to improve character evaluations relative to a denial.

Experiment 1: "What I said was sexist."

In Experiment 1, we measured what people thought of a perpetrator's response and character after he made a prejudiced remark that dismissed the competence of a woman in the workplace. We attempted to manipulate whether the man's comment was based on a relatively more hostile or more benevolent form of sexism (Glick & Fiske, 1996). Hostile sexism refers to antipathy towards women and includes the belief that men ought to have more power than women and a corresponding fear that women will use subversive tactics to usurp men's power. For example, claiming that a female co-worker is likely to make false complaints about discrimination in the workplace is an act of hostile sexism (Glick & Fiske, 1996). In comparison to hostile sexism, benevolent sexism is characterized by subjectively favorable, yet patronizing beliefs about women. For example, complimenting a female co-worker's physical appearance may appear to be an act of flattery, but it can undermine women's performance on difficult tasks and therefore is more akin to an act of benevolent sexism (e.g., Kahalon et al, 2018).

After the male perpetrator in our scenario expressed hostile or benevolent sexism about a future female colleague, one of his male coworkers confronted him. We manipulated whether the perpetrator responded to the confrontation by acknowledging or denying that the comment he made was prejudiced. We predicted that participants would evaluate the man's response and character more favorably when he acknowledged prejudice than when he denied. By

manipulating whether the man's comment was relatively more hostile or benevolent, we were able to test whether the benefits of acknowledging extend to both hostile and benevolent expressions of sexism. Although we did not have firm hypotheses, we explored the possibility that people may respond more favorably to an acknowledgement of hostile sexism than benevolent sexism. Most people find hostile sexism more offensive than benevolent sexism (Glick & Fiske, 2001) and thus acknowledging hostile sexism may be a clearer admission of wrongdoing.

We report all conditions, all measures, and all data exclusions in all experiments. In addition, across experiments, we excluded people who completed less than 80% of scale items because we considered this a sign of inattention. We did not look at the results of any experiment until all data for that experiment were collected. Data, analysis scripts, and output are available at: <https://osf.io/cbeu7/>

Method

Participants

A total of 320 undergraduates at a private U.S. university completed the study for partial course credit. Our initial goal was to recruit at least 50 women per cell to ensure we had a sufficient sample to examine women's perceptions of men who acknowledge or deny sexism. However, we did not exclude men from data collection, and for the sake of completeness, we report results with the full sample.² We excluded one person's data for inattention, leaving 211 women and 108 men ($M_{age} = 18.94$, $SD = 1.53$). Like the university demographics, most participants were White (80% White, 8% Asian, 3% Black, 4% Hispanic/Latino, 5% Other). A

² An analysis with data from only women showed that women perceived acknowledgement as significantly more appropriate and sincere, but less typical than denial. Women also perceived a man who acknowledged as warmer, more competent, more moral, and less prejudiced than a man who denied. No significant effects of sexism type emerged. Thus, all hypotheses were supported in the female sample.

sensitivity analysis revealed that with 319 participants we had 80% power to detect an effect size of $\eta_p^2 = .024$ in a 2 (response type) $\times$ 2 (sexism type) $\times$ 2 (participant gender) Multivariate Analyses of Variance.

Procedure

Participants were randomly assigned to read one of four vignettes. In each vignette, participants read about Brandon, a project manager at a tech company who made a sexist comment about a female job applicant. Participants read that Brandon and other managers hire employees by meeting to discuss applicants, voting, and then providing a recommendation to the director of personnel. After the group voted to hire a woman, another male manager asked Brandon how he felt about the vote. In the hostile sexism condition Brandon said, “Well, you know this woman is probably under-qualified for the job and when she doesn’t get promoted or gets a bad performance review, she’ll say it’s because we’re all sexist.” In the benevolent sexism condition, Brandon responded by saying, “Well, you know this woman is probably under-qualified for the job, but at least with her around there will be a pretty face to look at.”

Next, participants read that the other manager disagreed and confronted Brandon for using stereotypes about women. Brandon’s response to the confrontation constituted the manipulation of acknowledgement or denial. In the acknowledgment condition, participants read that Brandon responded with, “I see what you mean. I was just trying to be funny and it was the first thing that came to mind. On second thought, I admit that was prejudiced of me and unfair.” In the denial condition, Brandon said, “I don’t know what you mean. I was just trying to be funny and it was the first thing that came to mind. I don’t think it was prejudiced at all. It was just a joke.” After reading the vignette, participants answered items measuring their perceptions of Brandon’s response and character.

Measures

All dependent variables were measured on 7-point Likert-type scales (1 = *strongly disagree*, 7 = *strongly agree*), except where noted. In all experiments, we report dependent measures in the order they were assessed. Our supplementary online materials (SOM) contain a description of additional, exploratory measures beyond the key measures reported here.

Perceptions of the Response

We measured what people thought of Brandon's response with 19 adjectives capturing positive and negative reactions. From these items, we created three scales based on a pilot experiment and an exploratory factor analysis described in the SOM: appropriate (appropriate, immature (R), persuasive, effective, professional, compelling, and fitting), sincere (honest, genuine, sincere, untrustworthy, heartfelt, dishonest (R), and candid) and typical (surprising (R), unanticipated (R), predictable, common, and typical). Table 1 displays reliability coefficients for all dependent variables. The grand means, standard deviations, and correlations between all measures are reported in the SOM.

Perceptions of Character

We measured perceptions of character with 25 adjectives. Ten adjectives were adopted from Klein & Epley (2014) to measure perceptions of warmth (tolerant, warm, good-natured, sincere, and caring) and competence (competent, confident, independent, competitive, and intelligent). The remaining items were a) based on stereotypes of majority group members (Vorauer et al., 1998) and captured how prejudiced Brandon appeared (biased, racist, prejudiced, close-minded, ignorant), b) intended to measure how moral Brandon appeared (trustworthy, moral, principled; Goodwin et al., 2014), or c) exploratory (e.g., professional, funny, creepy). Analyses of the exploratory items are not reported here but the data are on OSF.

Results

We tested our hypotheses with 2 (response type: acknowledge vs. deny) $\times 2$ (sexism type: benevolent vs. hostile) $\times 2$ (participant gender: male vs. female) between-subjects MANOVA on perceptions of the response and perceptions of character.

Perceptions of the Response

As predicted, the multivariate main effect of response type was significant, Wilks' $\Lambda = .52$, $F(3, 309) = 96.55$, $p < .001$, $\eta_p^2 = .484$. Participants thought it was more appropriate, more sincere, and less typical when Brandon responded with an acknowledgement than a denial. See Table 1 for conditional means and univariate tests of response type.

Table 1

Experiment 1: Means, standard deviations, and inferential statistics

Response	Response Type		$F(1, 311)$	p	η_p^2 [90% CI]
	Acknowledge	Deny			
	$M(SD)$	$M(SD)$			
Appropriate ($\alpha=.92$)	4.26 _a (1.32)	2.21 _b (0.70)	278.78	<.001	.485 [.409, .527]
Sincere ($\alpha=.89$)	3.95 _a (1.24)	3.23 _b (1.13)	26.82	<.001	.079 [.038, .131]
Typical ($\alpha=.84$)	4.10 _a (1.16)	4.61 _b (1.10)	14.49	<.001	.045 [.014, .187]
Character					
Warmth ($\alpha=.83$)	3.49 _a (0.89)	3.04 _b (0.90)	18.72	<.001	.057 [.022, .103]
Competence ($\alpha=.63$)	4.29 _a (0.79)	4.19 _a (0.75)	1.30	.255	.004 [0, .024]
Prejudice ($\alpha=.83$)	5.30 _a (0.89)	5.60 _b (0.85)	9.99	.002	.031 [.007, .069]
Morality ($\alpha=.81$)	3.33 _a (0.92)	2.99 _b (0.94)	10.69	.001	.033 [.008, .072]

Note. Means that do not share a subscript within a row differ at a significance of $p < .05$.

There was also an unpredicted significant multivariate effect of gender, Wilks' $\Lambda=.97$, $F(3, 309) = 2.80$, $p = .040$, $\eta_p^2 = .026$. Men perceived Brandon's response as more sincere than

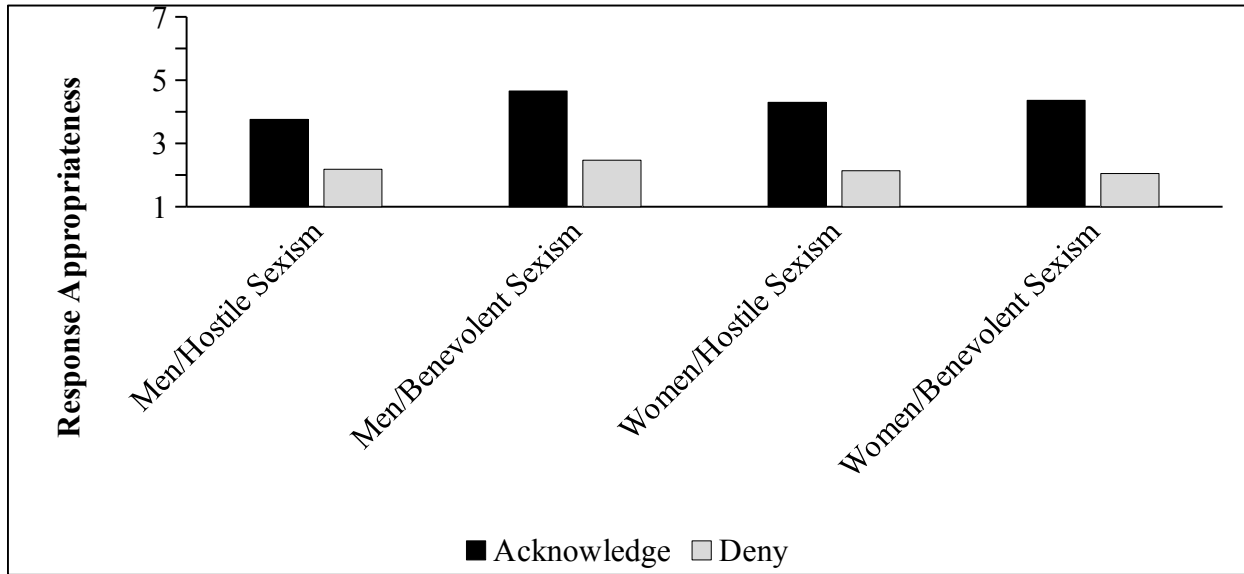
women, $F(1, 311) = 7.13, p = .008, \eta_p^2 = .022$, but there were no gender differences in appropriateness or typicality, $F_s < 1.83, p_s > .177, \eta_p^2 < .006$.

Finally, there were two other unexpected significant multivariate effects: a two-way response type $\times$ sexism type interaction, Wilks' $\Lambda = .97, F(3, 309) = 2.93, p = .034, \eta_p^2 = .028$, and a three-way response type $\times$ sexism type $\times$ gender interaction, Wilks' $\Lambda = .98, F(3, 309) = 2.68, p = .047, \eta_p^2 = .03$. All other multivariate tests were not significant, Wilks' $\Lambda_s > .97, F_s(3, 309) < 2.40, p_s > .084, \eta_p^2 < .021$.

To unpack these unpredicted multivariate interactions, we examined the univariate response type $\times$ sexism type interactions and response type $\times$ sexism type $\times$ gender interactions for appropriate, sincere, and typical. The univariate interactions were not significant for either sincere or typical, $F_s(1, 311) < 0.33, p_s > .568, \eta_p^2 < .001$. However, for appropriate, there was a significant response type $\times$ sexism type interaction, $F(1, 311) = 5.52, p = .019, \eta_p^2 = .017$, that was qualified by a significant response type $\times$ sexism type $\times$ gender interaction, $F(1, 311) = 6.05, p = .014, \eta_p^2 = .019$. To understand the three-way interaction, we tested for a response type $\times$ sexism type interaction separately among men and women. Among men, the two-way interaction was significant, $F(1, 104) = 8.69, p = .004, \eta_p^2 = .004$. The effect of acknowledging prejudice was weaker among men in the hostile sexism condition, $F(1, 104) = 20.53, p < .001, \eta_p^2 = .165$, than in the benevolent sexism condition, $F(1, 104) = 75.60, p < .001, \eta_p^2 = .421$. In comparison, the two-way interaction between response type and sexism type was not significant for women, $F(1, 207) = 0.01, p = .926, \eta_p^2 = .000$. That is, the effect of acknowledging prejudice was similarly strong for women in both the hostile sexism, $F(1, 207) = 121.53, p < .001, \eta_p^2 = .370$, and the benevolent sexism conditions, $F(1, 208) = 122.10, p < .001, \eta_p^2 = .370$, (see Figure 1).

Figure 1

Experiment 1: Effects of response type, sexism type, and participant gender on perceptions of response appropriateness



Perceptions of Character

The MANOVA revealed the predicted significant effect of response type, Wilks' $\Lambda = .94$, $F(4, 308) = 5.14$, $p = .001$, $\eta_p^2 = .063$. Participants thought Brandon was warmer, more moral, and less prejudiced when he acknowledged his sexism than when he denied it. However, acknowledging did not have a statistically significant effect on perceptions of competence. See Table 1 for conditional means and univariate tests of response type.

There was also an unpredicted multivariate effect of participant gender, Wilks' $\Lambda = .83$, $F(4, 308) = 15.68$, $p < .001$, $\eta_p^2 = .169$. Men thought Brandon was warmer, $F(1, 311) = 32.04$, $p < .001$, $\eta_p^2 = .093$, more moral, $F(1, 311) = 33.98$, $p < .001$, $\eta_p^2 = .098$, and less prejudiced, $F(1, 311) = 49.43$, $p < .001$, $\eta_p^2 = .137$, than women did. There were no gender differences in ratings of competence, $F(1, 311) = 0.23$, $p = .629$, $\eta_p^2 = .001$.

Although the multivariate effect of sexism type was significant, Wilks' $\Lambda = .96$, $F(4, 308) = 2.87$, $p = .023$, $\eta_p^2 = .036$, none of the univariate tests were, $F_s(1, 311) < 2.21$, $p_s > .139$, η_p^2 s $< .007$. Finally, all multivariate interactions were not significant, Wilks' Λ s $> .97$, $F_s(4, 308) < 0.30$, $p_s > .108$, η_p^2 s $< .024$.

Discussion

People perceived acknowledging sexism as more appropriate and sincere than denying, even though denying was perceived as more typical. People also perceived the perpetrator's character more favorably when he acknowledged than when he denied, and they evaluated acknowledgements of sexism more favorably than denials regardless of whether the initial expression of sexism was relatively more hostile or benevolent. Finally, although men generally had more positive evaluations of Brandon's character than women, acknowledging sexism was perceived more favorably by women and men alike.

Although Experiment 1 provided support for the hypothesis, it has several limitations. First, in addition to complimenting a woman's appearance, the benevolent sexism manipulation also impugned her competence, which is not typically an element of benevolent sexism (Glick & Fiske, 2001). Therefore, future research should replicate this finding with a more precise manipulation to ensure that men who acknowledge benevolent sexism are perceived more positively than men who deny such sexism. Second, in comparison to sexism, racism is generally taken more seriously (Czopp & Monteith, 2003). Thus, it is important to examine whether acknowledging racism is also perceived more favorably than denying sexism. Third, Experiment 1 lacked a true control condition. Therefore, it was unclear whether evaluations of the perpetrator suffered when he denied prejudice or benefitted when he acknowledged. Finally, Experiment 1 does not shed light on the mechanisms that lead to more positive evaluations when the

perpetrator acknowledged compared to when he denied his prejudice. Experiments 2a and 2b were designed to address all these limitations.

Experiments 2a and 2b: “I See What You Mean...that was Prejudiced of Me”

Experiments 2a and 2b examined responses to an expression of racial prejudice. Specifically, we measured what people thought of a White man who impugned the competence of a new Black employee by referring to him as “probably underqualified” and subsequently making a stereotypical comment about his likely ability at basketball. A White co-worker then confronted the perpetrator for using stereotypes about Black people. Unlike Experiment 1, Experiments 2a and 2b included a control condition in which participants were unaware of the perpetrator’s response to the confrontation. Finally, Experiments 2a and 2b tested the effect of acknowledging and denying prejudice on people’s perceptions of an actor’s learning and performance orientation. In Experiment 2b, we also assessed perceptions of the actor’s character.

Because acknowledging prejudice generally communicates a willingness to take the confronter’s perspective (e.g., Neel & Shapiro, 2012), we predicted that people would view an acknowledgment of prejudice as more learning-oriented than a denial or no response. In addition, we also initially predicted that people would view someone who denied prejudice as more performance-oriented than someone who acknowledged or said nothing. In Experiment 2b, we predicted that when someone responds to being confronted by acknowledging prejudice, people will perceive that person as more warm, less prejudiced, and more moral than when they deny or say nothing. Finally, we predicted that acknowledging would have an indirect effect on character (i.e., warmth, prejudice, and morality) that operates through perceptions of learning orientation. We preregistered our hypotheses for Experiment 2b: <https://osf.io/cbeu7/>.

Method

Experiment 2a Participants

A total of 319 undergraduates at a private U.S. university participated in the experiment for partial course credit. Our goal was to collect as much data as possible over the course of a semester. We excluded two people for inattention, leaving 317 people in the final sample (208 women, 108 men, 1 not reporting, $M_{age} = 19.03$, $SD = 1.30$). Most participants were White (76% White, 8% Asian, 5% Black, 6% Hispanic/Latino, 5% Other).³ A sensitivity analysis revealed that with 317 participants we had 80% power to detect an interaction effect size of $\eta_p^2 = .013$ in a 3 (Condition) $\times$ 2 (Motivation Type) mixed model ANOVA with repeated measures on the last factor.

Experiment 2a Procedure

In all conditions, participants read a modified version of the vignette used in Experiment 1. In this version of the vignette, participants read about Brandon, a White project manager at a tech company who made a prejudiced comment about a Black applicant. After the group of managers voted to hire a Black man named DeShawn, another White manager asked Brandon how he felt about the vote. Brandon responded by saying, “Well, you know this guy is probably under-qualified for the job, but at least with him our office basketball team will be more competitive.” Next, participants read that the other manager “disagreed and confronted Brandon for using stereotypes about Black people.”

Brandon’s response constituted the manipulation and participants were randomly assigned to condition. In the acknowledgment condition, participants read that Brandon responded with, “I see what you mean. I was just trying to be funny and it was the first thing that came to mind. On second thought, I admit that was prejudiced of me and unfair.” In the denial condition, Brandon said, “I don’t know what you mean. I was just trying to be funny and it was

³ The pattern of results and conclusions are unaltered if only the responses of White participants are retained in the analyses.

the first thing that came to mind. I don't think it was prejudiced at all. It was just a joke." Finally, in the control condition the scenario ended after the confrontation, meaning there was no response.

After reading the scenario, participants answered questions about whether Brandon is generally oriented toward performance or learning behaviors in interracial interactions (Neel & Shapiro, 2012; Simon et al., 2019). We also collected additional exploratory measures described in the SOM and available in our open datasets.

Experiment 2a Measures

We measured perceptions of Brandon's performance and learning orientation. Performance items tapped people's perceptions that Brandon is oriented toward avoiding the appearance of prejudice in interracial interactions while learning items tapped the perception that Brandon is oriented toward understanding the situation and his interaction partner.

Participants were asked to imagine what Brandon would do if he were in an interaction with a Black person that was going poorly and becoming racially tense. The five performance items included statements like, "Be extremely nice—nicer than he normally would be," and "Try to get out of the interaction as quickly as possible." The six learning items included statements like, "Try to understand things from his partner's point of view," and "Ask his partner what he or she is thinking." The five performance items and four of the learning items were adapted directly from Neel and Shapiro (2012). We created two additional learning items to capture the idea that Brandon would be motivated to respond to his partner's point of view (e.g., Try to place himself in his partner's shoes). In line with past work (Neel & Shapiro, 2012; Simon et al., 2018), the scales were only modestly correlated, $r(317) = .15, p = .010$. Table 2 displays scale reliabilities.

Experiment 2b: Participants, Procedure, and Measures

A total of 328 U.S. adults ($M_{age} = 34.67$, $SD = 11.79$, 183 men, 141 women, 4 non-binary/other) from Mechanical Turk (MTurk) participated in an experiment we expected to take seven minutes in exchange for \$0.75. Most participants were White (64% White, 10% Asian, 10% Hispanic/Latino, 14% Black, 1% Native American, 2% Other).⁴ Our original goal was to collect data from 320 people based on an a priori power analysis and allowing for the possibility that we might have to omit some participants. A sensitivity analysis revealed that with 328 participants we had 80% power to detect an interaction effect size of $\eta_p^2 = .014$ in a 3 (response type) $\times 2$ (motivation type) mixed model ANOVA with repeated measures on the last factor.

We managed data collection using TurkPrime (now CloudResearch, Litman et al., 2017). Participants were randomly assigned to condition, read the same scenario used in Experiment 2a, and answered the same perceived performance and learning-orientation items. In addition, participants completed the character (e.g., warmth, prejudice, and morality) scales from Experiment 1. Table 2 displays scale reliabilities for all measures. The grand means, standard deviations, and correlations between measures are reported in the SOM.

⁴ The pattern of results and conclusions are unaltered if only the responses of White participants are retained in the analyses.

Results

Experiment 2a

We tested our hypothesis with a 3 (response type: acknowledge vs. deny vs. control) $\times$ 2 (motivation type: performance vs. learning) mixed model ANOVA, with response type as a between subjects variable and motivation type as a within subjects variable.

There was a significant effect of motivation type. People generally perceived the perpetrator as more performance-oriented than learning-oriented, $F(1, 314) = 624.46, p < .001, \eta_p^2 = .665$. There was also a significant main effect of response type, $F(2, 314) = 16.67, p < .001, \eta_p^2 = .096$, that was qualified by the predicted interaction with motivation type, $F(2, 314) = 7.81, p < .001, \eta_p^2 = .047, 90\% \text{ CI } [.014, .087]$. See Table 2 for conditional means and univariate tests of response type.

Consistent with predictions, Tukey's post hoc comparisons showed that people viewed Brandon as more learning-oriented when he acknowledged his prejudice than when he denied or said nothing, $qs(314) > 7.71, ps < .001, ds > 0.70$ (see SOM for full reporting of all post hoc comparisons).⁵ However, there was no difference in perceived learning-orientation when Brandon denied as compared to when he said nothing, $q(314) = 0.06, p = .999, d = 0.01$.

Contrary to predictions, people perceived Brandon as equally performance-oriented when he acknowledged or denied, $q(314) = 0.83, p = .829, d = 0.08$. However, people perceived Brandon as less performance-oriented when he said nothing than when he acknowledged, $q(314) = 3.76, p = .022, d = 0.35$ or denied, $q(314) = 2.95, p = .094, d = 0.29$.

⁵ Except where noted, $k = 3$ for this and all subsequent post hoc tests.

Table 2*Experiments 2a and 2b: Means, standard deviations, and inferential statistics*

Experiment 2A	Response Type					
	Acknowledge	Deny	Control			
Motivation	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>F(2, 314)</i>	<i>p</i>	η_p^2 [90% CI]
Learning ($\alpha = .90$)	3.49 (1.24) _a	2.67 (0.92) _b	2.67 (1.11) _b	19.81	<.001	.112 [.060, .165]
Perform ($\alpha = .68$)	5.03(0.99) _a	4.95 (0.87) _{ab}	4.67 (1.07) _b	3.91	.021	.024 [.002, .055]
Experiment 2B	Acknowledge	Deny	Control			
Motivation	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>			
Learning ($\alpha = .95$)	4.09 (1.28) _a	3.33 1.48) _b	3.26 (1.48) _b	11.66	<.001	.067 [.027, .111]
Perform ($\alpha = .71$)	4.83 (1.02) _a	5.06 (0.98) _{ab}	4.66 (0.92) _a	4.60	.011	.028 [.004, .060]
Character	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>F(2, 325)</i>	<i>p</i>	η_p^2 [90% CI]
Warmth ($\alpha = .91$)	3.89 (1.10) _a	3.38 (1.16) _b	3.41 (1.27) _b	6.37	.002	.038 [.009, .074]
Prejudice ($\alpha = .89$)	4.84 (1.23) _a	5.39 (1.09) _b	5.42 (1.14) _b	8.59	<.001	.050 [.016, .090]
Morality ($\alpha = .87$)	3.92 (1.08) _a	3.50 (1.16) _b	3.47 (1.29) _b	5.02	.007	.030 [.005, .063]

Note. Means that do not share a subscript in a row differ at $p < .05$. Post hocs are Tukey tests.

Overall, these findings support our hypothesis that acknowledging signals more of a learning orientation than denying does. Interestingly, we predicted that denying would cause people to see the actor as more performance oriented. Across conditions, however, people saw relatively similar levels of performance orientation—albeit lower in the condition without a response. People also tended to perceive more performance orientation than learning orientation, overall, perhaps because people expect someone confronted for prejudice to be generally performance oriented. Regardless of the reason, the finding supporting our hypothesis was that acknowledging prejudice increased people's perceptions of learning orientation. Next, we sought to test whether perceptions of learning orientation are related to favorable evaluations of character. Although we continued to expect that acknowledging prejudice would lead to higher perceptions of learning orientation, we dropped our prediction that denying prejudice would lead to higher perceptions of performance orientation based on Experiment 2a's results.

Experiment 2b

Perceived Performance and Learning Motivation

Once again, there were significant effects of motivation type, $F(1, 325) = 197.07, p < .001, \eta_p^2 = .377$, response type, $F(2, 325) = 8.76, p < .001, \eta_p^2 = .051$, and a significant response type by motivation type interaction, $F(2, 325) = 10.12, p < .001, \eta_p^2 = .059$, 90% CI [.021, .101]. See Table 2 for conditional means and univariate tests of response type.

Replicating Experiment 2a, Tukey's post hoc tests revealed that people viewed Brandon as more learning-oriented when he acknowledged his prejudice than when he denied or said nothing, $qs(325) > 5.63, ps < .001, ds > 0.55$. There was no difference in perceptions of learning orientation when Brandon denied or said nothing, $q(325) = 0.49, p = .936, d = 0.04$.

The manipulation also affected performance motivation. People perceived Brandon as equally performance-oriented when he acknowledged or denied or acknowledged and said nothing, $qs(325) < |1.83|$, $ps > .193$, $d = |0.17|$. However, people perceived Brandon as more performance-oriented when he denied than said nothing, $(q(325) = 4.28, p = .008, d = 0.42$.

Perceptions of Character

Consistent with the hypothesis, the MANOVA testing the effect of Brandon's response on people's impression of his character was significant, Wilks' $\Lambda = .94$, $F(6, 646) = 3.23$, $p = .004$, $\eta_p^2 = .029$. See Table 2 for conditional means and univariate tests of response type. As expected, Tukey's post hoc tests revealed that people viewed Brandon as warmer, less prejudiced, and more moral when he acknowledged than denied or said nothing, $qs(325) > |3.72|$, $ps < .024$, $ds > |0.37|$. There was, however, consistently no difference between when Brandon denied and said nothing, $qs(325) < |0.28|$, $ps > .978$, $ds < |0.03|$.

Indirect Effect Analyses

We hypothesized that acknowledging prejudice increases perceptions of learning orientation and that, in turn, has a positive effect on perceptions of character. We tested this hypothesis using PROCESS Model 4 (Hayes, 2017). PROCESS uses ordinary least squares regression to create a 95% confidence interval for the indirect effect based on 5,000 bootstrapped samples. We specified the acknowledgment condition as the reference (i.e., coded as 0) and compared it to both the denial (1) and control conditions (1), and tested separate models for each character measure (e.g., warmth, prejudice, and morality).

Warmth. The relative indirect effects of whether Brandon acknowledged prejudice versus denied, $b = -.40$, $SE = .10$, 95% CI $[-.61, -.20]$, and acknowledged versus said nothing $b =$

-.43, $SE = .10$, 95% CI [-.63, -.24], on perceptions of warmth were both significant. See SOM Figure 1S.

Prejudice. The relative indirect effects of whether Brandon acknowledged versus denied, $b = .26$, $SE = .07$, 95% CI [.13, .41], and acknowledged versus said nothing, $b = .29$, $SE = .07$, 95% CI [.15, .44], on perceptions of his prejudice were both significant.

Morality. The relative indirect effects of whether Brandon acknowledged versus denied prejudice, $b = -.39$, $SE = .10$, 95% CI [-.59, -.20], and acknowledged versus said nothing, $b = -.42$, $SE = .10$, 95% CI [-.62, -.23], on perceptions of his morality were both significant.

In addition, we also tested an alternate model in which acknowledging decreases perceptions of performance orientation, which in turn, has a negative impact on perceptions of character. However, there was no evidence for an indirect effect of acknowledging on character that operated through performance orientation. See the SOM and Figure 2S for details.

Discussion 2a and 2b

Experiments 2a and 2b suggest that acknowledging prejudice signals a learning orientation. In Experiment 2a, people judged a man who acknowledged racial prejudice as more learning-oriented than a man who denied or said nothing. Experiment 2b replicated this effect and found that increased perceptions of learning orientation had an indirect effect on perceptions of character. Although we initially predicted that denying prejudice would be perceived as more performance-oriented than acknowledging in Experiment 2a, that prediction was not supported. Thus, we dropped that prediction for Experiment 2b.

Experiment 2b expanded on Experiment 1 in important ways. First, it demonstrated that acknowledging racial bias, in addition to gender bias, is perceived more favorably than denying. Second, by adding a control condition, Experiments 2a and 2b suggest that acknowledging bias

may improve character assessments of perpetrators, rather than denying prejudice harming character assessments. Finally, Experiment 2b provides some initial evidence to suggest that the benefits of acknowledging prejudice on character evaluations may operate by altering people's perceptions of an actor's learning orientation.

Collectively, Experiments 1, 2a, and 2b provide evidence that acknowledging both gender and racial prejudice may be perceived more favorably than denying gender and racial prejudice. Nonetheless, these initial experiments leave important questions unaddressed. For example, the actor in Experiments 1, 2a, and 2b made a statement that denigrated the competence of either a woman or a Black man. Would the benefits of acknowledging extend to situations in which an actor made a prejudiced statement alluding to competence stereotypes of Asian Americans as "model minorities"? We deemed this an important question since people are less likely to recognize stereotypes that emphasize a group's competence as prejudiced and harmful (Czopp, Kay, & Cheryan, 2015). Another limitation of the first three experiments is that, because most participants were White, they do not systematically address the question of whether people who belong to different racial groups will similarly view acknowledgements of prejudice as learning oriented. Experiment 3 addresses these limitations.

Experiment 3: "On second thought...that was prejudiced and unfair"

Experiment 3 tested whether acknowledging anti-Asian prejudice signals a learning orientation and leads to more positive character evaluations compared to denying prejudice or saying nothing. Experiment 3 extends Experiments 2a and 2b by moving beyond the Black-White binary (Plaut, 2010) that dominates the study of race and by testing whether people who belong to both the perpetrator's group and the target's group both perceive acknowledging prejudice as a sign of learning orientation. Based on Experiment 1, which suggested that women

and men perceived acknowledging similarly, we predicted that White and Asian people alike would perceive someone who acknowledged prejudice as more learning-oriented than someone who denied prejudice or said nothing. In addition, we predicted that, regardless of their race, people would make more positive character inferences about someone who acknowledged prejudice as opposed to someone who denied or said nothing. Finally, we predicted that acknowledging would have an indirect effect on character that operates through perceptions of learning orientation. Our preregistered hypotheses and analysis plan are available at:

<https://osf.io/cbeu7/>

Method

Participants

There were 1,339 participants ($M_{age} = 49.32$, $SD = 16.47$, 384 men, 949 women, 6 other/nonbinary) included in the data analyses. Based on an a priori power analysis and anticipating the need to exclude some participants, we attempted to recruit 372 adults (186 White and 186 Asian adults) in the U.S. using Prime Panels—a combination of online research panels offered through CloudResearch (see Chandler et al., 2019). Due to a programming error, 1,565 people completed the study (see SOM for details). We excluded data from two participants who had excessive missing data and 224 participants who did not indicate their race as either monoracial White or Asian (e.g., a person who identified as White and Latino would be excluded). Most participants were White ($n = 1,139$, 99.2% U.S. citizens). Among Asian participants ($n = 200$, 85.9% U.S. citizens) there was considerable diversity (46% Chinese, 16% Indian, 12% Japanese, 8% Korean, 7% Vietnamese, 7% Filipino, 5% other). A sensitivity analysis revealed that with 1,339 participants we had 80% power to detect an interaction effect size of $\eta_p^2 = .004$ in a 3 (response type) $\times$ 2 (participant race) $\times$ 2 (motivation type) mixed model ANOVA with repeated measures on the last factor.

Procedure

Participants were randomly assigned to condition. Across all three conditions, participants read about Jeff, a White man who works as a trader at a hedge fund. During a meeting to discuss who should analyze a potential investment, Jeff suggested the work be assigned to an Asian man and made a prejudiced comment: “Assign it to Jing. He’s a hard worker and those Asian brains were built for things like this.” Next, participants read that a White manager confronted Jeff. In response, Jeff either acknowledged, denied, or said nothing about his prejudice dependent on condition.

In the acknowledgment condition, Jeff said, “You know what, I just said the first thing that came to mind. But, I see what you mean, that may have come out wrong. On second thought, I admit that was prejudiced of me and unfair.” In the denial condition, Jeff said, “You know what, I just said the first thing that came to mind and I don’t see what you mean. It may have come out wrong, but I hardly think it qualifies as prejudiced.” Finally, in the control condition the scenario ended after the confrontation.

Measures

People completed the following measures from previous experiments: perceived learning orientation, perceived performance orientation, response appropriateness, response typicality, warmth, competence, morality, and prejudice. We added two items—just and fair—to the scale measuring perceptions of morality. Of note, we did not measure perceptions of the response in the control condition (because there was no response). Table 3 displays scale reliabilities for all measures. The grand means, standard deviations, and correlations between measures are in the SOM.

Results

Perceived Performance and Learning Orientation

We performed a 3 (response type: acknowledge, deny, or control) $\times$ 2 (participant race: Asian vs. White) $\times$ 2 (motivation type: performance vs. learning) mixed model ANOVA with repeated measures on the last factor. There were significant effects of response type, $F(2, 1333) = 12.73, p < .001, \eta_p^2 = .019$, and motivation type, $F(1, 1333) = 92.36, p < .001, \eta_p^2 = .065$, that were qualified by the predicted response type by motivation type interaction, $F(2, 1333) = 25.49, p < .001, \eta_p^2 = .036$, 90% CI [.021, .054]. See Table 3 for conditional means and univariate tests for response type.

Tukey's post hoc comparisons showed that people viewed Jeff as more learning-oriented when he acknowledged his prejudice than when denied or said nothing, $qs(1333) > 14.19, ps < .001, ds > 0.69$. There was no difference when Jeff denied and said nothing, $q(1333) = -0.47, p = .949, d = -0.02$. Because the univariate effect of response type on performance orientation was not significant, we did not perform Tukey's post hoc tests.

Of note, the main effect of race and all interactions involving race in the mixed model ANOVA were not significant, $F_s > 2.59, ps > .074, \eta_p^2 < .004$.

Perceptions of the Response

A 2 (response type: acknowledge vs. deny) $\times$ 2 (participant race: Asian vs. White) MANOVA on perceptions of Jeff's response revealed only the predicted main effect of response type, Wilks' $\Lambda = .83, F(2, 874) = 90.30, p < .001, \eta_p^2 = .171$. The multivariate main effect of race and response type $\times$ race interaction were both nonsignificant, Wilks' $\Lambda_s > .99, F_s(2, 874) < 1.36, ps > .258, \eta_p^2_s < .003$. Participants perceived Jeff's response as more appropriate when he acknowledged than denied, and at the same time, perceived Jeff's response as more typical when

he denied than acknowledged. Perceptions of the response were not measured in the control condition. See Table 3 for conditional means and univariate tests of response type.

Table 3

Experiment 3: Means, standard deviations, and inferential statistics

	Response Type					
	Acknowledge	Deny	Control			
Motivation	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (2, 1333)	<i>p</i>	η_p^2 [90% CI]
Learning ($\alpha = .95$)	4.56 _a (1.23)	3.58 _b (1.52)	3.61 _b (1.49)	25.52	<.001	.037 [.021, .054]
Performance ($\alpha = .66$)	4.35 (0.90)	4.53 (0.91)	4.38 (0.99)	0.46	.630	.001 [0, .004]
Response	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (1, 875)	<i>p</i>	η_p^2 [90% CI]
Appropriate ($\alpha = .95$)	4.58 _a (1.29)	2.81 _b (1.40)	-	174.26	<.001	.166 [.131, .203]
Typical ($\alpha = .71$)	3.81 _a (0.94)	4.12 _b (1.02)	-	8.43	.004	.010 [.002, .023]
Character	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (2, 1333)	<i>p</i>	η_p^2 [90% CI]
Warmth ($\alpha = .92$)	4.41 _a (1.11)	3.60 _b (1.27)	3.75 _b (1.22)	25.28	<.001	.037 [.021, .054]
Competence ($\alpha = .78$)	4.53 _a (0.89)	4.20 _b (1.0)	4.31 _b (0.94)	5.44	.004	.008 [.002, .017]
Prejudice ($\alpha = .91$)	4.12 _a (1.27)	4.79 _b (1.35)	4.79 _b (1.27)	18.75	<.001	.027 [.014, .043]
Morality ($\alpha = .92$)	4.39 _a (1.09)	3.54 _b (1.25)	3.69 _b (1.15)	30.58	<.001	.044 [.027, .062]

Note. Means that do not share a subscript in a row differ at $p < .05$. Post hoc tests are Tukey tests.

Perceptions of Character

Consistent with the hypothesis, a 3 (response type: acknowledge, deny, or control) $\times$ 2 (participant race: Asian vs. White) MANOVA on perceptions of Jeff's character revealed a main effect of response type, Wilks' $\Lambda = .95$, $F(8, 2660) = 9.21$, $p < .001$, $\eta_p^2 = .027$. See Table 3 for conditional means and univariate tests of response type. As expected, Tukey's post hoc tests revealed that people viewed Jeff as warmer, less prejudiced, and more moral when he acknowledged his prejudice than when he denied or said nothing, $qs(1333) > |10.39|$, $ps < .001$, $ds > |0.62|$. There were, however, consistently no differences between when Jeff denied and said nothing, $qs(1333) < -2.73$, $ps > .120$, $ds < -0.12$. Finally, although unpredicted, Jeff's response had a similar effect on perceptions of competence. People perceived Jeff as more competent when he acknowledged than when he denied or said nothing, $qs(1333) > 4.90$, $ps < .001$, $ds > 0.24$, and there was no difference between when he denied or said nothing, $q(1333) = -2.51$, $p = .179$, $d = -0.11$.

There was also an unpredicted significant multivariate effect of participant race, Wilks' $\Lambda = .99$, $F(4, 1330) = 3.48$, $p = .008$, $\eta_p^2 = .010$. Compared to Asian participants, White participants thought Jeff was warmer, $F(1, 1333) = 9.42$, $p = .002$, $\eta_p^2 = .007$, more moral, $F(1, 1333) = 7.78$, $p = .005$, $\eta_p^2 = .006$, and more competent, $F(1, 1333) = 13.06$, $p < .001$, $\eta_p^2 = .010$. There was, however, no difference in perceptions of prejudice, $F(1, 1333) = 1.66$, $p = .198$, $\eta_p^2 = .001$.

Finally, the multivariate interaction between race and response type was not significant, Wilks' $\Lambda = .99$, $F(8, 2662) = 0.59$, $p = .789$, $\eta_p^2 = .002$.

Indirect Effect Analyses

We hypothesized that acknowledging prejudice increases perceptions of learning orientation which, in turn, has a positive effect on perceptions of character. As in Experiment 2b, we tested whether acknowledging prejudice had an indirect effect on character that operated through perceptions of learning orientation. The acknowledgment condition was specified as the reference group (0) and compared to both the denial (1) and control conditions (1). In addition, participant race was added to the model as a covariate. We tested separate models for each outcome.

Warmth. The relative indirect effects of whether Jeff acknowledged prejudice versus denied, $b = -.52$, $SE = .06$, 95% CI $[-.63, -.41]$, and acknowledged versus said nothing, $b = -.50$, $SE = .05$, 95% CI $[-.61, -.40]$, on perceptions of warmth were both significant. See SOM Figure 3S.

Prejudice. The relative indirect effects of whether Jeff acknowledged prejudice versus denied, $b = .41$, $SE = .05$, 95% CI $[.31, .51]$, and acknowledged versus said nothing, $b = .39$, $SE = .05$, 95% CI $[.30, .49]$, on perceptions of Jeff's prejudice were both significant.

Morality. The relative indirect effects of whether Jeff acknowledged prejudice versus denied, $b = -.50$, $SE = .05$, 95% CI $[-.60, -.39]$ and acknowledged versus said nothing, $b = -.48$, $SE = .05$, 95% CI $[-.58, -.38]$, on perceptions of Jeff's morality were both significant.

In addition, we also tested alternate models in which acknowledging prejudice lowers perceptions of performance orientation, which in turn, has a negative impact on perceptions of character. However, there was very limited support for this alternative model (i.e., one out of six hypothesis tests showed support, see the SOM for details).

Discussion

Overall, Experiment 3 demonstrated that acknowledging anti-Asian prejudice was perceived more favorably than denying anti-Asian prejudice or saying nothing following a confrontation. Moreover, although White people tended to view the actor's character more positively than Asian people, White and Asian people alike perceived a man who acknowledged his prejudice more positively and as more learning-oriented than someone who denied or said nothing. Finally, replicating Experiment 2b, Experiment 3 suggested that acknowledging has an indirect effect on character evaluations that operates through perceptions of learning orientation.

Experiment 4: "I can understand why someone would be upset by that"

The goal of Experiment 4 was to compare two different types of acknowledgments to better understand why acknowledgement leads to more positive character inferences. At a minimum, an acknowledgement of prejudice must include a recognition that one's behavior was influenced by prejudice. Across the first four experiments, however, the actor's acknowledgment included not only a recognition of prejudiced behavior, but it also communicated a willingness to take the confronter's perspective (e.g., "I see what you mean"). We initially hypothesized that a validation of the confronter's perspective may need to accompany the recognition of prejudiced behavior for an acknowledgment to be perceived as more learning-oriented than a denial. This hypothesis would be consistent with the literature on mindsets and prejudice which characterizes learning orientation as an outward focus on other people. People who adopt learning goals often consider the perspective of others (Carr et al., 2012; Murphy et al., 2011; Neel & Shapiro, 2012).

In Experiment 4, we directly examined how a focus on the other person and their perspective affects perceptions of learning motivation and character evaluations. We manipulated whether the perpetrator recognized his behavior was influenced by prejudice and

subsequently validated the confronter's perspective (other-focused acknowledgment) or, instead, recognized his behavior was influenced by prejudice and subsequently communicated he was unwilling to take the confronter's perspective (self-focused acknowledgment). In practice, we suspect that self-focused acknowledgements in which people recognize their behavior was influenced by prejudice but subsequently communicate they are unwilling to take the confronters' perspective are rare. Nonetheless, by comparing two different types of acknowledgments to an outright denial of prejudice, the present design allows for a more fine-grained examination of the mechanism that leads to more positive character evaluations when people acknowledge their prejudice. In Experiment 4, we omitted the no information control condition to conserve power and because three previous experiments indicated there were no differences between the denial condition and the no-information control condition.

In Experiment 4, White and Black people judged a man who acknowledged his prejudice with an other-focus, acknowledged with a self-focus, or denied. Next, they evaluated the man's learning and performance orientation and his character. We predicted that an other-focused acknowledgment would be perceived as more learning-oriented than either of the other two conditions. We also predicted that people would evaluate the man as warmer, less prejudiced, and more moral when he acknowledged with an other-focus than when he acknowledged with a self-focus or denied. Finally, we predicted the same pattern of results among White and Black participants. We preregistered our hypotheses and analysis plan: <https://osf.io/cbeu7/>

Method

Participants and Procedure

There were 657 people—403 Black, 254 White—included in the final sample (283 men, 370 women, 2 non-binary/other, $M_{age}=44.54$, $SD=17.09$). Based on the same power analysis as

Experiment 3, we sought to collect data from 372 adults in the U.S. using Prime Panels (Chandler et al., 2019). The same programming error once again resulted in additional participants. Out of 743 participants, we dropped 84 people who did not indicate their race as either monoracial Black or White and two people for excessive missing data. A sensitivity analysis revealed that with 657 participants we had 80% power to detect an interaction effect size of $\eta_p^2 = .008$ in a 3 (response type) $\times$ 2 (participant race) $\times$ 2 (motivation type) mixed model ANOVA with repeated measures on the last factor.

Participants read the scenario from Experiments 2a and 2b with two modifications. We changed the name of the perpetrator from Brandon to Jeff, and we changed Jeff's response to the confrontation. After Jeff was confronted, people were randomly assigned to read an other-focused acknowledgement, a self-focused acknowledgement, or a denial. In the other-focused condition, Jeff said, "You're right, that was prejudiced of me. I was just trying to be funny and it was the first thing that came to mind. On second thought, I can understand why someone would be upset by that." In the self-focused condition Jeff said, "You're right, that was prejudiced of me. But, I was just trying to be funny and it was the first thing that came to mind. I really can't understand why anyone would be upset by that." Thus, the other-focus and the self-focused acknowledgement conditions both contained an admission of prejudice. However, in the other-focused condition, the perpetrator validated the confronter's perspective whereas in the latter case he indicated an unwillingness to take the confronter's perspective. Finally, in the denial condition, Jeff said: "I disagree, I don't think that was prejudiced at all. I was just trying to be funny and it was the first thing that came to mind. I really can't understand why anyone would be upset by that."

Measures

People completed the same measures used in Experiment 3; Table 4 displays scale reliabilities. The grand means, standard deviations, and correlations between measures are in the SOM.

Results

Perceived Performance and Learning Orientation

We performed a 3 (response type: other-focused acknowledgement, self-focused acknowledgement, denial) $\times$ 2 (participant race: Black vs. White) $\times$ 2 (motivation type: performance vs. learning) mixed model ANOVA with repeated measures on the last factor. There was a significant effect of motivation type, $F(1, 651) = 184.13, p < .001, \eta_p^2 = .220$. Overall, people perceived Jeff as more performance-oriented than learning-oriented. There was also a significant main effect of response type, $F(2, 651) = 11.38, p < .001, \eta_p^2 = .034$. More importantly, the predicted motivation type $\times$ response type interaction was significant, $F(2, 651) = 7.81, p < .001, \eta_p^2 = .023, 90\% \text{ CI } [.007, .044]$ See Table 4 for conditional means and univariate tests of response type.

Tukey's post hoc tests revealed that people viewed Jeff as more learning-oriented when he acknowledged with an other-focus than when he acknowledged with a self-focus or denied, $qs(651) > 4.63, ps < .01, ds > 0.32$. There was no difference when Jeff acknowledged with a self-focus versus denied, $q(651) = 2.57, p = .166, d = 0.17$. Because the univariate effect of response type on performance motivation was not significant, we did not perform post hoc tests.

Of note, the main effect of race and all interactions involving race in the mixed model ANOVA were not significant, $F_s(2, 651) < 2.03, ps > .154, \eta_p^2s < .003$.

Table 4*Experiment 4: Means, standard deviations, and inferential statistics*

	Response Type					
	Other-Focus	Self-Focus	Deny			
Motivation	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (2, 651)	<i>p</i>	η_p^2 [90% CI]
Learning ($\alpha = .93$)	4.11 _a (1.41)	3.64 _b (1.52)	3.39 _b (1.55)	13.74	<.001	.041 [.018, .066]
Perform ($\alpha = .73$)	4.70 _a (1.10)	4.76 _a (1.08)	4.55 _a (1.09)	2.49	.084	.008 [0, .021]
Response	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (2, 651)	<i>p</i>	η_p^2 [90% CI]
Appropriate ($\alpha = .92$)	4.01 _a (1.32)	2.82 _b (1.51)	2.77 _b (1.44)	57.91	<.001	.151 [.110, .191]
Typical ($\alpha = .86$)	4.10 _a (1.11)	4.48 _b (1.06)	4.45 _b (1.00)	7.23	<.001	.022 [.006, .042]
Character	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (2, 651)	<i>p</i>	η_p^2 [90% CI]
Warmth ($\alpha = .87$)	3.70 _a (1.31)	3.36 _b (1.37)	3.15 _b (1.41)	11.63	<.001	.035 [.014, .059]
Competence ($\alpha = .70$)	4.16 _a (1.08)	3.96 _{ab} (1.13)	3.79 _b (1.17)	6.48	.002	.020 [.005, .039]
Prejudice ($\alpha = .81$)	4.79 _a (1.29)	5.22 _b (1.18)	5.20 _b (1.28)	10.14	<.001	.030 [.011, .053]
Morality ($\alpha = .80$)	3.67 _a (1.34)	3.29 _{ab} (1.34)	3.10 _b (1.42)	11.75	<.001	.035 [.014, .059]

Note: Means that do not share a subscript in a row differ at $p < .05$. Post hocs are Tukey tests.

Perceptions of the Response

A 3(response type: other-focused acknowledgement, self-focused acknowledgement, denial) $\times$ 2(participant race: Black vs. White) MANOVA on perceptions of Jeff's response revealed a main effect of response type, Wilks' $\Lambda = .85$, $F(2, 1300) = 28.59$, $p < .001$, $\eta_p^2 = .081$, a main effect of participant race, Wilks' $\Lambda = .96$, $F(2, 650) = 12.38$, $p < .001$, $\eta_p^2 = .037$, and an unexpected response type $\times$ race interaction, Wilks' $\Lambda = .98$, $F(4, 1300) = 3.22$, $p = .012$, $\eta_p^2 = .010$.

Univariate analyses showed that, for appropriateness, there was a significant effect of response type, $F(2, 651) = 57.91$, $p < .001$, $\eta_p^2 = .151$, race, $F(1, 651) = 9.89$, $p = .002$, $\eta_p^2 = .015$, and a response type $\times$ race interaction, $F(2, 651) = 4.75$, $p = .009$, $\eta_p^2 = .014$. See Figure 2. Black participants rated the other-focused acknowledgement as more appropriate than both the self-focused acknowledgement and the denial, $qs(651) > |7.39|$, $ps < .001$, $ds > |0.62|$; however, ratings of the self-focused acknowledgement and the denial did not differ, $q(651) = |0.48|$, $p = .999$, $d = |0.04|$.⁶

White participants also rated the other-focused acknowledgement as more appropriate than both the self-focused acknowledgement and the denial, $qs(651) > |10.77|$, $ps < .001$, $ds > |1.28|$. White participants' ratings of the self-focused acknowledgement and the denial also did not differ, $q(651) = |0.14|$, $p = .999$, $ds = |0.01|$.

For typicality, there was a significant effect of response type, $F(2, 651) = 7.23$, $p = .001$, $\eta_p^2 = .022$, and race, $F(2, 651) = 18.97$, $p < .001$, $\eta_p^2 = .028$, but the interaction was not significant, $F(2, 651) = 0.94$, $p = .393$, $\eta_p^2 = .003$. Participants perceived an other-focused acknowledgement as less typical than a self-focused acknowledgement or a denial, $qs(651) > |$

⁶ $k = 6$ for all response type $\times$ race interactions.

4.91|, $ps < .002$, $ds > |0.33|$ and their perceptions of a self-focused acknowledgement and a denial did not differ, $q(651) = |0.17|$, $p = .992$, $d = |0.01|$. See Table 4.

Perceptions of Character

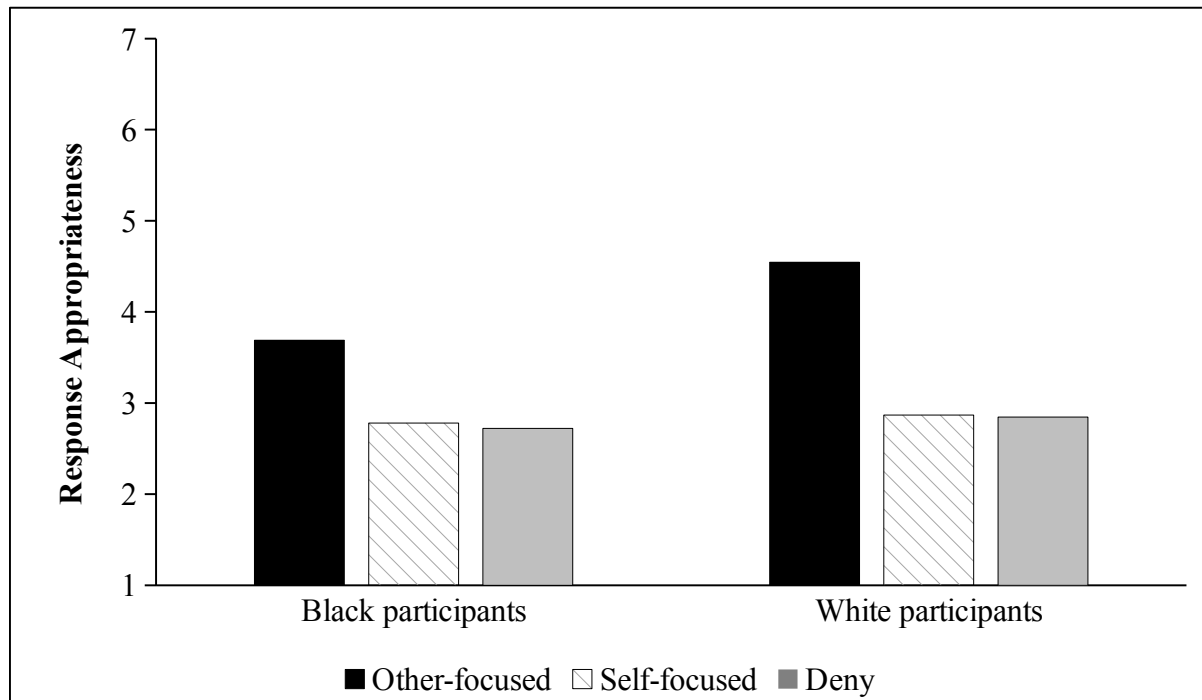
A 3 (response type: other-focused acknowledgement, self-focused acknowledgement, denial) $\times$ 2 (participant Race: Black vs. White) MANOVA on perceptions of character revealed the predicted main effect of response type, Wilks' $\Lambda = .95$, $F(8, 1298) = 4.06$, $p < .001$. (See Table 4 for conditional means and univariate tests of response type). Tukey's post hoc tests revealed that people perceived Jeff as significantly warmer, less prejudiced, and more moral when he acknowledged with an other-focus as compared to when he acknowledged with a self-focus or denied, $qs(651) > |3.82|$, $ps < .020$, $ds > |0.26|$. For all measures, there was no difference between acknowledging with a self-focus and denying, $qs(651) < 2.34$, $ps > .224$, $ds < 0.15$. Finally, although we did not make hypotheses about competence, people saw Jeff as more competent when he acknowledged with an other-focus as compared to when he denied, $q(651) = 4.93$, $p < .002$, $d = 0.33$.

There was also an unpredicted significant multivariate effect of participant race, Wilks' $\Lambda = .94$, $F(4, 648) = 10.85$, $p < .001$, $\eta_p^2 = .063$. White people thought Jeff was warmer, $F(1, 651) = 35.69$, $p < .001$, $\eta_p^2 = .052$, less prejudiced, $F(1, 651) = 18.96$, $p < .001$, $\eta_p^2 = .028$, more moral, $F(1, 651) = 22.17$, $p < .001$, $\eta_p^2 = .033$, and more competent, $F(1, 651) = 22.95$, $p < .001$, $\eta_p^2 = .034$, than Black people.

Of note, the multivariate interaction between response type and race was not significant, $F(8, 1298) = 1.07$, $p = .379$.

Figure 2

Experiment 4: Effects of acknowledging and participant race on perceptions of an appropriate response



Indirect Effect Analyses

We did not pre-register indirect effect analyses as part of our analysis plan. Nonetheless, we report the results here for comparison with previous experiments. We tested whether acknowledging with an other-focus had an indirect effect on character that operated through perceptions of learning motivation. We specified the other-focused acknowledgment condition (0) as the reference group and compared it to both the self-focused acknowledgement (1) and the denial conditions (1) and tested separate models for each measure of character. In addition, participant race was added to the model as a covariate.

Warmth. The relative indirect effects of whether Jeff acknowledged prejudice with an other-focus versus acknowledged with a self-focus, $b = -.22$, $SE = .07$, 95% CI $[-.36, -.09]$, and

acknowledged with an other-focus versus denied, $b = -.34$, $SE = .07$, 95% CI $[-.49, -.20]$, on perceptions of warmth were both significant. See SOM Figure 5S.

Prejudice. The relative indirect effects of whether Jeff acknowledged prejudice with an other-focus versus acknowledged with a self-focus, $b = .12$, $SE = .04$, 95% CI $[.05, .21]$, and acknowledged with an other-focus versus denied, $b = .19$, $SE = .05$, 95% CI $[.11, .29]$, on perceptions of prejudice were both significant.

Morality. The relative indirect effects of whether Jeff acknowledged prejudice with an other-focus versus acknowledged with a self-focus, $b = -.24$, $SE = .07$, 95% CI $[-.39, -.10]$ and acknowledged with an other-focus versus denied, $b = -.36$, $SE = .08$, 95% CI $[-.52, -.22]$, on perceptions of morality were both significant.

In addition, we also tested an alternate model in which other-focused acknowledgment lowers perceptions of performance motivation, which in turn, has a negative impact on perceptions of character. However, there no support for this alternative model, see the SOM for details.

Discussion

Both White and Black participants perceived more learning orientation in an other-focused acknowledgment than a self-focused acknowledgement. Even though Black participants viewed the actor more negatively than White participants, an other-focused acknowledgment led to more positive evaluations of character among Black and White participants alike as compared to a self-focused acknowledgment or denying. In fact, when an actor admitted to having acted with prejudice, but refused to take the perspective of the confronter, the actor was evaluated just as negatively as if he denied that he acted with prejudice altogether. Our indirect effect analysis

suggested that the positive effects of other-focused acknowledgments on character evaluations operated through perceptions of learning orientation.

We initially hypothesized that a validation of the confronter's perspective may need to accompany a recognition of prejudiced behavior for an acknowledgment to be perceived as more learning-oriented than a denial. The findings of Experiment 4 are broadly consistent with this hypothesis. However, an alternative explanation is that when a perpetrator recognizes prejudiced behavior but is explicitly unwilling to take the confronter's perspective, that unwillingness to perspective-take cancels the positive effects of recognizing prejudiced behavior. One limitation of Experiment 4 is that it did not include a condition in which the actor merely acknowledged that their behavior was prejudiced. Thus, one question left unaddressed by Experiment 4 is whether a simple recognition that one's behavior was influenced by prejudiced, in the absence of a validation (or refutation) of the confronter's perspective, leads to more positive character evaluations than a denial of prejudice. We turned to this question in our final experiment.

Experiment 5: "Yes, that was prejudiced"

Like Experiment 4, the goal of Experiment 5 was to compare two types of acknowledgments to understand why acknowledgement leads to more positive character inferences. However, instead of comparing an other-focused acknowledgement to a self-focused acknowledgement, Experiment 5 compared an other-focused acknowledgement to an acknowledgement in which the actor merely recognized his behavior was influenced by prejudice. In addition, we simplified the denial condition so that the actor only stated that he did not think his behavior was influenced by prejudiced. Thus, in Experiment 5, White and Black people judged a man who acknowledged his prejudice with an other-focus, merely

acknowledged his prejudice, or denied. Next, they evaluated the man's learning and performance orientation and his character.

We initially hypothesized that a validation of the confronter's perspective may need to accompany a recognition of prejudiced behavior for an acknowledgment to be perceived as more learning-oriented than a denial. If this hypothesis is correct, participants will have more positive character evaluations of a man who makes an other-focused acknowledgement as compared to a mere acknowledgement, whereas a man who makes a mere acknowledgement will be evaluated similarly as man who denies. However, an alternative hypothesis is that a mere acknowledgement of prejudice is viewed as more learning-oriented than an outright denial, and that when acknowledgment is coupled with an other-focus it further enhances the perceived learning-orientation of an acknowledgment. If this hypothesis is correct, then participants will have more positive character evaluations of a man who makes a mere acknowledgement as compared to a denial, and participants will have more positive character evaluations of a man who makes an other-focused acknowledgement than a man who merely acknowledges. Experiment 5 was designed to test between these two competing hypotheses. Accordingly, we pre-registered them along with our analysis plan: <https://osf.io/cbeu7/>

Method

Participants and Procedure

A total of 384 U.S. adults ($M_{age} = 41.28$, $SD = 12.72$, 141 men, 239 women, 4 non-binary/other) from CloudResearch's Connect platform participated in an experiment we expected to take five minutes. Participants were paid \$1.25. We excluded 29 participants who reported their race as something other than monoracial White or Black, leaving 177 White people and 178 Black people in the final sample. A sensitivity analysis revealed that with 355 participants we

had 80% power to detect an interaction effect size of $\eta_p^2 = .016$ in a 3 (response type) $\times$ 2 (participant race) $\times$ 2 (motivation type) mixed model ANOVA with repeated measures on the last factor.

Participants read the scenario from Experiment 4 with two modifications. We changed how the perpetrator in the scenario responded to the confrontation and we added manipulation checks to measure the effectiveness of our manipulation. Based on extensive pilot testing (see SOM), we created three responses to the confrontation that a) merely acknowledged prejudice, b) acknowledged prejudice while focusing on the confronter's perspective, or c) denied prejudice. Participants were randomly assigned to condition.

In the mere acknowledgement condition Jeff said, "Yes, I think that was prejudiced of me." In the other-focused acknowledgement condition Jeff said, "Yes, I think that was prejudiced of me. On second thought, I can understand why someone would be upset by that." Finally, in the denial condition, Jeff said: "No, I don't think that was prejudiced of me."

Measures

People completed the same measures used in Experiment 4 plus the manipulation checks described below. Table 5 displays scale reliabilities. The grand means, standard deviations, and correlations between measures are in the SOM.

Manipulation Checks

We measured the effectiveness of our manipulations with two items. One item measured the type of response Jeff gave ("To what extent would you describe Jeff's response as an acknowledgement that he engaged in prejudiced behavior?") and the other item measured whether the response was other-focused ("Jeff's response shows that he tried to consider the perspective of other people"). Four additional manipulation check items that are conceptually in

line with what is reported here and that show a similar pattern of results are available in our open datasets.

Results

Manipulation Checks

Participants perceived large differences between conditions based on Jeff's response. For the question measuring how much the response appeared to acknowledge prejudice there was a large effect of condition, $F(2, 352) = 316.21, p < .001, \eta_p^2 = .642$. Participants perceived Jeff as acknowledging more in the mere acknowledgment ($M = 5.90, SD = 1.21$) and other-focused acknowledgment conditions ($M = 5.76, SD = 1.31$) than the denial condition ($M = 1.96, SD = 1.54$), $qs(352) > 30.37, ps < .001, ds > 2.66$. Participants did not perceive differences in acknowledging in the mere-acknowledgment and the other-focused acknowledgment conditions, $q(352) = 1.13, p = .704, d = 0.11$.

There were also large group differences in how much people thought Jeff tried to consider the perspective of other people, $F(2, 381) = 152.41, p < .001, \eta_p^2 = .464$. Participants perceived Jeff as taking the confronter's perspective more in the other-focused acknowledgment ($M = 5.03, SD = 1.44$) than in the mere acknowledgment condition ($M = 3.86, SD = 1.87$), $q(352) = 8.55, p < .001, d = 0.71$, and they perceived Jeff to perspective-take significantly more in the mere acknowledgment than the denial condition ($M = 1.65, SD = 1.10$), $q(352) = 15.83, p < .001, d = 1.43$.

Perceived Performance and Learning Orientation

We performed a 3 (response type: mere acknowledgement, other-focused acknowledgement, denial) $\times$ 2 (participant race: Black vs. White) $\times$ 2 (motivation type: performance vs. learning) mixed model ANOVA with repeated measures on the last factor.

There was a significant effect of motivation type, $F(1, 349) = 403.73, p < .001, \eta_p^2 = .536$.

Overall, people perceived Jeff as more performance-oriented than learning-oriented. There was also a significant main effect of response type, $F(2, 349) = 34.18, p < .001, \eta_p^2 = .164$. More importantly, the predicted motivation type $\times$ response type interaction was significant, $F(2, 349) = 28.15, p < .001, \eta_p^2 = .139, 90\% \text{ CI } [.084, .192]$. See Table 4 for conditional means and univariate tests of response type.

Tukey's post hoc tests revealed that people viewed Jeff as more learning-oriented in the other-focused acknowledgement condition than in the mere acknowledgement or denial conditions, $qs(349) > 6.97, ps < .01, ds > 0.60$. However, people also viewed Jeff as more learning-oriented when he merely acknowledged versus denied, $q(349) = 7.35, p < .001, d = 0.70$. Because the univariate effect of response type on performance motivation was not significant, we did not perform post hoc tests.

Of note, the main effect of race was significant in the mixed model ANOVA, $F(1, 349) = 5.50, p = .02$, but all interactions involving race in the mixed model ANOVA were not, $Fs(2, 349) < 1.74, ps > .177$.

Perceptions of the Response

A 3 (response type: mere acknowledgement, other-focused acknowledgement, denial) $\times$ 2 (participant race: Black vs. White) MANOVA on perceptions of Jeff's response revealed a main effect of response type, Wilks' $\Lambda = .56, F(4, 696) = 58.46, p < .001, \eta_p^2 = .251$, a main effect of race, Wilks' $\Lambda = .95, F(2, 348) = 8.90, p < .001, \eta_p^2 = .049$, and an unexpected race $\times$ response type interaction, Wilks' $\Lambda = .93, F(4, 696) = 6.23, p < .001, \eta_p^2 = .035$.

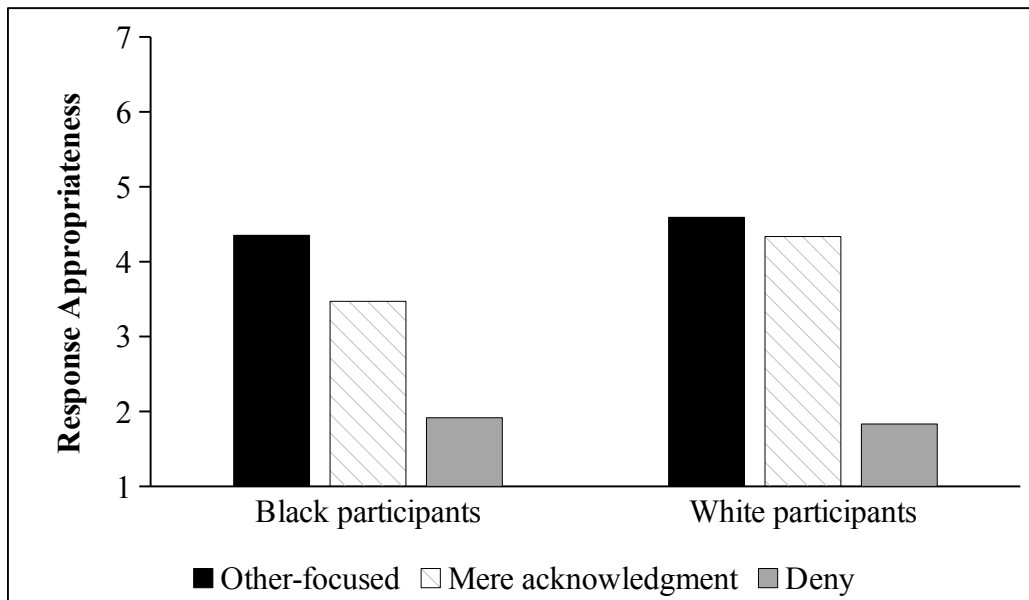
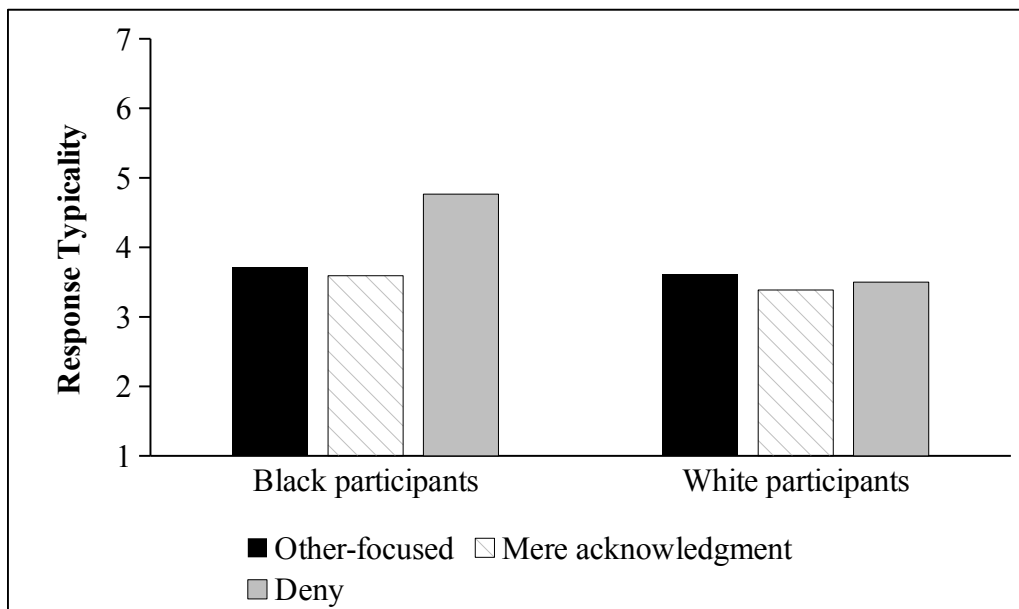
Table 5*Experiment 5: Means, standard deviations, and inferential statistics*

	Response Type					
	Mere Acknowledge	Other-focused Acknowledge	Deny			
Motivation	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (2, 349)	<i>p</i>	η_p^2 [90% CI]
Learning ($\alpha = .94$)	2.98 _a (1.32)	3.77 _b (1.31)	2.14 _c (1.07)	50.95	<.001	.226 [.163, .284]
Perform ($\alpha = .65$)	4.49 _a (1.06)	4.74 _a (0.93)	4.61 _a (1.05)	1.80	.167	.010 [0, .031]
Response	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (2, 349)	<i>p</i>	η_p^2 90% CI]
Appropriate ($\alpha = .94$)	3.89 _a (1.47)	4.47 _b (1.46)	1.87 _c (0.86)	133.70	<.001	.434 [.370, .486]
Typical ($\alpha = .86$)	3.49 _a (1.27)	3.66 _a (1.34)	4.13 _b (1.48)	7.46	<.001	.041 [.011, .077]
Character	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>F</i> (2, 349)	<i>p</i>	η_p^2 [90% CI]
Warmth ($\alpha = .91$)	3.17 _a (1.28)	3.81 _b (1.16)	2.71 _c (1.20)	26.25	<.001	.131 [.078, .183]
Competence ($\alpha = .75$)	4.22 _b (1.07)	4.60 _a (0.96)	4.17 _b (1.09)	6.37	.002	.035 [.008, .069]
Prejudice ($\alpha = .88$)	5.63 _a (1.09)	5.16 _b (1.08)	6.10 _c (0.97)	24.62	<.001	.124 [.071, .173]
Morality ($\alpha = .91$)	2.98 _a (1.27)	3.58 _b (1.13)	2.53 _c (1.15)	25.19	<.001	.126 [.074, .178]

Note: Means that do not share a subscript in a row differ at $p < .05$. Post hocs are Tukey tests.

Univariate analyses showed that, for appropriateness, there was a main effect of response type, $F(2, 349) = 133.70, p < .001, \eta_p^2 = .434$, a main effect of race, $F(1, 349) = 6.29, p < .013, \eta_p^2 = .018$, and a significant interaction, $F(2, 349) = 4.19, p = .016, \eta_p^2 = .023$. See Figure 4. Among Black participants, a mere acknowledgement was rated as more appropriate than a denial, $q(349) = 9.34, p < .001, d = 1.24$, and an other-focused acknowledgement was rated as more appropriate than a mere acknowledgement, $q(349) = 5.37, p = .002, d = 0.58$. Among White participants, a mere acknowledgement was also rated as more appropriate than a denial, $q(349) = 14.92, p < .001, d = 2.29$; however, an other-focused acknowledgement was not rated as significantly more appropriate than a mere acknowledgement, $q(349) = 1.55, p = .880, d = .19$.

For typicality, there was also a main effect of response type, $F(2, 349) = 7.46, p < .001, \eta_p^2 = .041$, a main effect of race, $F(1, 349) = 14.03, p < .001, \eta_p^2 = .039$, and a significant interaction, $F(2, 349) = 6.99, p = .001, \eta_p^2 = .039$. Among Black participants, a denial was seen as more typical than a mere acknowledgement, $q(349) = 6.83, p < .001, d = .89$, and an other-focused acknowledgement, $q(349) = 6.12, p < .001, d = .79$; however a mere acknowledgement and an other-focused acknowledgment did not differ in perceived typicality, $q(349) = .69, p = .997, d = .09$. Among White participants, there were no differences between any of the conditions in perceived typicality, $qs(349) < 1.30, ps > .94, ds < 0.18$.

Figure 4*Experiment 5: Perceptions of Response Appropriateness***Figure 5***Experiment 5: Perceptions of Response Typicality*

Perceptions of Character

A 3(response type: mere acknowledgement, other-focused acknowledgement, denial) $\times$ 2(participant race: Black vs. White) MANOVA on perceptions of character revealed the predicted main effect of response type, Wilks' $\Lambda = .85$, $F(8, 692) = 7.55$, $p < .001$, $\eta_p^2 = .080$. See Table 5 for conditional means and univariate tests of response type. Tukey's post hoc tests revealed that people perceived Jeff as significantly warmer, less prejudiced, more moral, and more competent when he acknowledged with an other-focus as compared to when he merely acknowledged or denied, $qs(349) > |3.70|$, $ps < .01$, $ds > |0.37|$. In addition, people also perceived Jeff as significantly warmer, less prejudiced, and more moral when he merely acknowledged as compared to when he denied, $qs(349) > 4.18$, $ps < .01$, $ds < 0.35$, but there was no difference in perceptions of competence, $q(349) = 0.53$, $p = .60$, $d = 0.05$.

There was also an unpredicted significant multivariate effect of participant race, Wilks' $\Lambda = .91$, $F(4, 346) = 8.91$, $p < .001$, $\eta_p^2 = .093$. White people thought Jeff was warmer, $F(1, 349) = 28.95$, $p < .001$, $\eta_p^2 = .077$, less prejudiced, $F(1, 349) = 7.11$, $p = .008$, $\eta_p^2 = .020$, more moral, $F(1, 349) = 20.63$, $p < .001$, $\eta_p^2 = .056$, and more competent, $F(1, 349) = 25.56$, $p < .001$, $\eta_p^2 = .068$, than Black people.

Of note, the multivariate interaction between response type and race was not significant, Wilks' $\Lambda = .97$, $F(8, 692) = 1.37$, $p = .207$, $\eta_p^2 = .016$.

Indirect Effect Analyses

We did not pre-register indirect effect analyses as part of our analysis plan. Nonetheless, we report the results here for comparison with previous experiments. We tested whether acknowledging with an other-focus had an indirect effect on character that operated through perceptions of learning motivation. We specified the other-focused acknowledgment condition

(0) as the reference group and compared it to both the mere acknowledgement (1) and the denial conditions (1) and tested separate models for each measure of character. In addition, participant race was added to the model as a covariate.

Warmth. The relative indirect effects of whether Jeff acknowledged prejudice with an other-focus versus merely acknowledged, $b = -.43$, $SE = .10$, 95% CI $[-.64, -.24]$, and acknowledged with an other-focus versus denied, $b = -.89$, $SE = .11$, 95% CI $[-1.14, -.69]$, on perceptions of warmth were both significant. See SOM Figure 7S.

Prejudice. The relative indirect effects of whether Jeff acknowledged prejudice with an other-focus versus merely acknowledged, $b = .34$, $SE = .08$, 95% CI $[.19, .50]$, and acknowledged with an other-focus versus denied, $b = .70$, $SE = .09$, 95% CI $[.52, .89]$, on perceptions of prejudice were both significant.

Morality. The relative indirect effects of whether Jeff acknowledged prejudice with an other-focus versus merely acknowledged, $b = -.42$, $SE = .10$, 95% CI $[-.62, -.24]$ and acknowledged with an other-focus versus denied, $b = -.86$, $SE = .11$, 95% CI $[-1.09, -.66]$, on perceptions of morality were both significant.

In addition, we also tested an alternate model in which other-focused acknowledgment lowers perceptions of performance motivation, which in turn, has a negative impact on perceptions of character. However, there no support for this alternative model, see the SOM for details.

Discussion

The goal of Experiment 5 was to determine whether a mere acknowledgement of biased behavior was sufficient to improve character evaluations relative to a denial, or whether a validation of the confronter's perspective must accompany a recognition of biased behavior to be

effective at improving character evaluations. Although we initially hypothesized that validation of the confronter's perspective must accompany recognition of biased behavior to improve character evaluations, this hypothesis was wrong. Instead, Experiment 5 found support for an alternative hypothesis. Mere acknowledgement of prejudiced behavior was perceived as more learning-oriented than an outright denial and led to more positive character evaluations than a denial. An other-focused acknowledgment that coupled recognition of biased behavior with validation of the confronter's perspective was perceived as more learning-oriented and led to more positive character evaluations than a mere acknowledgement. These results suggest that a mere acknowledgment of prejudice following confrontation is sufficient to improve character evaluations relative to an outright denial, but that including a validation of the confronter's perspective further enhances the value of acknowledgment, leading to the most positive outcomes. Moreover, the results of the indirect effect analysis are consistent with our argument that the positive effects of acknowledgment on character evaluation may operate through perceived learning-orientation. Finally, it is interesting to note the racial differences in how participants evaluated the confronter's response. Black participants distinguished between the two types of acknowledgment and the denial when considering how appropriate the response was, whereas White participants perceived the two types of acknowledgment as equally appropriate. Black participants also viewed denial as a much more typical response to confrontation while White participants viewed all responses as equally typical.

General Discussion

People generally form negative impressions of perpetrators who are confronted for expressing prejudice. The present series of experiments, however, suggest that a perpetrator's response to confrontation can influence how other people evaluate the perpetrator's character.

People evaluated a perpetrator who responded to confrontation by acknowledging prejudice as warmer, more moral, and somewhat ironically, as less prejudiced than a perpetrator who responded by denying or a perpetrator who did not respond. Moreover, the benefit of acknowledging prejudice on evaluations of a perpetrator's character occurred for expressions of sexism (Experiment 1), anti-Black (Experiment 2b, 4, 5), and anti-Asian (Experiment 3) racism. Finally, people who belonged to the perpetrator's group (i.e., men, White people) as well as people who belonged to the target's group (i.e., women, Black people, and Asian people) all evaluated a perpetrator who acknowledged prejudice more positively than a perpetrator who denied.

Beyond the impact of acknowledgment on character evaluations, people also perceived acknowledging prejudice as a more appropriate response to confrontation than denying. This finding is noteworthy given evidence that people generally perceived acknowledgements as a less typical response to confrontation than denial, a phenomenon that is consistent with studies showing that people confronted for prejudice may be more likely to respond by denying than acknowledging (see Czopp et al., 2006; Winslow, 2004; although cf. Chaney & Sanchez, 2018). Indeed, people may deny prejudice in response to confrontation because being confronted with evidence of one's own bias makes people defensive (Howell et al., 2015; Howell et al., 2017). Although denials may be a relatively common response to prejudice confrontation, denials are likely to be perceived especially negatively by other people.

We theorized that acknowledgements of prejudice in response to confrontation lead to more positive character evaluations than denials because acknowledgements signal that the perpetrator has a learning orientation to prejudice and intergroup relations. Acknowledging that one has expressed prejudice may communicate to others that one is willing to take the

confronter's perspective and to learn about the cause of the offense. Consistent with this argument, a perpetrator who acknowledged his prejudice was perceived as having a more learning-oriented approach to prejudice and intergroup relations than a perpetrator who denied his prejudice or a perpetrator who did not respond to confrontation. Moreover, statistical mediation analyses suggested that acknowledging had an indirect effect on evaluations of character (e.g., warmth, prejudice, and morality) that operated through perceived learning orientation (Experiments 2b, 3, 4, and 5).

Although the mediational analyses are consistent with our hypothesis and suggestive of a mechanism, it is important to recognize the limitations of testing causal models with statistical mediation analyses (Fiedler et al., 2018). It is possible that alternative mediators explain the relationship between acknowledging prejudice and positive character evaluations. Early in our research (Experiment 2A), we hypothesized that a perpetrator who denied prejudice would be perceived as having a more performance-oriented approach to prejudice and intergroup relations than a perpetrator who acknowledged their prejudice. Thus, it is theoretically possible that denying prejudice increases perceptions of a performance orientation which, in turn, leads to more negative character evaluations (e.g., Simon et al., 2019). However, mediation analyses that tested for an indirect effect of acknowledging vs. denying on evaluations of character that operated through perceived performance orientation were almost completely unsupported (see SOM).

Of course, other mediators that were not tested in our experiments may explain the relationship between acknowledging prejudice and positive evaluations of a perpetrator's character. One possibility is that, when someone acknowledges prejudice, others assume that the prejudice was unintentional or the result of implicit bias whereas when someone denies

prejudice, people may assume that the perpetrator holds explicit prejudices. Recent research suggests that people are less likely to hold perpetrators accountable for implicit bias as compared to explicit bias (Daumeyer et al., 2019; Daumeyer et al., 2021). Because the present experiments did not assess whether participants believed the prejudice expression was the result of implicit or explicit bias, ultimately, we are not able to rule out changes in the perceived nature of the bias (i.e., explicit vs. implicit) as a possible mediating mechanism.

The desire to better understand why acknowledging prejudice leads to more positive evaluations of character led to our final two experiments that compared different types of acknowledgements. Experiment 4 compared an other-focused acknowledgement (recognition of prejudiced behavior + validation of the confronter's perspective) to a self-focused acknowledgement (recognition of prejudiced behavior + refusal to take the confronter's perspective). Experiment 5 compared an other-focused acknowledgement to a mere acknowledgement (recognition of prejudiced behavior only). We initially hypothesized that only an other-focused acknowledgement would lead to more positive character evaluations than a denial. Experiment 4 showed that a perpetrator who made a self-focused acknowledgement was perceived similarly to one who denied, and that a perpetrator who made an other-focused acknowledgement was perceived more positively than one who either made a self-focused acknowledgement or denial. However, Experiment 5 showed that a perpetrator who merely acknowledged prejudice was perceived more positively one who denied, whereas a perpetrator who made an other-focused acknowledgement was perceived more positively than one who merely acknowledged. Thus, our initial hypothesis that only an other-focused acknowledgement would lead to more positive character evaluations than a denial was not supported by Experiment 5.

Instead, Experiments 4 and 5 suggest three general conclusions. First, a mere recognition of biased behavior may be sufficient to increase perceptions of learning orientation and improve perceptions of a perpetrator's character relative to a denial. Second, focusing on the confronter and validating their perspective may further increase perceptions of learning orientation and improve perceptions of a perpetrator's character relative to merely recognizing biased behavior. Finally, recognizing biased behavior, but then refusing to take the confronter's perspective, may cancel the otherwise beneficial effects of recognizing biased behavior for perceptions of learning orientation and character.

Limitations and Future Directions

Although we initially expected people to see the perpetrator as less performance oriented when he acknowledged than when he denied, we found little difference in people's perceptions of performance orientation regardless of how the perpetrator responded. This may be because people generally expect someone confronted for prejudice to be performance-oriented; indeed, perceptions of performance orientation were consistently high across all conditions. This finding is also consistent with recent research that suggests, at least among Black perceivers, people generally expect White people to have more of a performance orientation than a learning orientation to prejudice and intergroup relations (e.g., Simon et al., 2019). Furthermore, it is also possible that, within the context of prejudice and intergroup relations, lay people do not view performance and learning orientations as incompatible within the context of intergroup relations. In support of this latter point, there was consistently either no correlation between perceptions of performance and learning orientation (Experiments 2B and 5, see also Simon et al., 2019) or a small, but positive correlation (Experiments 2A, 3, and 4). The question of how various responses to confrontation affect people's perceptions of a perpetrator's performance orientation

and whether people perceive learning and performance orientations as incompatible within the context of intergroup relations are opportunities for future research.

A second limitation of this work is that the findings may not generalize to more extreme expressions of sexism or racism (e.g., expressions that include slurs, etc.). In the present research, the actor expressed prejudice in the form of disparaging humor (Ford, 2015). When people express prejudice in the form of humor, other people are less likely to find it offensive and to hold the perpetrator accountable (Mallett et al., 2016; Woodzicka & Ford, 2010). In fact, people frequently express their prejudice through humor because it is challenging for other people to confront (Woodzicka & Mallett, 2021). Thus, we believe the prejudice expressions in the present experiments resemble common, everyday experiences, and several pilot experiments reported in the SOM confirm that people perceived the comments as believable, prejudiced, and stereotypical. Of course, the comments' familiar nature does not mean they are not harmful to members of marginalized groups (Ford, 2015). A White man who makes a statement that a new Black colleague is incompetent or who suggests assigning extra work to an Asian colleague may contribute to an unwelcome and hostile working environment even if the target is not present at the time the comment is made.

Although we think the statements made by perpetrators in our studies are representative of prejudiced statements commonly made in the workplace, it is unclear whether our results would be replicated if the perpetrator made more extreme statements. For example, if a White person used a racial slur, was confronted, and subsequently responded by recognizing their prejudiced behavior and validating the perspective of the confronter, the response may be perceived as insincere. After all, most people know that racial slurs are unacceptable. In cases when people express extreme prejudice, they may be perceived negatively regardless of their

subsequent response to confrontation. Although the results of the present studies may not generalize to more extreme statements of bias, we nonetheless believe that our results are likely to generalize to other situations in which people express common, everyday forms of prejudice.

Finally, a third limitation is that the perpetrator acknowledged that prejudice had influenced his *behavior*, but not that he was a prejudiced *person*. In other research, people have evaluated those who admit implicit bias, discomfort in interactions with Black people, and discomfort with Black people in positions of leadership more negatively than those who deny such biases (Perry et al., 2020). Thus, acknowledging bias may not be universally perceived more positively than denying. We suspect one boundary condition of our findings may be whether someone acknowledges a behavior or attitude that can be changed (i.e., something malleable) or acknowledges a bias that others perceive as fixed (e.g., Neel & Shaprio, 2012; Simon et al., 2019).

Our findings contribute to a growing body of research examining the important role of mindsets in prejudice and intergroup relations (e.g., Carr et al., 2012; Murphy et al. 2011; Neel & Shapiro, 2012; Rattan & Dweck, 2010; Rattan & Dweck, 2018; Simon et al., 2019). For example, research suggests that mindsets can affect targets' decisions to confront prejudice (Rattan & Dweck, 2010) and that, among targets who confront prejudice, a growth mindset fosters a more positive outlook about repairing workplace relationships (Rattan & Dweck, 2018). One direction for future work then is investigating whether mindsets also affect perpetrator's decisions to acknowledge prejudice in response to confrontation. Our theoretical framework casts acknowledging as a learning-oriented behavior and denying as a performance-oriented behavior that people may engage in depending on their motivational mindset. Applied to personal transgressions, people with a malleable view of personality are more likely than those

with a fixed view to accept responsibility for their misdeeds (Schumann & Dweck, 2014). Thus, the direct implication of this research is that people who believe prejudice is malleable should be more likely to acknowledge their own prejudiced behavior than those who see prejudice as fixed.

Future work should also investigate the consequences of acknowledgment versus denial for people who are directly targeted by prejudice. Compared to hearing someone deny that their obviously prejudiced behavior was, in fact, prejudiced, hearing someone acknowledge may be a relatively less negative experience for members of target groups. Acknowledgements may indicate the perpetrator's willingness to learn from their mistake and change future behavior. And, because a willingness to learn and change are generally perceived as signs of a low prejudiced person, acknowledgments of prejudice may reduce stress in intergroup interactions when compared to hearing perpetrators deny.

Finally, like all confrontation research, examining people's reactions to stories and vignettes is a good place to begin. But, future work can build from our studies, explore other factors, and investigate whether real life interactions replicate what we have found.

Conclusion

People confronted for expressing prejudice may struggle to determine the appropriate response. Moreover, in many situations, confrontation will not be the last interaction that two people have and a perpetrator may wish to repair their relationship with the confronter and other people present. The present research indicates that people generally find acknowledging prejudice to be a much more appropriate response than denying prejudice. People are less likely to infer negative things about a person who acknowledges prejudiced behavior than a person who denies it and acknowledging may improve people's standing in the eyes of others because acknowledging signals more learning motivation than denying does. Thus, the clear implication

from our experiments is that after someone says or does something prejudiced, they may be able to move toward reconciliation and improve the interaction going forward if they can find the upside in genuinely acknowledging their prejudiced behavior.

References

- 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, 332-342. <https://doi.org/10.5465/amle.2008.34251671>
- Ashburn-Nardo, L., & Karim, M. F. A. (2019). The CPR model: Decisions involved in confronting prejudiced responses. In R. K. Mallett and M. J. Montieth (Eds.) *Confronting Prejudice and Discrimination* (pp. 29-47). Academic Press.
- Boysen, G. A. (2013). Confronting math stereotypes in the classroom: Its effect on female college students' sexism and perceptions of confronters. *Sex roles*, 69, 297-307. <https://doi.org/10.1007/s11199-013-0287-y>
- 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, 452-471. doi:10.1037/a0028849
- Chandler, J., Rosenzweig, C., Moss, A.J. et al. Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior Research Methods*, 51, 2022–2038 (2019). <https://doi.org/10.3758/s13428-019-01273-7>
- Chaney, K. E., & Sanchez, D. T. (2018). The endurance of interpersonal confrontations as a prejudice reduction strategy. *Personality and Social Psychology Bulletin*, 44, 418-429. <https://doi.org/10.1177%2F0146167217741344>
- Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin*, 129, 414–446. <https://doi.org/10.1037/0033-2909.129.3.414>

- Czopp, A. M. (2019). The consequences of confronting prejudice. In R. K. Mallett and M. J. Monteith (Eds.) *Confronting Prejudice and Discrimination* (pp. 201-221). Academic Press.
- Czopp, A. M., Kay, A. C., & Cheryan, S. (2015). Positive stereotypes are pervasive and powerful. *Perspectives on Psychological Science*, 10(4), 451-463.
<https://doi.org/10.1177%2F1745691615588091>
- Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin*, 29, 532-544. doi:10.1177/0146167202250923
- Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*, 90, 784-803. <http://dx.doi.org/10.1037/0022-3514.90.5.784>
- Daumeyer, N. M., Onyeador, I. N., Brown, X., & Richeson, J. A. (2019). Consequences of attributing discrimination to implicit vs. explicit bias. *Journal of Experimental Social Psychology*, 84, 103812. <https://doi.org/10.1016/j.jesp.2019.04.010>
- Daumeyer, N. M., Onyeador, I. N., & Richeson, J. A. (2021). Does shared gender group membership mitigate the effect of implicit bias attributions on accountability for gender-based discrimination?. *Personality and Social Psychology Bulletin*, 47(9), 1343-1357.
<https://doi.org/10.1177%2F0146167220965306>
- Dickter, C. L., Kittel, J. A., & Gyurovski, I. I. (2012). Perceptions of non-target confronters in response to racist and heterosexist remarks. *European Journal of Social Psychology*, 42, 112-119. <https://doi.org/10.1002/ejsp.855>

- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273. doi:10.1037/0033-295X.95.2.256
- Ford, T. E. (2015). The social consequences of disparagement humor: Introduction and overview. *Humor*, 28(2), 163-169. <https://doi.org/10.1515/humor-2015-0016>
- Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70, 491-512. <http://dx.doi.org/10.1037/0022-3514.70.3.491>
- Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56, 109. <https://psycnet.apa.org/doi/10.1037/0003-066X.56.2.109>
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148-168. <http://dx.doi.org/10.1037/a0034726>
- Hayes, A. F. (2017). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford Publications.
- Howell, J. L., Gaither, S. E., & Ratliff, K. A. (2015). Caught in the middle: Defensive responses to IAT feedback among whites, blacks, and biracial black/whites. *Social Psychological and Personality Science*, 6(4), 373-381. <https://doi.org/10.1177%2F1948550614561127>
- Howell, J. L., Redford, L., Pogge, G., & Ratliff, K. A. (2017). Defensive responding to IAT feedback. *Social Cognition*, 35(5), 520-562. <https://doi.org/10.1521/soco.2017.35.5.520>
- Kahalon, R., Shnabel, N., & Becker, J. C. (2018). Positive stereotypes, negative outcomes: Reminders of the positive components of complementary gender stereotypes impair

- performance in counter-stereotypical tasks. *British Journal of Social Psychology*, 57, 482-502. <https://doi.org/10.1111/bjso.12240>
- Klein, N., & Epley, N. (2014). The topography of generosity: Asymmetric evaluations of prosocial actions. *Journal of Experimental Psychology: General*, 143(6), 2366-2379. <http://dx.doi.org/10.1037/xge0000025>
- Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 49, 433-442.
- Mae, L., & Carlston, D. E. (2005). Hoist on your own petard: When prejudiced remarks are recognized and backfire on speakers. *Journal of Experimental Social Psychology*, 41(3), 240-255. <https://doi.org/10.1016/j.jesp.2004.06.011>
- Mallett, R. K., & Monteith, M. J. (Eds.). (2019). *Confronting prejudice and discrimination: The science of changing minds and behaviors*. Academic Press.
- Mallett, R. K., & Wagner, D. E. (2011). The unexpectedly positive consequences of confronting sexism. *Journal of Experimental Social Psychology*, 47, 215-220. <https://doi.org/10.1016/j.jesp.2010.10.001>
- Martinez, L. R., Hebl, M. R., Smith, N. A., & Sabat, I. E. (2017). Standing up and speaking out against prejudice toward gay men in the workplace. *Journal of Vocational Behavior*, 103, 71-85. <https://doi.org/10.1016/j.jvb.2017.08.001>
- Murphy, M. C., Richeson, J. A., & Molden, D. C. (2011). Leveraging motivational mindsets to foster positive interracial interactions. *Social and Personality Psychology Compass*, 5, 118-131. doi:10.1111/j.1751-9004.2010.00338.x

- 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, 101-120. <http://dx.doi.org/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, 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>
- Perry, S. P., Skinner, A. L., Wages, J. E., & Parzonka, J. (2020, November). *Ambivalent reactions to people who deny and admit their gender and racial biases*. Working paper. PsyArxiv: <https://doi.org/10.31234/osf.io/x47q3>
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75, 811-832. <http://dx.doi.org/10.1037/0022-3514.75.3.811>
- Plaut, V. C. (2010). Diversity science: Why and how difference makes a difference. *Psychological Inquiry*, 21, 77-99. <https://doi.org/10.1080/10478401003676501>
- 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, 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, 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, 676. <https://psycnet.apa.org/doi/10.1037/apl0000287>
- Schumann, K., & Dweck, C. S. (2014). Who accepts responsibility for their transgressions? *Personality and Social Psychology Bulletin*, 40, 1598-1610. <https://doi.org/10.1177/0146167214552789>
- 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*. <https://doi.org/10.1177/1948550618778584>
- Trawalter, S., Richeson, J. A., & Shelton, J. N. (2009). Predicting behavior during interracial interactions: A stress and coping approach. *Personality and Social Psychology Review*, 13, 243-268. doi:10.1177/1088868309345850
- Vaccarino, E., & Kawakami, K. (2020). In the office or at the gym: The impact of confronting sexism in specific contexts on support for confrontation and perceptions of others. *Self and Identity*, 1-20. <https://doi.org/10.1080/15298868.2020.1832566>
- Vorauer, J. D., Main, K. J., & O'Connell, G. B. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *Journal of Personality and Social Psychology*, 75, 917-937. doi: 10.1037/0022-3514.75.4.917
- Winslow, M. P. (2004). Reactions to the imputation of prejudice. *Basic and Applied Social Psychology*, 26, 289-297. http://dx.doi.org/10.1207/s15324834basp2604_5

Woodzicka, J. A., & Ford, T. E. (2010). A framework for thinking about the (not-so-funny) effects of sexist humor. *Europe's Journal of Psychology*, 6(3), 174-195.

<https://doi.org/10.5964/ejop.v6i3.217>

Woodzicka, J. A., & Mallett, R. K. (2021). Addressing the challenges of confronting disparagement humor. In M Strick & T. Ford (Eds.) *The Social Psychology of Humor* (pp. 189-205). Routledge.

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

<https://doi.org/10.1515/humor-2015-0025>