Promoting concern about gender bias with evidence-based confrontation

Laura R. Parker\textsuperscript{a}, Margo J. Monteith\textsuperscript{a, \Dagger}, Corinne A. Moss-Racusin\textsuperscript{b}, Amanda R. Van Camp\textsuperscript{c}

\textsuperscript{a} Purdue University, Department of Psychological Sciences, 703 Third Street, West Lafayette, IN 47907, United States
\textsuperscript{b} Skidmore College, Department of Psychology, 815 North Broadway, Saratoga Springs, NY 12866, United States
\textsuperscript{c} Tulane University, Department of Psychology, 2007 Percival Stern Hall, New Orleans, LA 70118, United States

\textbf{ARTICLE INFO}

\textbf{ABSTRACT}

Whereas confrontations of racial bias prompt negative self-directed affect (e.g., guilt) and reduce bias, confrontations of gender bias are typically disregarded. We examined the effectiveness of an evidence-based confrontation in which participants received concrete evidence of their discriminatory gender-biased evaluations. Participants were confronted with evidence that they evaluated a female applicant for a lab manager position negatively, which would have resulted in her not being hired, and that their evaluations were more negative than an otherwise identical applicant who was male. Experiment 1 found that this confrontation activated greater guilt and, in turn, concern about expressing and regulating gender bias in the future (hereafter “future concern”), relative to participants who were confronted about gender bias but without concrete evidence of bias. Experiment 2 showed that confrontation with concrete evidence of gender bias activated guilt and future concern whether the bias was framed as viewing women as not belonging in science or as unintelligent. Experiment 3 showed that the positive effects of confronting gender bias with evidence were just as strong as a parallel confrontation of racial bias. Experiment 4 showed that the positive effects also held in a non-STEM domain, and that these effects were not dependent on the situational salience of norms opposing gender bias. Thus, a robust strategy for effective confrontation of gender bias was established. However, we discuss difficulties with implementing this strategy in many real world situations, along with implications for how best to curb gender bias through confrontation.

Unequal gender representation is commonplace across many science, technology, engineering, and mathematics (STEM) fields (National Science Foundation, 2017). Although many factors contribute to this disparity (Ceci & Williams, 2011), biases against women likely play a significant role (Knobloch-Westercick, Glynn, & Huge, 2013; Reuben, Sapienza, & Zingales, 2014). For instance, gender bias was recently demonstrated experimentally in Moss-Racusin, Dovidio, Brescoll, Graham, and Handelsman’s (2012) research. Male and female science professors at research-intensive universities were asked to evaluate the application of an undergraduate student seeking a science lab manager position. All faculty participants received an identical application, with the applicant specified as either male or female. Results revealed robust gender bias, in that the female applicant was rated as less competent and less desirable to mentor than the male applicant. She was also conferred a significantly lower starting salary and ultimately rated as less hirable. These results were framed in terms of the subtle operation of gender bias, and the authors suggested that many of the faculty likely were unaware that gender had influenced their evaluations. These findings are disturbing and demonstrate the profound impact gender bias has on women’s progress in STEM.

This research also begs the questions: What if people were made aware of the fact that they disadvantaged female applicants in this way? Would they become more concerned about their propensity for gender bias and monitor future behavior to guard against it? Within social psychology, these questions fall squarely into the study of people’s reactions to being confronted about their intergroup biases. Confrontation can be a powerful tool for raising people’s awareness of their proneness to subtle but consequential biases, thereby highlighting discrepancies between personal values and actual behaviors (Czopp, Monteith, & Mark, 2006; Rokeach, 1973). Confrontation also communicatates that bias is unacceptable, which establishes situational norms opposing bias (Paluck, 2011). Thus, when people are confronted about things they have said or done that are biased, they experience self-reflective negative affect (e.g., disappointed with the self and guilt), become more concerned about their biases, and reduce subsequent biased responding (Czopp \textit{et al.}, 2006; Czopp & Monteith, 2003; Guiker, Mark, \& Monteith, 2013). However, research to date suggests that confrontations of gender bias are not met with the same reactions as

\textsuperscript{\Dagger} Corresponding author.
\textit{E-mail address: mmonteith@purdue.edu (M.J. Monteith).}

http://dx.doi.org/10.1016/j.jesp.2017.07.009
Received 15 June 2016; Received in revised form 12 July 2017; Accepted 13 July 2017
0022-1031/ © 2017 Elsevier Inc. All rights reserved.
confrontations of racial bias; instead, they are often ignored (Czopp & Montenheit, 2003; Gulker et al., 2013; Simon & O’Brien, 2015; Woodzicka, Mallett, Hendricks, and Pruitt, 2015). This is likely because norms against sexism are relatively weak (Fiske & Stevens, 1993), and people believe that their positive regard for women (the “women are wonderful” effect, see Eagly & Mladinic, 1989) protects against any possibility that they behave in sexist ways.

We sought to determine whether a new confrontation strategy could improve people’s reactions to being confronted about their own gender bias. Importantly, as explained in greater detail below, past gender bias confrontation studies did not provide concrete evidence of participants’ gender bias, and thus allowed people to dismiss or discount the confrontation. In contrast, we used an evidence-based gender bias confrontation strategy to investigate whether presenting people with concrete evidence that they discriminated against a female lab manager applicant would prompt concern and intentions to guard against future gender bias. Each experiment also examined potential moderating and mediating variables. Altogether, this research advances current understanding about the potential and limitations of evidence-based gender confrontation for curbing gender bias.

1. Raising awareness of STEM gender bias with scientific evidence

Educating people about the scientific evidence of gender bias is one potential method of raising awareness of the bias and inspiring personal change. Unfortunately, simply reporting experimental evidence of gender bias does not appear to be sufficient to eradicate it, likely because people are often motivated to justify the existing social system (Jost & Banaji, 1994) and rationalize or minimize evidence that contradicts their existing worldview (Darley & Gross, 1983). For example, researchers (Moss-Racusin, Molenda, & Cramer, 2015) content analyzed 831 comments written by members of the public in response to three wonderful people believe that their positive regard for women (the Woodzicka, Mallett, Hendricks, and Pruitt, 2015). This is likely because (Czopp & Monteith, 2003; Gulker et al., 2013; Simon & O’Brien, 2015; Monteith, Ashburn-Nardo, Voils, & Czopp, 2002). When another person identifies one’s biases through confrontation, do similar consequences occur? Past research in which non-Black participants were confronted about responding in biased ways toward Blacks has supported the efficacy of confrontation for curbing racial bias (Czopp & Montenheit, 2003; Gulker et al., 2013) as mediated through negative self-directed affect (Czopp et al., 2006).

However, confrontation may not work as effectively for gender bias. For example, whereas participants who imagined engaging in race-biased behaviors felt negative self-directed affect, participants who imagined engaging in gender-biased behaviors felt amused (Czopp & Montenheit, 2003). In other research (Gulker et al., 2013), participants read a persuasive article that addressed the prevalence and consequences of implicit bias either toward women or Blacks, and that called on people to work toward personal changes. Participants in the gender bias condition appeared to trivialize the confrontation message and were unpersuaded, compared to participants confronted in the race bias condition. Still other research has shown that racist jokes and statements were rated as more offensive and confrontation-worthy than sexist jokes and statements, and confronters of racism were liked more than confronters of sexism (Woodzicka, Mallett, Hendricks, and Pruitt, 2015). Finally, men initially confronted about a sexist comment expressed just as much gender prejudice later as men who were not initially confronted (Simon & O’Brien, 2015; but see Mallet & Wagner, 2011).

What underlies the marked ineffectiveness of gender bias confrontation? In general, beliefs about women are positive (Eagly, Mladinic, & Otto, 1991); indeed, the so-called “women are wonderful” effect reflects the fact that people often report liking women more than men (Eagly and Mladinic, 1989). Likely because of their role as caregivers, women are generally viewed as warm and likable (Fiske, Cuddy, Glick, & Xu, 2002), and many people hold benevolent (but nonetheless patronizing and restrictive) attitudes toward women (Glick & Fiske, 1996). Furthermore, Fiske and Stevens (1993) argued that the close, communal relationships between men and women, coupled with extensive contact with women regardless of one’s own gender, gives prejudice toward women a unique character. At the societal level, these views of women can contribute to weak norms opposing sexism (Fiske & Stevens, 1993). At the individual level, people who believe they hold very positive attitudes about women may not believe they can be prone to sexism. These arguments are consistent with the finding that internal and external motivations to control one’s bias are weaker for sexism than for racism (Klonis, Plant, & Devine, 2005, Study 2; see also Cowan & Hodge, 1996; Rodin, Price, Bryson, & Sanchez, 1990).

Because people may doubt that they could be sexist or that sexism is a societal problem, the effectiveness of confrontation of gender bias may hinge on presenting people with clear, conclusive evidence of their own gender bias and its negative consequences. Presenting people with direct evidence that they exhibited bias with discriminatory outcomes may be effective at prompting self-reflective affect and self-regulatory processes. Importantly, this type of evidence-based confrontation has not been used in past gender confrontation research, which gave people greater latitude to trivialize and dismiss the confrontations.

2. Confronting people about their STEM gender bias

When people recognize that they are prone to responding in biased ways that conflict with their egalitarian personal standards, they often experience self-directed negative affect, such as guilt and disappointment with the self (e.g., Burns, Montenheit, & Parker, 2017; Montenheit, Devine, & Zuwerink, 1993; Montenheit & Voils, 1998). This particular type of affect leads to reparative action (e.g., Tagney, Miller, Flicker, and Barlow, 1996). In the context of prejudice, negative self-directed affect activates a variety of self-regulatory processes that facilitate the monitoring, control and reduction of bias in the future (e.g., Amadio, Devine, & Harmon-Jones, 2007; Montenheit, 1993; Montenheit, Ashburn-Nardo, Voils, & Czopp, 2002). We utilized Moss-Racusin et al.’s (2012) pattern of results to investigate how people respond to being confronted with evidence of their own gender bias. Specifically, we examined participants’ reactions to the news that they had provided gender biased evaluations of a female applicant for a lab manager position, relative to participants who were not confronted for gender-biased evaluations.

All four experiments examined whether an evidence-based confrontation about gender bias increased negative self-directed affect and, in turn, concern about gender bias and intentions to monitor one’s
future behavior for possible gender bias. We also measured defensive reactions (i.e., participants perceiving their gender bias as reasonable and the researchers as oversensitive). Given past findings that men reacted more defensively than women to scientific evidence claiming gender bias (Handley et al., 2015; Moss-Racusin et al., 2015), we were also interested in examining possible gender effects.

To establish the unique effects of confrontation with evidence, Experiment 1 compared the impact of our novel evidence-based confrontation of gender bias in Physics to the impact of a traditional confrontation of gender bias without evidence, relative to a no confrontation condition. Experiment 2 again tested the effectiveness of an evidence-based confrontation and also included a bias characterization manipulation by explaining that bias against women in Physics could arise either from associating women with lack of intelligence, or from a tendency not to associate women with science. Regarding women as unintelligent is an old-fashioned form of sexism that many people now regard as personally unacceptable (Swim, Aikin, Hall, & Hunter, 1995). Therefore, exhibiting such bias should conflict with participants’ personal standards for responding, so they may experience increased negative self-directed affect and, in turn, increased concern and intentions for monitoring gender bias. In contrast, we reasoned that a bias rooted in failure to associate women with science would be more easily dismissed or justified (e.g., “After all, scientists are disproportionately male”). In sum, this manipulation tested whether the effectiveness of evidence-based confrontation depended on how the gender bias was characterized.

Experiment 3 also concerned gender bias in Physics, but varied whether participants were confronted about unfavorable evaluations based on the applicant’s race or gender. Given past research revealed weaker effects of gender than race confrontation (e.g., Gulker et al., 2013; Woodzicka, Mallett, Hendricks, & Pruitt, 2015), we sought to determine whether our evidence-based confrontation likewise yielded stronger effects for racial bias.

Experiment 4 explored whether evidence-based confrontation effects would generalize beyond a STEM context by using History as the confrontation context. In addition, we manipulated the salience of injunctive social norms opposing bias toward women in History. Prejudice researchers have long emphasized that norms opposing bias play an important role in reducing people’s expression of prejudice (Allport, 1954; Crandall & Eshleman, 2003). Understanding whether the effectiveness of confrontation depends on norm salience will be informative for understanding the contexts in which evidence-based confrontations may or may not be effective. Finally, Experiment 4 also examined whether the effects of confrontation depended on participants’ degree of internal and external motivation to respond without prejudice (IMS and EMS, respectively; Plant & Devine, 1998). Prior confrontation research has not (to our knowledge) investigated whether these motivations affect reactions to confrontation. However, past research has shown that people are more receptive to pressures to reduce bias as IMS increases and as EMS decreases (Plant & Devine, 2001).

In sum, the primary hypotheses tested across these experiments were that an evidence-based confrontation would increase concern about gender bias including intentions to self-regulate bias in the future, and that this effect would be mediated by negative self-directed affect (Czopp et al., 2006; Monteith et al., 2002). By also examining a broad array of possible moderating factors, we further sought to better understand this strategy’s limits or lack thereof. These factors include characterization of the bias [science vs. intelligence], bias type [gender vs. race], domain [Physics vs. History], norms [oppose bias vs. control], and individual differences [IMS and EMS]. The breadth of the research is intended to provide nuanced, practical answers to the question of how people react to evidence-based gender confrontations.

4. Experiment 1

For all experiments reported herein, all measures, manipulations, and exclusions are disclosed, and also the method for determining the final sample size.

4.1. Method

4.1.1. Power analysis

Confrontation yielded medium to large effect sizes in prior research (e.g., $d = 0.50$, Czopp et al., 2006). Past research has not found interactions between participant gender and confrontation (Czopp & Monteith, 2003; Gulker et al., 2013) nor that participant gender affects judgments gender discrimination (Rudman & Phelan, 2008). Nonetheless, we sought adequate power to detect medium-sized interaction effects ($d = 0.50$) if they were present. A power analysis conducted with G*Power (v. 3.1) to estimate the required sample size to have 0.80 power (Faul, Erdfelder, Buchner, and Lang, 2009) suggested a sample size of 158.

4.1.2. Participants

Participants were 177 Amazon Mechanical Turk (mTurk) workers from the United States who were compensated $1.00. Sixteen participants’ data were excluded because they incorrectly recalled the gender of the applicant they evaluated. Data from one participant who reported non-binary gender were also excluded given our interest in examining male versus female gender effects. One participant denied permission to use his data after debriefing concerning the study’s use of deception; this person's data were deleted. Thus, the final sample included 159 participants (55% female, $M_{age} = 37.03$, $SD_{age} = 11.73$, 77% White).

4.1.3. Design

A 2 (participant gender: male vs. female) × 3 (confrontation: not confronted vs. confronted without evidence vs. confronted with evidence) between-participants design was used.

4.1.4. Procedure

The online experiment was presented through Qualtrics. The instructions described the experiment as investigating evaluations of applicants in employment decision contexts. After providing consent, participants were asked to imagine that they were a professor making a selection for a laboratory manager for their own Physics lab. They were given detailed information about the potential impact of this choice on professional success and outcomes (e.g., publications, grants, salary, tenure and promotion), which was then summed up with the statement that “Ultimately, the success of your laboratory and your research depends on who you hire for the position.” Participants then reviewed a detailed description of the laboratory manager position including typical responsibilities and tasks. Next, participants were provided with an example of an application for an “outstanding applicant” so they could “see what a stellar candidate looked like.” Afterwards, participants viewed the application materials of the candidate they were to evaluate, which varied according to the experimental condition to which participants were randomly assigned.

4.1.4.1. Application materials. The application included a resume, excerpts from a personal statement, academic statement, and excerpts from a letter of recommendation (adapted from Moss-Racusin et al., 2012). Participants either evaluated a male (not confronted condition) or female (confronted with and without evidence conditions) applicant who had graduated from college with a Bachelor of Science degree. The application materials in each condition were identical and were designed to be ambiguous in strength (GPA of 3.2, ACT of 24, somewhat weak explanation of favorite college course) to enhance the believability of the feedback that we later provided to participants about their applicant evaluations.

4.1.4.2. Applicant evaluation. Using items adapted from Moss-Racusin...
et al. (2012), participants recorded 11 ratings concerning the applicant’s competence, likeability, hireability, and compensation. The applicants were evaluated similarly when specified as male (no confrontation condition) and female (confronted with evidence and confronted without evidence conditions). ps > 0.19. These ratings were substantively unimportant for the present purposes beyond their supposed use for providing evaluation feedback to participants.

4.1.4.3. Confrontation. Next, participants in the evidence-based confrontation condition were informed that everyone in the experiment had seen the same application, the application was very strong, and the applicant should have been hired. In addition, they read that the applicant was specified as male for some participants and as female for others. The instructions explained that the researchers were manipulating applicant gender to see whether gender bias operated to influence evaluations, noting that when the same application was used in past studies and attributed to a male, the applicant was invariably evaluated favorably, resulting in a decision to hire him. The instructions concluded that if a female applicant was not hired, some type of gender bias must be operating. Next, participants in this evidence-based confrontation condition saw that their evaluation results were supposedly being calculated, followed by a brief report (see Supplemental Materials). The report showed participants that their evaluations of the applicant were more negative than ratings obtained from past large samples using the same application, but specifying the applicant as male. The report concluded that the applicant would have been rejected from the position based on participants’ evaluations, and that “this indicates that you showed a gender bias against women.”

Participants in the confrontation without evidence condition were only informed that the study was exploring how demographic information, such as the applicant’s gender, might influence evaluations, and that failure to hire the female applicant indicated gender bias. Next, these participants saw that their evaluation results were supposedly being calculated. This was followed by a report indicating that the applicant would not have been hired for the position based on the ratings participants’ provided, and that “this indicates that you showed a gender bias against women.”

Finally, participants in the not confronted condition did not receive any information about gender or potential gender bias. Although these participants also saw that their evaluation results were calculated and received a brief report, the report simply indicated that the applicant (who was male in this condition) would have been rejected from the position based on their ratings.

4.1.4.4. Dependent variables. After viewing their feedback, participants completed a 25-item affect measure by rating the extent to which each affect item applied to how they felt at that moment on a scale from 1 (does not apply at all) to 7 (applies very much). Consistent with prior work involving this measure (Czopp et al., 2006; Czopp & Monteith, 2003; Monteith et al., 1993) and results of a principal components analysis that we performed, we created six affect indices: negative self-directed affect, or negsef (guilty, self-critical, angry with myself; disgusted; \( \alpha = 0.91 \)); discomfort (frustrated, anxious, embarrassed, surprised, tense, uncomfortable, and uneasy; \( \alpha = 0.88 \)); depressed (low, sad, depressed, helpless, \( \alpha = 0.86 \)); negative other-directed affect or negother (annoyed with others, angry with others, irritated at others, and fed up; \( \alpha = 0.91 \)); positive (happy, optimistic, friendly, and good, \( \alpha = 0.91 \)); and amusement (amused and entertained, \( r = 0.77, p < 0.001 \)).

Next, participants rated 10 items presented in random order concerning their future intentions to monitor and control their possible gender biases, using a response scale ranging from 1 (not at all true) to 7 (completely true). These items were constructed for the purpose of this research based on findings from past research related to the self-regulation of prejudice (e.g., Monteith et al., 2002; Monteith, Mark, & Ashburn-Nardo, 2010). Sample items include: “I am concerned about exhibiting gender biases in the future,” “After what I learned today, I will be more on guard for gender biased behavior,” and “I will now be more careful to monitor myself to make sure that I am not being gender biased.” These items were averaged to form a reliable composite, \( \alpha = 0.96 \).

Two types of items were also included to assess possible defensive reactions to the feedback. First, seven items involved claims that women are a poor fit for science careers (e.g., “Women are ultimately more interested in family and kids than science careers,” and “My ratings of the student reflect the fact that women simply do not have the talent necessary to be great scientists,” \( \alpha = 0.95 \)). Such claims can be a way of defending gender bias, and the items were constructed based on past research examining people’s reactions to confrontation (Czopp & Monteith, 2003) and to being presented with evidence of gender bias (Moss-Racusin et al., 2015). Second, three items tapped into the extent to which participants evaluated the researchers negatively (e.g., “The researchers for this study are complainers,” and “The researchers for this study are over-sensitive,” \( \alpha = 0.88 \)). These items were drawn from past research on perceptions of people who claim discrimination (e.g., Kaiser & Miller, 2001). All defensive reaction items were rated on a scale from 1 (not at all true) to 7 (completely true).

Participants also responded to four items assessing the extent to which they believed the feedback they had received (e.g., “The feedback about my ratings of the applicant was believable,” \( \alpha = 0.79 \)), using a 1 (not at all true) to 7 (completely true) response scale.

Finally, participants completed a brief demographic questionnaire and reported the gender of the applicant they had reviewed. Participants were then fully debriefed and asked for permission to use their data on a post-session consent form.

4.2. Results and discussion

4.2.1. Analytic procedure and treatment of participant gender effects

Each dependent variable was initially submitted to a 2 (participant gender: men vs. women) × 3 (confrontation: not confronted vs. confronted with evidence vs. confronted without evidence) ANOVA. However, in this and the subsequent three experiments, only a few interactions involving gender emerged, and they were neither consistent across studies nor theoretically interesting. Thus, for all reported studies, the main effect for participant gender was included in analyses but interactions involving gender were not, and significant gender main effects are reported when obtained.

4.2.2. Believability of the confrontation

A significant main effect for confrontation was obtained when predicting how much participants believed their evaluation feedback, \( F (2, 156) = 25.15, p < 0.001, \eta^2_p = 0.24 \). Participants who were not confronted believed their feedback more (\( M = 5.50, SD = 1.25 \)) than participants in either the confronted with evidence condition (\( M = 5.38, SD = 1.63 \)), \( t(156) = 5.32, p < 0.001, d = 1.18 \), or in the confronted without evidence condition (\( M = 3.88, SD = 1.76 \)), \( t(156) = 8.49, p < 0.001, d = 1.39 \). The two confrontation conditions did not differ from one another, \( t(156) = 1.36, p = 0.175, d = 0.23 \). Because participants in the confrontation conditions had been informed that they exhibited gender bias, we were unsurprised to find that they were less likely to believe their feedback than participants in the not confronted condition. However, the believability ratings even in the confronted conditions seemed sufficiently high to suggest that participants did not outright dismiss the feedback as unbelievable.

4.2.3. Affective reactions to feedback

Analysis of the negsef index revealed the anticipated main effect for confrontation condition, \( F(2, 156) = 3.43, p = 0.035, \eta^2_p = 0.04 \). As shown in Table 1, results supported the critical prediction that confronting participants about gender bias would activate greater negative self-directed affect when the confrontation was accompanied by
The main effect of confrontation condition was significant when predicting participants’ concern about their gender biases and monitoring them in the future, $F(2, 156) = 5.44, p = 0.005, \eta^2_p = 0.07$. As anticipated, participants confronted with evidence reported significantly greater concern about gender bias ($M = 3.62, SD = 2.03$) than participants who were not confronted ($M = 2.39, SD = 1.41$), $t(156) = 3.08, p = 0.002, d = 0.70$, and participants who were confronted without evidence ($M = 2.77, SD = 1.81$), $t(156) = 2.57, p = 0.011, d = 0.46$. In contrast, participants who were confronted about gender bias but without any accompanying evidence of their bias reported no greater future concern than participants in the not confronted condition, $t(156) = 0.82, p = 0.414, d = 0.23$.

### 4.2.5. Mediation: effect of confrontation on future concern through negself

Experiencing negative self-directed affect in relation to one’s biases is important for instigating future self-regulatory efforts (e.g., Burns et al., 2017; Monteith et al., 2002). The crucial test for this process in the present research involved examining whether the effect of confrontation on concern about expressing and regulating gender bias in the future was mediated by negself. This hypothesis was tested using Hayes’ (2013) PROCESS macro (Model 4) with 5000 bootstrapped samples. We specified confrontation condition as a multicategorical independent variable, which allowed us to examine mediation for both the not confronted vs. confrontation with evidence comparison, and for the not confronted vs. confrontation without evidence comparison. As shown in Fig. 1, results were consistent with predictions. The effect of the confrontation without evidence on concerns about expressing and monitoring for future bias was not mediated by negself, 95% CI [−0.40, 0.33]. However, the effect of confrontation with evidence on concerns about future bias was mediated by negself, 95% CI [0.02, 0.98].

Because confrontation also influenced participants’ experience of discomfort, negother, and positive affect in this study, we also examined the hypothesized mediating effect of negself beyond these other affective reactions. Specifically, the effect of confrontation on future concern was examined with negself, discomfort, negother, and positive affect, $F(2, 156) = 4.39, p = 0.014, \eta^2_p = 0.05$, and positive affect, $F(2, 156) = 4.58, p = 0.012, \eta^2_p = 0.05$. As shown in Table 1, the pattern observed across these indexes indicated that both types of confrontations activated greater discomfort and negother and less positive affect, relative to the not confronted condition, $t(156) > 2.15, ps < 0.034, ds > 0.51$. The only exception was that participants who were not confronted reported marginally less discomfort than those who were confronted without evidence, $t(156) = 1.87, p = 0.063, d = 0.45$. In contrast, the confrontation with evidence and without evidence conditions consistently did not differ from each other, $t(156) < 0.45, ps > 0.647, ds < 0.09$.

In sum, the affect results indicated that both confrontations elicited feelings of discomfort and negative affect directed at others, along with reduced positive affect. However, confrontation elicited heightened negative self-directed affect only when it was accompanied by concrete evidence of bias. This type of affect should be particularly influential in triggering concern about future prejudice and associated self-regulatory efforts.

### 4.2.6. Defensive reactions to feedback

Defensiveness was assessed by measuring perceptions that 1) women are a poor fit for science, and 2) evaluating the researchers as oversensitive and complaining. There was a significant main effect of confrontation condition on agreement that women are a poor fit for science, $F(2, 156) = 3.30, p = 0.039, \eta^2_p = 0.04$. Participants

![Fig. 1. The effect of confrontation (without evidence vs. not confronted; with evidence vs. not confronted) on future concern about expressing and monitoring bias mediated by negative self-directed affect, Experiment 1. Path values are unstandardized coefficients with standard errors presented in parentheses. The direct effect of the independent variable on the dependent variable (c) is reported above the line, and the indirect effect (ab) is reported below the line. *$p < 0.05$, **$p < 0.01$.](image-url)
confronted with evidence were significantly more likely to agree that women were a poor fit for science careers than participants who were confronted without evidence, \( t(156) = 2.56, p = 0.012, d = 0.47 \). However, this finding is difficult to interpret given that neither of the confrontation conditions differed significantly from the not confronted condition \((M = 1.63, SD = 1.20)\), \( ts(156) < 1.30, ps > 0.194, ds < 0.28 \).

We also obtained a significant main effect for confrontation when predicting ratings of the extent to which the researchers were oversensitive and complaining, \( R(2, 156) = 6.11, p = 0.003, \eta^2_p = 0.07 \). Consistent with past findings concerning reactions to confronters (e.g., Gulker et al., 2013; Kaiser & Miller, 2001), participants in the confrontation with evidence \((M = 3.19, SD = 1.86)\) and without evidence \((M = 3.49, SD = 1.99)\) conditions were more likely to view the researchers as oversensitive and complaining than participants in the not confronted condition \((M = 2.40, SD = 1.81)\), \( ts(156) = 2.48 \) and \( 3.44, ps = 0.014 \) and \( 0.001, ds = 0.43 \) and \( 0.60 \), respectively. The two confrontation conditions did not differ from one another, \( t(156) = 0.86, p = 0.387, d = 0.16 \).

4.3. Summary

Study 1 demonstrated that confronting participants with evidence of gender bias had affective consequences including negative self-directed affect, which, in turn prompted concern about expressing and intentions to regulate gender bias in the future. In contrast, participants simply confronted about gender bias but without evidence of their bias reported significantly less negative self-directed affect and future concern relative to the confrontation with evidence condition. Although participants confronted without evidence apparently did not like being confronted, as evidenced by increased discomfort and negative affect directed at others and reduced positive affect, these affective reactions did not prompt concern about future gender bias.

5. Experiment 2

Given that the evidence-based confrontation was effective in Experiment 1, we proceeded to examine whether its effectiveness would be moderated by the way in which gender bias is characterized during confrontation. Seeing women as less intelligent than men constitutes an old-fashioned form of sexism (Swim et al., 1995), and people may be especially responsive to an evidence-based confrontation highlighting that their gender bias is rooted in this type of bias. In contrast, a confrontation highlighting that gender bias is rooted in a tendency not to associate women with certain careers, such as science, may be viewed as more defensible and less objectionable. In sum, we wished to test whether the effectiveness of an evidence-based confrontation of gender bias depended on whether the gender bias was characterized as involving more old-fashioned sexism concerning the intelligence of women or as involving associations between women and science careers.

5.1. Method

5.1.1. Power analysis

Confrontation yielded medium to large effect sizes in Experiment 1 and prior research (e.g., \( ds \geq 0.50 \), Czopp et al., 2006). Although obtaining sufficient power to detect possible two-way interactions was important to our research questions, we did not expect a significant three-way interaction (i.e., between confrontation condition, bias characterization, and participant gender (see Gulker et al., 2013; Rudman & Phelan, 2008). Thus, we conducted a power analysis in G*Power (v. 3.1) to estimate the required sample size to have 0.80 power for detecting medium \((d = 0.50)\) two-way interaction effects. The analysis suggested a sample size of 128, which we surpassed.

5.1.2. Participants

Participants were 188 Introductory Psychology students at a large Midwestern University who received course credit. Five participants’ data were not recorded due to computer malfunction, and nine participants’ data were excluded because they failed a manipulation check (i.e., incorrectly recalled the gender of the applicant they evaluated or how gender bias was characterized in the confrontation). The final sample included 174 participants (42% female, \( M_{\text{age}} = 18.94 \)) who were primarily White (80%).

5.1.3. Design

A 2 (confrontation: not confronted vs. confronted) × 2 (bias characterization: gender-science vs. gender-intelligence) between-participants design was used. In this experiment and Experiments 3 and 4, all confronted participants were presented with evidence of gender bias.

5.1.4. Procedure

Up to six participants completed the experiment at a time at individual computer stations. The procedure for Experiment 2 was identical to Experiment 1 until the point at which the confrontation occurred, with one exception. As in Experiment 1, confronted participants evaluated a female applicant and not confronted participants evaluated a male applicant; however, in this experiment and also in Experiments 3 and 4, the male applicant had stronger credentials (i.e., higher ACT and GPA scores, stronger statement about favorite college course) than the female applicant. Drawing from past research using a similar method (Monteith, 1993, Study 1), we varied these aspects of the applicant’s materials so that ratings of the male applicant would be positive and thus consistent with feedback that he would have been hired based on evaluations. In contrast, ratings of the female applicant were expected to be less positive and thus consistent with feedback that she would not have been hired. Indeed, in this experiment and Experiments 3 and 4, the male applicant was evaluated significantly more positively than the female applicant \((p < 0.05)\).

5.1.4.1. Confrontation. After participants evaluated their applicant, the experimenter explained that they would learn whether the applicant would have been hired based on their ratings. All participants were told that everyone in the experiment had seen the same application, that the application was very strong, and thus, that the applicant should have been hired. They were further told that although the application was the same for all participants, the applicant was specified as male for some participants and as female for others. The experimenter explained that we were manipulating gender to see whether gender bias operated to influence applicant evaluations, noting that when the same application was used in past studies and attributed to a male, the applicant was invariably hired. So, the experimenter explained, “If we find that the female applicant is not hired, some type of gender bias must be operating.” In contrast, participants were told that hiring the male applicant would not indicate bias because the application was, after all, strong. In sum, participants in both confrontation conditions learned that negative evaluations of the female applicant would be indicative of gender bias; however, only participants in the confrontation condition had actually evaluated a female candidate.

We manipulated how gender bias was characterized when delivering the above explanations according to how the experimenter explained the potential for gender bias. In the gender-science condition, participants were told that any gender bias observed “must be because

\footnote{In the undergraduate samples used in Experiments 2-4, an a priori decision was made to exclude participants if they reported 1) being born in another country, 2) learning English as a second language, and 3) speaking English for less than five years. This exclusion criterion was set because the procedure relied so heavily on reading and responding to materials on written English. Consequently, 15 participants were excluded in Experiment 2 and two participants were excluded in Experiment 4. These participants are not included in the reports of sample sizes found in the main text.}
the position is for a lab manager in physics, and oftentimes women are not associated with science in people’s minds.” In the gender-intelligence condition, the explanation for bias was “because lab manager positions require high levels of intelligence, and typically women may be perceived as less intelligent than males.”

Next, participants’ computer appeared to calculate their results, and then displayed a brief report supposedly comparing participants’ own ratings of the applicant on various dimensions to ratings of the male that had been collected across large past samples (see Supplemental Materials). In the not-confronted condition, the results showed that participants’ evaluations of the male applicant were similar to prior research and would have resulted in the male applicant being hired. In the confrontation condition, the report showed that the participant’s evaluations (of the female applicant) were lower than in prior research involving the male applicant and would not have resulted in hiring the applicant.

5.1.4.2. Dependent variables. The dependent variables were the same as in Experiment 1. After viewing their feedback, participants completed the same 25-item affect measure as in Experiment 1 that yielded the following indexes: negself (α = 0.87), discomfort (α = 0.83), depressed (α = 0.83), negther (α = 0.87); positive (α = 0.87); and amusement (r = 0.64, p < 0.001). Participants then completed the 10-item measure tapping into concern about gender bias and intentions to monitor it in the future (α = 0.94), seven items concerning the poor fit of women in science careers (α = 0.91), three items related to perceptions of the researchers as complainers and oversensitive (α = 0.84), and four items assessing how much participants believed the feedback they received (α = 0.66).

Finally, participants completed a brief demographic questionnaire, manipulation checks (i.e., reported gender of applicant and how bias was characterized). Participants were then fully debriefed and asked for permission to use their data (all gave permission).

5.2. Results and discussion

5.2.1. Analytic procedure

Unless otherwise noted, each dependent variable was initially submitted to a 2 (confronted vs. not-confronted) x 2 (bias characterization: gender-science vs. gender-intelligence) ANOVA, with gender also entered as a main effect.

5.2.2. Believability of the confrontation

Confronted participants believed the feedback less (M = 4.85, SD = 1.05) than participants who were not confronted (M = 5.73, SD = 0.93), F(1, 169) = 34.63, p < 0.001, d = 0.89. Because agreeing with the believability items also implied accepting responsibility for sexist bias, this difference is unsurprising. However, the believability ratings even in the confronted condition were above the midpoint, suggesting sufficient believability. A main effect of participant gender indicated that women were more likely to believe their feedback (M = 5.38, SD = 0.98) than were men (M = 5.05, SD = 1.16), F(1, 169) = 6.28, p = 0.013, d = 0.31, regardless of confrontation condition.

5.2.3. Affective reactions to feedback

Results from analyses of the affect indexes are summarized in Table 2. As expected, confronted participants reported significantly greater negself than participants who were not confronted, F(1, 169) = 22.15, p < 0.001. Confronted participants also reported greater discomfort, F(1, 169) = 36.45, p < 0.001, depressed affect, F(1, 169) = 12.48, p = 0.001, and negther, F(1, 169) = 13.61, p < 0.001, and also less positive affect, F(1, 169) = 26.68, p < 0.001. Only amusement had no significant effects (ps > 0.19).

These main effects for confrontation were qualified by an interaction with bias characterization in the case of depressed affect only, F(1, 169) = 3.75, p = 0.054, αp = 0.02. Participants in the gender-science condition reported significantly greater depressed affect following a confrontation (M = 2.50, SD = 1.27) than participants who were not confronted (M = 1.58, SD = 0.85), t (81) = 3.57, p = 0.001, d = 0.85. Surprisingly, this effect was weaker and not significant in the gender-intelligence condition (confronted M = 2.06, SD = 0.96; not confronted M = 1.78, SD = 1.00), t(89) = 1.33, p = 0.187, d = 0.29.

In sum, the evidence-based confrontation clearly aroused a range of feelings, suggesting that this form of confrontation is powerful enough to elicit affective reactions, and regardless of how the bias was characterized.

5.2.4. Concern about future prejudice

Women reported greater concern about expressing and regulating gender bias in the future (M = 4.32, SD = 1.34) than men (M = 3.66, SD = 1.49), F(1, 169) = 7.95, p = 0.005, d = 0.47. In addition, confronted participants reported marginally greater concern about future prejudice (M = 4.12, SD = 1.43) compared to participants who were not confronted (M = 3.65, SD = 1.35), F(1, 169) = 3.63, p = 0.058, d = 0.35. Notably, even participants who were not confronted were moderately concerned about expressing and regulating their bias, with ratings around the midpoint of the scale. This likely was because participants who were not confronted had heard the experimenter explain that gender bias could subtly influence ratings of a female applicant, which may have heightened their concern even though they did not evaluate a female applicant. We examine this possibility directly in Experiment 4.

This effect of confrontation on future concern was not moderated by bias characterization. Participants reported similar concern about expressing and regulating bias in the future whether the confrontation was couched in terms of gender-STEM or gender-intelligence bias, F < 1.0.

5.2.5. Mediation: effect of confrontation on future concern through negself

As in Experiment 1, we tested whether the effect of confrontation on concern about expressing and regulating gender bias in the future was mediated by negself. Using Hayes’ (2013) PROCESS macro (Model 4) with 5000 bootstrapped samples, we found evidence of a significant indirect effect, 95% CI [0.30, 0.79] both in a simple mediation model (see Fig. 2) and when other affect indices (i.e., discomfort, depressed, negther, and positive) were included in a multiple simultaneous mediation model, 95% CI [0.25, 0.93]. Thus, being confronted about gender bias prompted future self-regulatory efforts to the extent that it induced negative self-directed affect. In addition, the simultaneous mediation analysis also revealed a unique role for negther, and in a direction opposite to that of negself, CI [−0.32, 0.02]. That is, increased negative other-directed affect through confrontation was associated with less concern about regulating gender bias.

5.2.6. Defensive reactions to feedback

Agreement that women are a poor fit for science was marginally greater among male (M = 1.74, SD = 0.98) than female (M = 1.50, SD = 0.70) participants, F(1, 169) = 3.43, p = 0.066, d = 0.28. Men (M = 2.13, SD = 1.12) were also more likely than women (M = 1.65, SD = 0.91) to report that the researchers were complainers and overly sensitive, F(1, 169) = 9.45, p = 0.002, d = 0.47. However, no effects involving confrontation condition were significant (ps > 0.13).

Because all participants (in the confronted and not-confronted conditions alike) knew that the research was testing for gender biases, defensiveness may have arisen among males simply because this was the
Experiment 2 revealed that evidence-based confrontations of gender bias increased feelings of negative self-directed affect and future concern about exhibiting and regulating gender bias, regardless of how the gender bias was characterized. When people were provided with clear evidence of their own biased responses, they were equally likely to accept the confrontation when bias was characterized in terms of old-fashioned, hostile sexism (i.e., women are less intelligent than men) or more subtle gender-stereotypes (i.e., women ≠ STEM). This suggests that confronting people with evidence of their bias can motivate future self-regulation for both old-fashioned forms of sexism that are contrary to many people's personal standards and for more modern forms of bias.

Although the confrontation did not elicit greater defensive reactions from confronted than not confronted participants, Experiment 2 revealed that male participants generally showed greater defensiveness and also less concern about gender bias than female participants. These latter results are consistent with Handley et al.'s (2015) finding that males are more threatened than women by research concerning gender

### Table 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>Experiment 2</th>
<th>Exp. 3</th>
<th>Exp. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not confronted</td>
<td>Confronted</td>
<td>d</td>
</tr>
<tr>
<td>Negself</td>
<td>1.86 (0.96)</td>
<td>2.84 (1.52)</td>
<td>0.77*</td>
</tr>
<tr>
<td>Discomfort</td>
<td>1.90 (0.79)</td>
<td>2.82 (1.07)</td>
<td>0.97*</td>
</tr>
<tr>
<td>Depressed</td>
<td>1.69 (0.93)</td>
<td>2.27 (1.39)</td>
<td>0.56*</td>
</tr>
<tr>
<td>Negoher</td>
<td>1.75 (0.90)</td>
<td>2.45 (1.39)</td>
<td>0.60*</td>
</tr>
<tr>
<td>Positive</td>
<td>4.81 (1.27)</td>
<td>3.72 (1.39)</td>
<td>0.82*</td>
</tr>
<tr>
<td>Amused</td>
<td>3.25 (1.38)</td>
<td>3.00 (1.49)</td>
<td>0.17</td>
</tr>
</tbody>
</table>

* Denotes means differ significantly at $p < 0.05$ at least. Standard deviations are presented in parentheses.

Fig. 2. The effect of confrontation on future concerns about expressing and monitoring bias mediated by negative self-directed affect, Experiments 2–4. Path values are unstandardized coefficients with standard errors presented in parentheses. The direct effect of the independent variable on the dependent variable ($c'$) is reported above the line, and the indirect effect (ab) is reported below the line. *$p < 0.001$.
bias.

6. Experiment 3

In Experiment 3 we compared the effectiveness of an evidence-based confrontation of gender bias to a comparable evidence-based confrontation of race bias. People view sexism as less offensive than racism (Cowan & Hodge, 1996; Rodin et al., 1990; Woodzicka et al., 2015), and prior research suggests that confrontations of racism are more effective than confrontations of sexism (Czopp & Monteith, 2003; Gulker et al., 2013). Thus, an evidence-based race bias confrontation may have a stronger effect on negative self-directed affect and future concern about bias than a gender bias confrontation. On the other hand, direct evidence of one's gender bias in a hiring context that has tangible negative consequences for women may “have what it takes” to prompt negative self-directed affect and future concern at levels that parallel racism confrontation.

6.1. Method

6.1.1. Power analysis

The G*Power (v. 3.1) power analysis from Experiment 2 applies to Experiment 3, estimating that a sample size of at least 179 would be ideal for 0.80 power for detecting medium ($d = 0.50$) effects in a two-way ANOVA.

6.1.2. Participants

Participants were 218 Introductory Psychology students at a large Midwestern university who received course credit for participating. Data for nineteen participants were excluded because they failed their manipulation check (i.e., incorrectly recalled the gender of the applicant in the gender bias condition or the race of the applicant in the race bias condition). The final sample included 199 participants (60% female, $M_{age} = 18.93, 98\%$ White).

6.1.3. Design

A 2 [(confrontation: confronted vs. not confronted) × 2 (bias type: race vs. gender)] between-participants design was used.

6.1.4. Procedure

The cover story, procedure, and materials were very similar to Experiment 2. Participants randomly assigned to the gender bias condition either evaluated a White male (not confronted) or a White female (confronted) applicant. Participants in the race bias condition either evaluated a White male (not confronted) or a Black male (confronted) applicant. As in Experiment 2, the application materials were varied so that participants who were later confronted evaluated a somewhat weaker application than participants who were not confronted. Participants evaluated their applicant with the same items as in Experiments 1 and 2.

6.2. Results and discussion

6.2.1. Analytic procedure

Unless otherwise noted, dependent variables were submitted to a 2 [(confrontation: confront vs. not confront) × 2 (bias type: race vs. gender bias)] ANOVA with gender entered as a main effect.

6.2.2. Believability of the confrontation

Confronted participants believed their feedback less than not confronted participants, $F(1, 194) = 52.3, p = 0.023, d = 0.27$. These main effects were qualified by a significant confrontation X bias type interaction, $F(1, 194) = 4.03, p = 0.046, \eta^2_p = 0.02$. Participants believed the feedback more if they were confronted about gender bias ($M = 4.86, SD = 0.94$) than race bias ($M = 4.32, SD = 1.07$), $t(99) = 2.74, p = 0.007, d = 0.54$. However, bias type did not affect belief ratings among participant who were not confronted (gender bias $M = 5.75, SD = 0.68$; race bias $M = 5.72, SD = 0.87$), $t(96) = 0.17, p = 0.869, d = 0.04$. The difference in the confrontation condition may reflect the normative uneasiness of being biased toward Blacks (e.g., Crandall, Eshleman, & O’Brien, 2002; Crandall et al., 2002) along with White’s typically strong non-racist self-images (e.g., Dovidio & Gaertner, 2004; Monteith & Walters, 1998). That is, participants may have felt especially reticent to accept the idea that they had provided racially biased evaluations. At any rate, the believability ratings were above the midpoint in the confronted conditions, suggesting that our procedure was generally believable.

6.2.3. Affective reactions to feedback

Analyses of the affect indexes revealed several significant main effects that were not related to our main research questions: Women reported greater negotier ($M = 2.27, SD = 1.50$) than men ($M = 1.98, SD = 0.75$), $F(1, 194) = 3.79, p = 0.053, d = 0.21$. Women also reported feeling less amused ($M = 2.51, SD = 1.40$) than men ($M = 3.18, SD = 1.45$), $F(1, 191) = 10.79, p = 0.001, d = 0.47$. In addition, participants reported less amusement in the race bias ($M = 2.32, SD = 1.29$) than the gender bias ($M = 3.22, SD = 1.47$) condition, $F(1, 194) = 21.62, p < 0.001, d = 0.65$.

More important for the main questions driving this research were significant effects of confrontation. As shown in Table 2, relative to participants who were not confronted, confronted participants experienced greater negotier, $F(1, 194) = 54.75, p < 0.001, d = 0.64$, and gender, $F(1, 194) = 59.06, p < 0.001$, depressed affect, $F(1, 194) = 36.31, p < 0.001$, and negotier, $F(1, 194) = 14.64, p < 0.001$, and less positive affect, $F(1, 194) = 38.51, p < 0.001$, and amusement, $F(1, 194) = 5.58, p = 0.019$. 
Were these main effects moderated by bias type, such that the racial bias confrontation elicited stronger affective reactions than the gender bias confrontation, as observed in previous research? We did not find a significant interaction between confrontation and bias type for the affect index of greatest interest, negself, $F(1, 194) = 2.72, p = 0.101$, but the interaction did reach significance for two other affect indexes. First, it was significant for the discomfort index, $F(1, 194) = 6.51, p = 0.011$, $\eta^2_p = 0.03$. Participants in the gender bias condition reported significantly greater discomfort if they were confronted ($M = 2.95, SD = 1.22$) than if they were not confronted ($M = 2.14, SD = 0.91$), $t(99) = 3.74, p < 0.001, d = 0.75$. However, the difference between the confronted ($M = 3.22, SD = 1.42$) and the not confronted ($M = 1.65, SD = 0.56$) conditions was even more pronounced in the case of racial bias, $t(96) = 7.15, p < 0.001, d = 1.45$. Second, confrontation and bias type interacted when predicting depressed affect, $F(1, 194) = 5.39, p = 0.021$, $\eta^2_p = 0.03$. As with discomfort, the not confronted ($M = 1.55, SD = 0.71$) and confronted ($M = 2.11, SD = 1.14$) conditions differed significantly when participants received gender bias feedback, $t(99) = 2.99, p = 0.04, d = 0.59$. However, this difference was more pronounced in the racial bias condition (not confronted $M = 1.32, SD = 0.50$; confronted $M = 2.54, SD = 1.49$), $t(96) = 5.36, p < 0.001, d = 1.10$.

In sum, in contrast to prior results obtained in response to non-evidence-based confrontation, an evidence-based confrontation elicited comparable levels of negative self-directed affect when gender and racially biased responses were confronted. Notably, the magnitude of this effect was large, as were the main effects when predicting discomfort, depressed affect, and positive affect. In addition, confrontation had a stronger effect on discomfort and depressed affect in the racism than the gender condition.

6.2.4. Concern about future prejudice

Replicating Experiment 2, women reported greater concern about expressing and monitoring possible bias in the future ($M = 4.25, SD = 1.22$) than did men ($M = 3.52, SD = 1.47$), $F(1, 194) = 14.78, p < 0.001, d = 0.54$. In addition, participants reported greater concern about future bias involving gender ($M = 4.20, SD = 1.21$) than race ($M = 3.72, SD = 1.48$), $F(1, 194) = 7.43, p = 0.007, d = 0.35$. This main effect may indicate a tendency for people to regard themselves as non-racist (Dovidio & Gaertner, 2004), which could lead to the conclusion that one need not be strongly vigilant for cues of racial bias in the future.

Most importantly, participants who were confronted about bias ($M = 4.16, SD = 1.50$) reported greater concern about expressing and regulating bias in the future than participants who were not confronted ($M = 3.76, SD = 1.18$), $F(1, 194) = 4.17, p = 0.043, d = 0.30$. Confrontation and bias type did not interact, $F < 1, p = 0.695, \eta^2_p = 0.001$, indicating that concern was heightened in the gender bias condition to the same degree as in the racial bias condition.

6.2.5. Mediation: effect of confrontation on future concern through negself

Using Hayes (2013) PROCESS macro (Model 4) with 5000 bootstrapped samples, we found that the effect of confrontation on concerns about expressing and monitoring for future bias was mediated by negself, 95% CI [0.39, 0.90] (see Fig. 2). In a simultaneous mediators analysis involving all affect indexes, we found that negself no longer was a significant mediator, 95% CI [−0.10, 0.71], and discomfort carried the mediation effect instead, 95% CI [0.39, 1.11]. Note that discomfort and negself were very strongly correlated in this sample, $r(199) = 0.85, p < 0.001$, thus making the test of a unique mediating role for negself particularly stringent. Replicating Experiment 2, we also found that negother was a significant mediator, such that confrontation was associated with less future concern as mediated through negother, 95% CI [−0.49, −0.09].

6.2.6. Defensive reaction to confrontation

Consistent with Experiments 1 and 2, confronted participants did not rate women/Blacks as having a poorer fit with science than participants who were not confronted, $p = 0.259$. However, we did find that, overall, participants in the gender bias condition rated women as a significantly poorer fit to science ($M = 2.46, SD = 0.60$), compared to ratings made about Blacks’ fit to science in the race bias type condition ($M = 2.23, SD = 0.44$), $F(1, 194) = 8.74, p = 0.004, d = 0.44$. Unlike in Experiment 2, we did not find a significant effect for participant gender, although there was a trend for male participants to evaluate the applicant as a poorer fit ($M = 1.62, SD = 0.81$) compared to female participants ($M = 1.47, SD = 0.68$), $F(1, 194) = 2.17, p = 0.143, d = 0.20$.

Turning to defensive reactions in the form of blaming the researchers for being “overly sensitive” or “complainers,” we found a marginally significant effect of confrontation, $F(1, 194) = 3.56, p = 0.061, d = 0.25$ (confronted participants; $M = 2.04, SD = 0.97$; not-confronted participants, $M = 1.81, SD = 0.83$). As with the poor fit ratings, we found a trend for male participants to be somewhat defensive in terms of viewing the researchers as complainers ($M = 2.05, SD = 0.98$) compared to female participants ($M = 1.84, SD = 0.85$), $F(1, 194) = 2.59, p = 0.109, d = 0.23$.

6.3. Summary

Experiment 3 revealed that reactions to evidence-based confrontations of gender and race bias were very similar. Although there was evidence of stronger discomfort and depressed affect following race than gender bias confrontation, both confrontations prompted greater negative self-directed affect and future concern about bias. Replicating Experiments 1 and 2, negative self-directed affect mediated the relation between evidence-based confrontation and future concern (although discomfort carried the effect in the simultaneously mediation analysis). These results suggest that an evidence-based confrontation of gender bias has similar effects on negative self-directed affect and future concern about bias as a confrontation of race bias.

Although the effects of confrontation on affect (negative self-directed affect and discomfort, in particular; see Table 2) were strong in Experiments 2 and 3, the effects were not as strong on future concern (Experiment 2, $p = 0.058$ and $d = 0.35$; Experiment 3, $p = 0.043$ and $d = 0.30$). It is possible that our procedure elevated future concern somewhat even in the not confronted condition. Specifically, all participants in Experiments 2 and 3 learned that we were testing for gender bias in STEM—a domain in which injunctive norms opposing gender bias may be fairly well-known given the frequency with which the problem of gender biases in STEM has been covered in the media of late (Morello, 2015). These circumstances may have heightened concern even in the absence of a confrontation. These considerations raise the question of whether injunctive norms opposing bias are critical for the effectiveness of evidence-based gender bias confrontation, which we examined in Experiment 4.

7. Experiment 4

We sought to manipulate injunctive norms opposing gender bias systematically in the context of an evidence-based confrontation of gender bias in Experiment 4. Past research has established the important role that norms opposing bias can play in reducing people’s prejudiced responses (e.g., Crandall et al., 2002; Monteith, Deneen, & Tooman, 1996), but prior confrontation research has not examined whether injunctive norms opposing bias are critical to confrontation effectiveness. To facilitate the manipulation of norms and generalize our research beyond the science domain, we switched our examination to applicants who had applied for a history lab manager position. We manipulated the salience of injunctive norms opposing sexism in history, such that these norms were made salient for some
participants and not for others. This manipulation was crossed with whether or not participants were confronted.

These design details and the resultant experiences of participants as a function of experimental condition are shown in Table 3. By comparing reactions of participants who simply believed they had evaluated the male applicant similarly to others (Cell 1) to reactions of participants who also learned that we were testing for gender biases that are normatively inappropriate (Cell 2), we could determine whether concern about future gender bias was heightened even without confrontation when participants knew we were investigating gender biases that were normatively unacceptable. Furthermore, we could test whether confrontation heightened concern and self-regulatory intentions even if a norm opposing bias was not made salient (Cell 3 vs. 4). In other words, comparing Cells 3 and 4 allowed us to determine how essential societal norms opposing bias were to the effectiveness of the confrontation.

Experiment 4 also examined whether internal and external motivations to respond without prejudice play a role in people’s reactions to confrontations of gender bias. Although people who are primarily externally motivated to respond without prejudice are driven to regulate their bias by social norms and social pressures (Plant & Devine, 1998), they may respond particularly negatively and defensively to the external pressures to regulate bias that are inherent in confrontation (Plant & Devine, 2001). In addition, people who are primarily motivated to respond without prejudice due to their own closely held values may show greater concern about bias and fewer defensive reactions following a confrontation.

### Table 3

<table>
<thead>
<tr>
<th>Norm</th>
<th>Confrontation</th>
<th>Control (no norm made salient)</th>
<th>Oppose Bias (norm opposing gender bias in history made salient)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Confronted</td>
<td>Confronted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cell 1: Evaluated male applicant similar to past evaluations of the male applicant.</td>
<td>Cell 3: Evaluated female applicant more negatively than the applicant—when specified as male—was evaluated in past research.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cell 2: Evaluated male applicant similar to past evaluations of the male applicant. PLUS Research goal is to test for gender biases that are normatively inappropriate.</td>
<td>Cell 4: Evaluated female applicant more negatively than the applicant—when specified as male—was evaluated in past research. PLUS Research goal is to test for gender biases that are normatively inappropriate.</td>
<td></td>
</tr>
</tbody>
</table>

7.1. Method

7.1.1. Power analysis

To estimate the sample size required to detect significant main effects, two-way interactions, and three-way interactions (with IMS and EMS as continuous variables), we conducted a power analysis in SAS (v. 9.4; Lane & Hennes, in press). (Four-way interactions were neither expected nor obtained.) We simulated data assuming medium effect sizes ($r = 0.40$). These simulations estimated that we would have at least 0.80 power to detect all hypothesized effects with 165 participants.

7.1.2. Participants

Participants were 189 Introductory Psychology students at a large Midwestern University who received course credit. Data were removed for three participants who did not follow instructions (e.g., responded with all 1 s) and 19 participants who incorrectly recalled the gender of the applicant they evaluated. The final sample included 167 participants (51% female, $M_{age} = 19.38$, 81% White). Note that 11 participants were missing IMS and EMS scores either because they entered their participant ID number incorrectly so it could not match with prescreen data or they skipped the IMS/EMS items in the prescreen.

7.1.3. Design

A 2 (confrontation: confronted vs. not confronted) × 2 (norm: control vs. oppose bias) between-participants design was used. In addition, participants varied in their motivations to respond without sexism.

7.1.4. Procedure

Participants completed internal and external motivations to respond without sexism scales (IMS and EMS; Klonis et al., 2005) during a departmental prescreen at the start of the semester. Due to prescreen length constraints, we included three (out of five) IMS items (e.g., “Because of my personal values, I believe that using stereotypes about women is wrong.” $M = 5.19$, $SD = 1.31$, $\alpha = 0.87$) and three (out of five) EMS items (e.g., “I try to act nonprejudiced toward women because of pressure from others,” $M = 2.96$, $SD = 1.55$, $\alpha = 0.87$). Responses were made on 1 (strongly disagree) to 7 (strongly agree) scales. Participants were unaware of the relation between their completion of these scales and recruitment for the experiment.

The cover story, materials, and procedure were very similar to those used in Experiments 2 and 3. However, participants in this experiment took the perspective of a History professor and all materials were modified to be appropriate for this domain. Depending on random assignment to confrontation condition, participants evaluated White male (not confronted) or White female (confronted) applicant using the same evaluation items as in the previous experiments.

Next, participants received feedback. Participants in the not confronted, control norm condition simply learned that their ratings of the (White male) applicant were very similar to ratings of this applicant in past research. Then they read a supposed excerpt from a Chronicle of Higher Education article concerning “The Importance of Studying Hiring Practices,” which did not mention anything about gender bias. Thus, participants in the not confronted, control norm condition heard no mention of possible gender bias throughout the experiment.

Participants in the not confronted, oppose bias norm condition were also told that their (male applicant) ratings were similar to the male applicant ratings in past research. However, they also learned that some participants that the applicant was male and others read that the applicant was female, and were told that if the female was not hired, then “some type of gender bias must be operating.” These participants then read a supposed excerpt from a Chronicle of Higher Education article titled “The Importance of Removing Gender Bias from the Hiring Process.”

After decades of hard-fought battles for gender equality, 98% of Americans surveyed in a recent nationally representative study agree very strongly that women should be treated as equals to men in all disciplines, including history. Currently, researchers and non-researchers alike agree that even unconscious bias favoring males should be kept in check because it can create inequality. People want to ensure that their reactions to others are not influenced by any form or sort of gender bias, conscious or unconscious, as they strive for social justice and equality for the genders. In today’s society, women are recognized as male’s equals and should be treated with equality and fairness.

Thus, participants in the not confronted, oppose bias norm condition did not think that they themselves had biased applicant evaluations, but
learned that we were testing for the possibility of gender biases that were considered normatively unacceptable.

Participants in the confronted, control norm condition learned that their ratings of a female applicant compared unfavorably to past ratings of an identical male applicant and would have resulted in rejection of the female applicant. Thus, when the research explained that such a pattern of results would “indicate some type of gender bias,” the implication was that participants had been biased. These participants then went on to read the control article concerning recruiting and hiring.

Finally, participants in the confronted, opposite norm condition were confronted in the same manner as described above and then read the article that made norms opposing gender bias in history salient.

Participants then went on to complete the measures of affect from our prior studies: negself (α = 0.82), discomfort (α = 0.88), depressed (α = 0.84), negother (α = 0.89), positive (α = 0.86), and amused (r = 0.72, p < 0.001); concern about gender bias and self-regulatory intentions (α = 0.94); perceived poor fit between women and the field of history (α = 0.87); perceptions of the researchers as complainers and oversensitive (α = 0.86); and believability of feedback (α = 0.74). Finally, participants answered manipulation check items (about applicant gender and norm information), were debriefed, and asked for permission to use their data (all agreed).

7.2. Results and discussion

7.2.1. Analytic procedure

Each dependent variable was analyzed with hierarchical linear regression. Confrontation condition (0 = not confronted, 1 = confronted), norm condition (0 = control; 1 = oppose bias), IMS (mean centered), EMS (mean centered) and gender (0 = male; 1 = female) were entered and assessed at the first step. Two-way interactions between confrontation condition, norm condition, IMS and EMS were entered and assessed at the second step. Three-way interactions were entered and assessed at the third step, and the interaction among all variables was considered at the fourth step.

7.2.2. Believability of the confrontation

Believability ratings were lower in the confronted (M = 4.42, SD = 1.26) than the not confronted (M = 5.44, SD = 1.12) condition, B = −1.00, SE = 0.19, t(150) = 5.18, p < 0.001. However, believability was above the midpoint in both conditions, suggesting that the cover story and feedback were sufficiently credible.

We also found a main effect of IMS, B = 0.15, SE = 0.07, t(150) = 1.97, p = 0.050, which was qualified by an interaction with EMS, B = 0.16, SE = 0.05, t(144) = 3.38, p = 0.001. Participants who scored high on the EMS were less likely to believe the feedback if they also had low IMS (Ŷ = 4.22) as compared to high IMS scores (Ŷ = 5.18), B = 0.37, SE = 0.11, t(149) = 3.33, p = 0.001. In contrast, participants who scored low on the EMS believed the feedback equally at low and high IMS levels (low EMS/low IMS Ŷ = 5.15; low EMS/high EMS Ŷ = 5.15), p = 0.984.

7.2.3. Affect

Confrontation once again consistently influenced affect (see Table 2). Compared to participants who were not confronted, confronted participants reported greater negself, B = 0.96, SE = 0.19, t(150) = 5.12, p < 0.001, discomfort, B = 0.86, SE = 0.19, t(150) = 4.64, p < 0.001, depressed affect, B = 0.53, SE = 0.16, t(150) = 3.37, p = 0.001, and negother, B = 0.61, SE = 0.20, t(150) = 3.07, p = 0.003, and less positive affect, B = −0.68, SE = 0.22, t(150) = 3.01, p = 0.003, and amusement, B = −0.46, SE = 0.24, t(150) = 1.95, p = 0.054.

A few other significant results also emerged in the affect analyses, although they were neither consistent across affect indexes nor qualified the confrontation main effects. These results are reported in the Supplementary material.

7.2.4. Concern about future prejudice

Replicating Experiments 2 and 3, women reported greater concern (M = 4.26, SD = 1.36) than men (M = 3.40, SD = 1.57), B = 0.85, SE = 0.24, t(150) = 3.62, p < 0.001. The effect of confrontation was marginally significant, B = 0.42, SE = 0.23, t(150) = 1.80, p = 0.074, and the confrontation × norm interaction was significant, B = −1.52, SE = 0.45, t(144) = 3.34, p < 0.001. As can be seen in Fig. 3, either learning that we were testing for gender bias that was normatively unacceptable or confrontation (whether the norm opposing bias was salient or not) elevated concern, relative to participants who were neither confronted or primed with the oppose bias norm. Follow-up comparisons indicated that the effect norm condition was significant among participants who were not confronted, B = 1.03, SE = 0.32, t(149) = 3.21, p = 0.002, but not among confronted participants, B = −0.48, SE = 0.32, t(149) = 1.51, p = 0.133. In addition, the effect of confrontation was significant among control norm participants, B = 1.16, SE = 0.32, t(149) = 3.66, p < 0.001, and not among opposite bias norm participants, B = −0.36, SE = 0.33, t(149) = 1.10, p = 0.275.

These findings establish that confrontation’s effectiveness is not dependent on the salient of injunctive norms opposing bias; indeed, having these norms salient did not even boost the effects of confrontation on future concern. The results also show that concern was heightened when participants knew we were investigating gender bias and normative opposition to such bias was primed. Note, however, that participants in the not confronted, oppose norm condition did not report heightened negself, which is important for instigating regulation of bias in the future (e.g., Monteith et al., 2002). We therefore suspect that the heightened concern in this condition reflects compliance with the norm that was experimentally salient, rather than being internalized and lasting (e.g., Kelman, 1958).

We also found a significant interaction between IMS and confrontation condition, B = 0.41, SE = 0.19, t(144) = 2.12, p = 0.036. As levels of IMS increased, participants who were confronted reported stronger concern about expressing and monitoring for possible gender bias in the future, B = 0.31, SE = 0.13, t(149) = 2.45, p = 0.016, whereas the relation between IMS and concern was not significant among not confronted participants, B = −0.11, SE = 0.12, t(149) = 0.90, p = 0.371. Viewed another way, at +1 SD on IMS scores, confronted participants reported more concern (Ŷ = 4.43) than not confronted participants (Ŷ = 3.45), B = 0.97, SE = 0.36, t(149) = 2.99, p = 0.003. However, at −1 SD on IMS scores, confronted (Ŷ = 3.60) and not confronted (Ŷ = 3.74) participants reported comparable levels of future concern, B = −0.14, SE = 0.33, t(149) = 0.42, p = 0.673. Thus, the confrontation heightened concern and intentions to monitor for future bias to a greater extent as participants’
internal motivation to respond in non-biased ways toward women increased.

7.2.5. Mediation: effect of confrontation on future concern through negself

Using Hayes (2013) PROCESS macro (Model 4) with 5000 bootstrapped samples, the effect of confrontation on concern about gender bias was mediated by negself, 95% CI [0.26, 0.76] (see Fig. 2). This effect remained significant when all other affect indexes were entered as simultaneous mediators, 95% CI [0.26, 1.11], whereas no other index was a significant mediator.

7.2.6. Defensive response to the feedback

Unlike Studies 1–3, confrontation increased agreement with the poor fit items (confronted $M = 1.73$, $SD = 0.93$; not confronted $M = 1.45$, $SD = 0.53$), $B = 0.24$, $SE = 0.11$, $t(150) = 2.12$, $p = 0.036$. Also, poor fit ratings and IMS were negatively related, $B = -0.15$, $SE = 0.04$, $t(150) = 3.43$, $p = 0.001$, whereas poor fit ratings and EMS were positively related, $B = 0.08$, $SE = 0.04$, $t(150) = 2.15$, $p = 0.033$.

Turning to the other index of defensiveness, males were significantly more inclined to claim that the researchers were oversensitive and complainers ($M = 2.51$, $SD = 1.46$) than females ($M = 1.90$, $SD = 1.08$), $B = -0.56$, $SE = 0.20$, $t(150) = 2.84$, $p = 0.005$. We also found significant main effects for confrontation, $B = 0.44$, $SE = 0.20$, $t(150) = 2.24$, $p = 0.026$, IMS, $B = -0.27$, $SE = 0.08$, $t(150) = 3.60$, $p < 0.001$, and EMS, $B = 0.14$, $SE = 0.06$, $t(150) = 2.12$, $p = 0.035$. However, these three main effects were qualified by a significant confrontation X IMS X EMS interaction, $B = -0.24$, $SE = 0.10$, $t(140) = 2.39$, $p = 0.018$. As reflected in Fig. 4, no effects approached significance among participants who were not confronted. However, in the confronted condition, defensive reactions were relatively high among lower IMS/higher EMS participants. Follow-up analyses among confronted participants indicated that the effect of IMS was not significant among relatively low ($−1$ SD) EMS participants, $B = −0.11$, $SE = 0.17$, $t(146) = 0.66$, $p = 0.508$, but was strong among high ($+1$ SD) EMS participants, $B = −0.59$, $SE = 0.13$, $t(146) = 4.41$, $p < 0.001$. Furthermore, the effect of EMS at relatively low levels of IMS was significant, $B = 0.49$, $SE = 0.16$, $t(146) = 2.98$, $p < 0.01$. Thus confronted participants who were highly externally but not internally motivated to avoid gender bias were especially likely to view the researchers as oversensitive and complainers.

7.3. Summary

The results of Experiment 4 establish that an evidence-based confrontation of gender bias has affective and regulatory implications in a non-STEM domain (i.e., history). Furthermore, the effect of confrontation was just as robust when injunctive norms against bias were not made salient in the confrontation context as when these norms were made salient. Also, Experiment 4 is the first study that we know of to examine the effects of individual differences in motivations to respond without gender bias on gender confrontation effectiveness. These motivations were influential in that confrontation generated greater concern about gender bias and intentions to regulate to the extent that participants were internally motivated to respond without bias toward women. Finally, participants who were especially externally but not internally motivated to respond without gender bias were most likely to demonstrate defensiveness toward the researchers.

8. Meta-analyses

To provide an overall summary of the effects of an evidence-based confrontation of gender bias on our primary dependent variables, we conducted meta-analyses using Cumming and Calin-Jageman’s (2017) Exploratory Software for Confidence Intervals. In each of the meta-analyses presented below the fixed effects were weighted by sample size. The meta-analyses included data from the not confronted and evidence-based confrontation conditions in Experiment 1, data from all participants in Experiment 2, data from the not confronted and confronted gender-bias conditions in Experiment 3, and data from all participants in Experiment 4.

8.1. Affect

The meta-analytic comparisons between the evidence-based confrontation and no confrontation condition revealed that confronted participants reported significantly greater negative affect in the form of negself, $d = 0.75$, 95% CI [0.58, 0.93], discomfort, $d = 0.78$, 95% CI [0.60, 0.95], depressed affect, $d = 0.52$, 95% CI [0.34, 0.69], and anger, $d = 0.49$, 95% CI [0.32, 0.67]. Confrontation also reduced positive affect, $d = −0.66$, 95% CI [$−0.84$, $−0.49$], and heightened amusement, $d = 0.24$, 95% CI [0.07, 0.41].

8.2. Future concern

The meta-analytic comparison between the evidence-based confrontation and no confrontation conditions revealed a significant effect for future concern about bias, $d = 0.40$, 95% CI [0.23, 0.57]. As expected, participants reported greater future concern about gender bias following an evidence-based confrontation than no confrontation.

8.3. Poor fit

The meta-analytic comparison between the evidence-based confrontation and no confrontation conditions revealed a small but significant effect when predicting poor fit beliefs, $d = 0.26$, 95% CI [0.09, 0.43], such that participants tended to endorse poor fit beliefs more strongly following an evidence-based confrontation than no confrontation.
8.4. Oversensitive

The meta-analytic comparison between the evidence-based confrontation and no confrontation conditions revealed a significant effect for viewing the researchers as oversensitive and complainers, $d = 0.27$, 95% CI [0.10, 0.44]. As expected, there was greater defensiveness toward the researchers when participants were confronted about gender bias than when they were not confronted.

8.5. Mediation: effect of confrontation on future concern through negself

In addition, we meta-analytically examined our hypothesized mediation model. First, we examined the indirect effect of evidence-based confrontation vs. no confrontation condition on future concern through negself. Following recommendations by Wen and Fan (2015, see also, Preacher and Kelley, 2011) we used fully standardized indirect effects to conduct the meta-analysis. We found that the meta-analytic indirect effect was significant, $\beta = 0.15$, 95% CI [0.07, 0.24]. This effect held even when other affect indices (i.e., discomfort, depressed affect, negother, positive affect) were simultaneously entered as mediators in the model, $\beta = 0.17$, 95% CI [0.08, 0.25]. In contrast, the meta-analyses indicated that no other affect index had a significant mediating role.

In sum, these results are consistent with our expectation that the evidence-based confrontations (vs. no confrontation) of gender bias promotes greater negative self-directed affect, which in turn promotes future concern about gender bias.

9. General discussion

Four studies revealed that a novel evidence-based approach to confronting participants about their gender bias led to a host of affective consequences, including increased negative self-directed affect. Consistent with past research on self-regulation (e.g., Burns et al., 2017; Montieth et al., 2002), this effect elevated the extent to which participants were concerned about their propensity for gender bias and their intentions to monitor and control gender bias in the future. These findings can be contrasted with past studies showing that gender confrontations that did not include evidence of bias were trivialized and dismissed (Czopp & Montieth, 2003; Gulker et al., 2013). The evidence-based confrontation showed participants that gender supposedly played a role in their evaluations of a female applicant and resulted in a discriminatory hiring decision. Unlike past gender confrontation research, a confrontation of this nature meant that participants no doubt found it difficult to claim that they were impervious to gender bias because they “love women” (Eagly et al., 1991; Glick and Fiske, 1996) or that the so-called gender bias was inconsequential (Woodzicka et al., 2015) or even amusing (Czopp & Montieth, 2003).

Furthermore, we found that an evidence-based confrontation had self-regulatory consequences whether it was linked to a gender-science or gender-intelligence bias (Experiment 2); both in a STEM (Physics; Experiments 1–3) and non-STEM (History; Experiment 4) fields, and whether norms opposing gender bias were made situationally salient or not (Experiment 4). We also found that gender bias confrontation produced similar levels of negative self-directed affect and concern about bias compared to racial bias confrontation. This latter finding is particularly important given past research has repeatedly established that people are more concerned about racism than sexism (Gowan & Hodge, 1996; Rodin et al., 1990; Woodzicka et al., 2015) and more responsive to racism than sexism confrontations (e.g., Gulker et al., 2013).

Some features of our experimental context deserve comment, as they may have contributed to the effectiveness of the evidence-based confrontation. From the participants’ perspective, the confrontation came from the investigators responsible for the research—presumably experts with authority and not individuals who were directly implicated by the purported gender bias. An evidence-based sexism confrontation may be less effective if performed by a lay person who was the actual target of the sexism. Also, although we led participants to believe that they exhibited gender bias so that they could be confronted, we actually found that male and female candidates with equivalent credentials were evaluated similarly (Experiment 1). The absence of actual gender bias may be explained in a variety of ways (e.g., the lab manager evaluation task had lower stakes and was less self-relevant among our lay participants than for the STEM faculty in Moss-Racusin et al., 2015, research, so that gender bias may have been less likely to be activated and applied). Regardless, reactions to an evidence-based confrontation may differ when people have actually just exhibited sexism. One possibility is that confrontation becomes more impactful as the extent of actual gender bias increases and becomes more plainly evident to all. There may be a limit, however, as even an evidence-based confrontation may fall on indifferent ears when perpetrators engage in overt biases that they consciously sanction.

Although we successfully identified an effective gender confrontation strategy, a variety of less encouraging caveats must be considered. First, opportunities for confronting people with concrete evidence that their treatment of women in particular situations (e.g., hiring) is more negative than their treatment of men are far from commonplace. In other words, the requisite conditions for instigating social change through evidenced-based confrontation will frequently be unmet. Perhaps greater applicability of evidence-based confrontation can be found in diversity training interventions. Researchers should direct efforts not only to documenting the existence of gender biases and disseminating their results, but also to designing effective experiential tasks and activities that reveal biases and underscore their discriminatory potential. This suggestion is consistent with research showing that experiential learning activities increase people’s receptiveness to evidence of bias more than passive tactics such as lectures (Moss-Racusin et al., 2014; Shields, Zawadzki, & Johnson, 2011). When using evidence-based confrontation, care must be taken not to pressure people so that backlash can be avoided (Plant & Devine, 2001). Also, several tasks concerning a variety of gender biases should be used to increase the chances of generalization to various situations in which gender biases require monitoring and control.

A second caveat stems from our findings that even evidence-based confrontation is less effective for some people than for others (Experiment 4). The confrontation was less likely to generate concern about future gender bias as internal motivation to respond without bias decreased. In addition, people who were low on internal motivation but high on external motivation to respond without prejudice reacted defensively to the confrontation by asserting that the researchers were overly sensitive and complainers (Experiment 4). Finally, Experiments 2 and 3 showed that participants became less concerned about gender bias following confrontation to the extent that it elicited negative affect directed toward others. These findings suggest that evidence-based confrontation of gender bias will not be uniformly effective, and lead us to echo past calls for multi-faceted interventions (Moss-Racusin et al., 2014).

Our results also suggested that men reacted less favorably to our focus on studying possible gender biases than women, even if men were not confronted. Consistent with past research (Handley et al., 2015; Moss-Racusin et al., 2015), we observed trends and a number of significant effects across our experiments indicating that men reacted more defensively than women. Also, men reported less concern than women about gender biases in Experiments 2–4.

A final caveat relates to our use of undergraduate and MTurk samples. Whether our findings would generalize to other types of samples, such as the science faculty who participated in the original Moss-Racusin et al. (2012) research, is unclear. Some research suggests that male STEM faculty can be particularly resistant to scientific evidence of gender bias (Handley et al., 2015, Study 2), which may also be true when they receive evidence-based confrontations about their own
gender biases. We believe these caveats point to the complexity of issues surrounding attempts to reduce gender bias. Nonetheless, our findings do support the efficacy of evidence-based confrontation for many individuals. Whether an initial evidence-based confrontation might produce slight changes in thinking and responding across time, resulting in broader cumulative effects, is a question for future research. Importantly, we suspect that lasting effects would not be observed from simply making norms opposing gender bias salient. Although Experiment 4 indicated that salient norms heightened concern about gender bias, negative self-directed affect was not heightened. Such affect is critical to instigating self-regulatory processes in the future (Experiment 1–4 results; Burns et al., 2017; Monteith et al., 2002), and we suspected that heightened concern in the absence of compunction would not be associated with lowered bias in the future.

In sum, the present research identified a type of confrontation that proved effective at promoting people’s concern about gender bias and their intentions to monitor and control it in the future. By presenting people with concrete evidence of their own gender bias and under-scoring its discriminatory implications, the confrontation could not simply be disregarded. This type of confrontation may be required to discourage discrimination in various contexts, including the sort of disturbing gender bias in hiring practices (Moss-Racusin et al., 2012) that provided the impetus for the present work. We hope that this research will provide inspiration for designing tasks that can convince people of their own susceptibility to harmful gender biases and their intentions to self-regulate in relation them.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jesp.2017.07.009.

References


