

# A “Scientific Diversity” Intervention to Reduce Gender Bias in a Sample of Life Scientists

Corinne A. Moss-Racusin,<sup>\*,†</sup> Jojanneke van der Toorn,<sup>‡</sup> John F. Dovidio,<sup>§</sup>  
Victoria L. Brescoll,<sup>||</sup> Mark J. Graham,<sup>¶</sup> and Jo Handelsman<sup>#</sup>

<sup>†</sup>Department of Psychology, Skidmore College, Saratoga Springs, NY 12866; <sup>‡</sup>Department of Psychology, Leiden University, 2333 AK Leiden, The Netherlands; <sup>§</sup>Department of Psychology, School of Management, <sup>¶</sup>Center for Teaching and Learning, and <sup>#</sup>Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520

## ABSTRACT

Mounting experimental evidence suggests that subtle gender biases favoring men contribute to the underrepresentation of women in science, technology, engineering, and mathematics (STEM), including many subfields of the life sciences. However, there are relatively few evaluations of diversity interventions designed to reduce gender biases within the STEM community. Because gender biases distort the meritocratic evaluation and advancement of students, interventions targeting instructors' biases are particularly needed. We evaluated one such intervention, a workshop called “Scientific Diversity” that was consistent with an established framework guiding the development of diversity interventions designed to reduce biases and was administered to a sample of life science instructors ( $N = 126$ ) at several sessions of the National Academies Summer Institute for Undergraduate Education held nationwide. Evidence emerged indicating the efficacy of the “Scientific Diversity” workshop, such that participants were more aware of gender bias, expressed less gender bias, and were more willing to engage in actions to reduce gender bias 2 weeks after participating in the intervention compared with 2 weeks before the intervention. Implications for diversity interventions aimed at reducing gender bias and broadening the participation of women in the life sciences are discussed.

At the 2015 World Conference of Science Journalists, biochemist and Nobel laureate Tim Hunt delivered controversial remarks suggesting that the inclusion of women in science labs may undermine scientific productivity (Thomason, 2015). After many colleagues critiqued these remarks as inappropriate and undermining women's advancement (*Nature*, 2015), Hunt was asked to resign from several professional positions, including an honorary professorship (Bever, 2015). And yet, his remarks highlight the continued prevalence of subtle gender biases in science, technology, engineering, and mathematics (STEM). Indeed, many have called for the implementation of validated diversity interventions designed to educate science faculty and reduce pernicious gender biases (Al-Gazali *et al.*, 2013; Moss-Racusin *et al.*, 2014; *Nature*, 2015). However, relatively few validated interventions targeting gender bias are available for use in the STEM community (Moss-Racusin *et al.*, 2014). Although some recent studies suggest that existing interventions can reduce gender bias (Jackson *et al.*, 2014; Carnes *et al.*, 2015), relatively little empirical research has assessed the effectiveness of existing STEM gender bias interventions. The current work addresses this gap by examining life science instructors' responses to an evidence-based intervention designed to target subtle gender bias.

Kenneth Gibbs, *Monitoring Editor*

Submitted September 5, 2015; Revised November 12, 2015; Accepted November 12, 2015

CBE Life Sci Educ September 1, 2016 15:ar29

DOI:10.1187/cbe.15-09-0187

\*Address correspondence to: Corinne A. Moss-Racusin (cmossrac@skidmore.edu).

© 2016 C. A. Moss-Racusin *et al.* CBE—Life Sciences Education © 2016 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

## GENDER BIAS IN ACADEMIC SCIENCE

Women's participation in the life sciences has increased over recent decades, particularly relative to other STEM fields. For example, women earned 52% of doctoral degrees in the life sciences in 2012, up from 37% in 1993; in comparison, women earned just 20% of doctoral degrees in physics in 2012, up slightly from 13% in 1993 (National Science Foundation [NSF], 2012). And yet, female life scientists remain underrepresented among senior faculty, grantees, and award winners. For example, as of 2006, women comprised only 26% of full professors in the life sciences (NSF, 2008) and constituted only 5% of Nobel laureates in physiology or medicine (Nobel Foundation, 2015). Additionally, female undergraduate life science students appear to be underrepresented in large-classroom discussions; although they comprise ~60% of students enrolled in large biology classes, one study found that their voices reflect fewer than 40% of responses to instructor questions, a critical method of engaging with and mastering course material (Eddy *et al.*, 2014). To the extent that women's restricted involvement undermines access to talent and slows scientific progress, the scientific community stands to benefit from boosting women's participation in the life sciences and other STEM fields.

Researchers have sought to identify factors responsible for the lingering underrepresentation of women in academic science. For example, some evidence suggests that early sex differences in attitudes toward and interest in STEM fields (such that boys are more likely than girls to express positive attitudes and interest in STEM) may contribute to the gender disparity, though these attitudes are likely associated with socialization processes communicating that STEM is more appropriate for boys and men (vs. girls and women; Ceci *et al.*, 2014). Additionally, women's participation may be undermined by a lack of female role models (Dasgupta, 2011; Stout *et al.*, 2011) and peers (Dasgupta *et al.*, 2015).

In addition to these factors, mounting evidence suggests that persistent gender biases favoring men may undermine women's progress in STEM. For example, across cultures and developmental stages, children are more likely to produce images of men than women when asked to draw a scientist (Finson *et al.*, 1995). Similarly, adults are more likely to associate science with men than with women (Nosek *et al.*, 2002). Finally, journal articles with female authors are less likely to be cited than articles with male authors (Lariviere *et al.*, 2013), and Dutch female scientists' grant applications are less likely to be funded and receive high "quality of researcher" scores relative to male colleagues (Van der Lee and Ellemers, 2015).

Experimental evidence bears out these correlational trends. For example, a laboratory experiment demonstrated that undergraduate participants were more likely to hire a male applicant for a mathematical task relative to the identical female, even when objective performance data were provided (Reuben *et al.*, 2014). Although it is reasonable to expect that STEM professionals might not demonstrate these types of biases (due to their rigorous training in scientific objectivity), further experimental evidence suggests that gender biases are observed among STEM participants as well as the general population. For example, both male and female STEM faculty favored a male lab manager applicant relative to the identical female applicant (Moss-Racusin *et al.*, 2012)

and were more likely to agree to a mentoring meeting with a male doctoral candidate than with the identical female candidate (Milkman *et al.*, 2015). Thus, persistent gender biases may significantly undermine women's advancement within STEM fields; conversely, developing methods to reduce gender bias may help to broaden women's participation.

## THE NEED FOR EFFECTIVE GENDER BIAS INTERVENTIONS

Because STEM educators are responsible for training, evaluating, and mentoring the next generation of scientists, they are a particularly critical target group for evidence-based interventions designed to reduce gender biases. Simply put, STEM educators have the opportunity to either perpetuate or interrupt the transmission of existing gender biases to the next generation of scientists. Moreover, their willingness to mentor female students (or lack thereof; Sheltzer and Smith, 2014) may directly affect women's persistence in academia (Pfund *et al.*, 2015). Although this population of dedicated educators may be committed to scientific meritocracy and fair treatment, experimental evidence suggests that they are subject to the same subtle gender biases observed in the general population (Steinpreis *et al.*, 1999; Moss-Racusin *et al.*, 2012; Milkman *et al.*, 2015). Moreover, experimental tests (Moss-Racusin *et al.*, 2012; Milkman *et al.*, 2015) reveal that life science faculty (e.g., biologists) are just as likely to demonstrate gender biases as faculty in other STEM fields (e.g., physics, mathematics). Thus, it is critical to implement tested gender bias interventions among life science instructors, with the broader aim of increasing the participation of talented female scientists (Moss-Racusin *et al.*, 2014).

There are few validated diversity interventions available, and almost none targeting STEM instructors' gender biases (Moss-Racusin *et al.*, 2014). Problematically, because the majority of existing interventions have not been subject to systematic evaluation, the efficacy of current approaches remains largely unknown (Paluck, 2006; Paluck and Green, 2009). Of further concern, evidence suggests that some interventions may actually backfire, exacerbating biases and undermining diversification efforts (Legault *et al.*, 2011; Dobbin and Kalev, 2013; Dobbin *et al.*, 2015).

To our knowledge, only two evidence-based interventions designed to reduce STEM faculty gender biases have been systematically evaluated and validated in randomized controlled trials (RCTs). Specifically, one intervention increased awareness of and motivation to address gender bias among University of Wisconsin–Madison faculty relative to wait-list controls (Carnes *et al.*, 2015). Another intervention significantly reduced male faculty's implicit (or automatic, nonconscious; Nosek *et al.*, 2002) gender biases relative to controls (Jackson *et al.*, 2014). A third RCT revealed that established social psychological bias-reduction techniques (e.g., stereotype replacement, counterstereotypic imaging) reduced implicit racial bias, but did not examine gender bias, and used undergraduate psychology student participants (Devine *et al.*, 2012). This promising evidence demonstrates the potential for diversity interventions to bring about meaningful bias-related changes and highlights the need for further work assessing the effectiveness of gender bias-reduction interventions among STEM educators.

## ACTION READINESS

Egalitarian attitudes themselves are insufficient to produce positive changes if they are not also accompanied by relevant behaviors designed to combat gender bias. Indeed, people who believe that they are objective often paradoxically express high levels of gender discrimination, suggesting that biases are frequently unintentional and that egalitarian attitudes do not necessarily ward off inequitable behaviors (Uhlmann and Cohen, 2007). Thus, in addition to increasing awareness of diversity issues and reducing subtle gender bias, successful interventions should also increase attendees' readiness to take action on diversity-related issues (Moss-Racusin *et al.*, 2014). With this in mind, we turned to the literature to identify a theoretically grounded measure of action readiness.

A large body of literature on self-regulation (i.e., the ways in which people maintain motivation and focus throughout a task; Crowe and Higgins, 1997) has identified two primary strategies for managing goal-related behavior, such as tackling diversity challenges in a professional academic context. An approach orientation (or *promotion focus*) emphasizes goal attainment and accomplishment, such as working hard to receive a coveted job offer. In a diversity context, someone utilizing a promotion focus might strive to create positive interactions and parity between diverse students in the classroom and actively seek out opportunities to mentor talented students from a variety of demographic backgrounds. In contrast, an avoidance orientation (or *prevention focus*) is a more conservative approach focused on risk aversion, such as working hard to avoid being fired. Someone employing a prevention focus might attempt to evade discussion of diversity-related topics in the classroom or avoid potentially uncomfortable diversity-related conversations with mentees.

Previous research has demonstrated that the use of a promotion focus (relative to a prevention focus) is often associated with superior performance across a host of different types of tasks (Crowe and Higgins, 1997; Higgins *et al.*, 1997; Moss-Racusin and Rudman, 2010; Rudman *et al.*, 2012), including behavior related to bias and diversity (Trawalter and Richeson, 2006; Does *et al.*, 2011). This work suggests that increased promotion focus (rather than decreased prevention focus) is the critical predictor of behavioral outcomes. Thus, we were primarily interested in determining whether promotion focus increased after exposure to a diversity-training intervention (i.e., that participants became more focused on attaining positive diversity-related outcomes) and also predicted that prevention focus would not increase (i.e., that participants did not become more likely to avoid tackling diversity-related challenges).

## THE CURRENT RESEARCH

We sought to address gaps in the existing literature by providing an evaluation of a workshop called "Scientific Diversity." This brief, evidence-based gender bias intervention has been widely implemented among life science instructors at the National Academies Summer Institutes (NASIs), a series of annual nationwide workshops that train life science faculty in the principles of scientific teaching (Handelsman *et al.*, 2004, 2007; Wood and Handelsman, 2004; Miller *et al.*, 2008; Pfund *et al.*, 2009; Couch *et al.*, 2015). To recruit life science instructors as participants, we partnered with NASI leaders, inviting life scientists who attended a NASI in the Summer of 2012 to

participate in our study. Those who elected to participate completed baseline measures (pretest) 2 weeks before attending the "Scientific Diversity" workshop at a NASI. They then completed the same set of measures 2 weeks after completing the workshop (posttest), allowing us to assess change over time.

Our objective was to assess the extent to which the "Scientific Diversity" workshop was associated with improvements in STEM faculty participants' awareness of diversity issues, gender bias, and readiness to take action on diversity issues. We tested the hypothesis that the "Scientific Diversity" workshop increases participants' awareness of gender bias, reduces participants' gender bias, and increases participants' readiness to take action on diversity issues. Specific predictions were as follows:

Prediction 1: Participants' awareness of diversity issues will increase from pretest to posttest.

Prediction 2: Participants' gender bias will decrease from pretest to posttest.

Prediction 3: Participants will demonstrate optimized action readiness at pretest relative to posttest, such that their approach orientation will increase, while their avoidance orientation will remain stable or decrease.

## MATERIALS AND METHOD

### Participants

Of 212 total eligible participants (i.e., NASI attendees), 150 completed the survey at pretest and posttest, for an overall participation rate of 71%. This is significantly higher than participation rates typically obtained in field studies of professionals, which can dip as low as 30% (Steinpreis *et al.*, 1999; Moss-Racusin *et al.*, 2012). Of these 150 participants, 24 were participants of color. In keeping with other studies evaluating the efficacy of diversity interventions (Devine *et al.*, 2012; Prime *et al.*, 2012), the current analyses were restricted to data from white participants. The reasoning behind this decision was twofold. First, previous research has indicated that participants of color often respond differently than white participants to issues pertaining to diversity (Purdie-Vaughns *et al.*, 2008). Thus, including both white participants and participants of color in the current study without statistically accounting for participant race would likely have obscured results. Second, because the NASI attendees were majority white, it was not possible to collect responses from enough participants from any single racial minority group to obtain sufficient statistical power to adequately test for racial differences. Thus, to avoid reporting inaccurate results, we restricted the current analyses to white participants, resulting in a final sample of 126 (70% female) for our primary analyses (although results were unchanged when participants of color were included in analyses<sup>1</sup>). We strongly

<sup>1</sup>Despite these important caveats, we examined whether the overall pattern of results differed if participants of color were included in analyses. For these analyses, we included all 150 participants. Of importance, the observed pattern of results did not deviate from those obtained with the smaller sample of 126 white participants. Supporting prediction 1, participants were more aware of diversity issues following the intervention,  $b = 0.31$ ,  $SE = 0.11$ ,  $t(147.99) = 2.96$ ,  $p = 0.01$ . Supporting prediction 2, participants expressed less subtle gender bias following the intervention,  $b = -0.11$ ,  $SE = 0.06$ ,  $t(148.56) = -1.99$ ,  $p = 0.04$ . Finally, supporting prediction 3, participants expressed greater promotion focus following the intervention,  $b = 0.30$ ,  $SE = 0.06$ ,  $t(183.57) = 5.25$ ,  $p < 0.001$ , but their prevention focus remained stable,  $b = -0.04$ ,  $SE = 0.07$ ,  $t(148.00) = -0.52$ ,  $p = 0.61$ .

suggest that future research attempt to recruit a diverse faculty participant sample, enabling more sophisticated tests to determine whether attendees from different racial backgrounds respond differently to diversity-training interventions.

Participants were academic scientists working primarily in the life sciences (88% life/biological sciences), with a minority drawn from other related STEM fields (7% chemistry, 4% physics, 1% other). Participants' average age was 45 (SD = 10.48, range 28–70). Of participants, 13% were full professors, 22% were associate professors, 19% were assistant professors, 28% were non-tenure-track instructors, 9% were postdoctoral associates, and 9% reported another title. Using the institutional categories, 71% were from “very high research activity” research universities (R1), 15% were from “high research activity” or doctoral/research universities (R2), 10% were from liberal arts colleges, 2% were from community colleges, and 2% were from other types of institutions; 77% taught at public institutions. Of importance, no significant effects were associated with these participant demographic variables ( $t < 1$  and  $p > 0.05$  for all analyses); as a result, they are not discussed further.

### “Scientific Diversity” Workshop

Our intervention took the form of an evidence-based workshop entitled “Scientific Diversity,” which lasted for ~120 min and has been implemented by more than a dozen different leaders with more than 1400 scientist participants since 2004 (Wood and Handelsman, 2004; Pfund *et al.*, 2009). As its name implies, the goal of the workshop was to take a scientific approach to diversity intervention. Specifically, trained facilitators introduced attendees to empirical social science literature on the existence and nature of subtle biases and the benefits of inclusive, heterogeneous learning environments. Of importance, the workshop was consistent with a framework derived from a systematic review of the existing literature on the efficacy of existing diversity interventions (Moss-Racusin *et al.*, 2014).

This framework identified four design elements shared by successful interventions and recommended the targeting and assessment of at least three key measurable outcome variables. Essential design elements were as follows: A) interventions based on theory and empirical evidence, rather than intuition; B) approaches that utilize active learning to foster participants' dynamic engagement with workshop content; C) presentation of diversity as a shared goal and responsibility rather than as the fault of one group or individual; and D) incorporation of rigorous evaluation to assess the intervention's efficacy. Essential outcomes were 1) increase in participants' awareness of diversity issues; 2) reduction of participants' biases; and 3) preparedness of participants to take action on diversity-related issues rather than avoid diversity challenges (see Table 1 for a list of all intervention design elements, outcome variables, and associated key results).

Individual NASI leaders and “Scientific Diversity” instructors were free to tailor the specific workshop activities and assign preparatory readings based on the particular needs of attendees (and we accounted for potential outcome differences associated with this modest pedagogical variability in our analyses—see *Analytic Strategy* below), but the “Scientific Diversity” workshop followed the same general format at each NASI. This format was consistent with the four design ele-

**TABLE 1. Design elements (A–D) and measured outcomes (1–3) of the “Scientific Diversity” workshop intervention, as well as key results associated with each prediction (1 after 3)<sup>a</sup>**

Intervention design elements		
A	Based on existing theory and empirical evidence	
B	Utilized active-learning techniques to foster participants' engagement	
C	Presented diversity as a shared goal; avoided assigning blame for existing challenges	
D	Included evaluation to assess intervention efficacy	
Intervention outcomes		Key result
1	Participants' awareness of diversity issues	Awareness of diversity issues increased significantly from pretest to posttest.
2	Participants' gender bias	Gender bias decreased significantly from pretest to posttest.
3	Participants' readiness to take action on diversity-related issues	Action readiness was optimized, such that approach orientation increased significantly from pretest to posttest, while avoidance orientation remained constant.

<sup>a</sup>The pretest occurred 2 weeks before participants attended the “Scientific Diversity” workshop, and the posttest occurred 2 weeks after the workshop.

ments and three outcomes identified in the diversity intervention framework discussed above (Moss-Racusin *et al.*, 2014). Specifically, addressing design elements A and B, the workshop was designed from the theoretical perspective of “scientific teaching,” an evidence-based approach to teaching that combines frequent assessment, inclusiveness for diverse students, and *active learning* (a collection of teaching methods that engage learners and provide practice in scientific thinking; Handelsman *et al.*, 2004, 2007; Miller *et al.*, 2008; for a detailed list of scientific teaching goals and a taxonomy of observable practices, see Couch *et al.*, 2015). For example, rather than simply stating in a lecture format that prejudice can be subtle or that diverse classroom environments optimize learning, instructors gave a detailed, interactive presentation of the empirical research on prevalent automatic or *implicit* biases (Greenwald *et al.*, 2015) and the benefits of heterogeneous learning environments (Chamany *et al.*, 2008).

From this, instructors generated active discussion, encouraging participants to draw their own conclusions from the data. Specifically, instructors created structured small-group activities to facilitate participants' engagement with and critical analysis of the empirical evidence (targeting outcomes 1 and 2) before returning to a larger group discussion. To help participants prepare to tackle diversity-related issues (targeting outcome 3), instructors and participants practiced techniques for creating an inclusive academic environment, such as classroom exercises targeting diverse learning styles, mentoring approaches tailored to the needs of individual students,

and approaches to mastering difficult conversations about sensitive or uncomfortable diversity topics.

Addressing design element C, the workshop emphasized shared responsibility for addressing diversity challenges and avoided assigning blame for diversity issues. For example, facilitators shared empirical findings indicating that subtle cultural stereotypes are pervasive (Devine, 1989); thus, biases are often expressed even by well-intentioned individuals who value egalitarianism and fairness (Uhlmann and Cohen, 2007). Additionally, instructors stressed that the literature typically does not indicate differences between people belonging to different demographic groups (e.g., male faculty are no more likely than female faculty to exhibit bias against women in STEM; Moss-Racusin *et al.*, 2012; Milkman *et al.*, 2015), suggesting that the blame for diversity challenges does not lie with members of specific groups. Finally, addressing design element D, this paper reports the findings of the first evaluation of the intervention's efficacy. Additional systematic assessments are ongoing.

We measured the three key outcome variables identified by the diversity intervention framework discussed in the introduction (Moss-Racusin *et al.*, 2014). Although the workshop touched on the importance of diversity broadly defined (including diversity associated with racial background, age, learning style, etc.), it particularly emphasized literature and activities on gender diversity and bias. For this reason, we focused the assessment on variables pertaining to gender bias, though future work should assess additional potential effects on other types of bias.

#### Awareness of Diversity Issues (Outcome 1)

Previous research has indicated that increasing participants' awareness and knowledge of diversity issues is a critical component of successful diversity interventions (Shields *et al.*, 2011; Carnes *et al.*, 2012, 2015; Prime *et al.*, 2012; Zawadzki *et al.*, 2012). Although prior work has assessed this construct by directly asking participants to self-report the extent to which they are aware of diversity issues (e.g., Prime *et al.*, 2012), these measures may be hampered by social desirability concerns, particularly after repeated measurements. That is, such explicit measures may alert participants to the nature of the construct being measured and lead them to edit their responses accordingly (see the *Discussion* section for an analysis of social desirability concerns pertinent to the current research).

Thus, drawing upon the existing literature, we chose to assess this construct in a slightly more indirect way. Specifically, one indicator of increased diversity awareness is attendees' heightened ability to accurately detect the relative homogeneity of their environment (Case, 2007). We predicted that participants would be more likely to perceive the relative lack of diversity in their home departments (which likely reflected the consistent nationwide underrepresentation of women and people of color in STEM fields; NSF, 2012) following their exposure to the "Scientific Diversity" workshop. We chose to focus on the perceived diversity of participants' departments (rather than their larger academic institutions) due to the fact that, although college and university communities as a whole now often approach gender parity, women remain significantly underrepresented within many STEM departments (NSF, 2012). Thus, assessment at the department rather than institutional level provided a more precise measurement of the relevant underlying

construct. Specifically, we utilized the item "To what extent do you think that your own department is diverse?" Participants responded using the scale 1 (not at all diverse) to 7 (very diverse). This scale was reverse-coded, such that higher numbers indicate more awareness of diversity issues.

#### Subtle Gender Bias (Outcome 2)

In keeping with previous research (Moss-Racusin *et al.*, 2012) we measured a subtle, contemporary form of bias. We reasoned that, as committed instructors and mentors, participants would be unlikely to display traditional, "old-fashioned" intentional forms of bias, characterized by explicit hostility toward and an overt intention to prevent the success of stigmatized groups (such as racial minorities and women; Glick and Fiske, 1996; Dovidio and Gaertner, 2004). Rather, we investigated a more subtle form of modern bias, which stems from pervasive cultural stereotypes, frequently exists beyond individuals' conscious awareness, and does not reflect an overt intention to discriminate or cause harm. Thus, we employed the widely used and well-validated Modern Sexism Scale (Swim *et al.*, 1995, 2005) to assess subtle gender bias. This scale includes eight items, such as "Discrimination against women is no longer a problem in the United States" and "Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences," to which participants responded on a scale of 1 (disagree strongly) to 7 (agree strongly). Endorsement of these items reflects subtle, contemporary gender bias (e.g., dismissing or minimizing concerns about ongoing gender discrimination). Responses to items were averaged to create the Modern Sexism Scale, on which higher numbers indicated greater levels of subtle gender bias ( $\alpha_{\text{preintervention}} = 0.84$ ,  $\alpha_{\text{postintervention}} = 0.85$ ).

#### Action Readiness (Outcome 3)

We used a previously validated scale (Rudman *et al.*, 2012) tailored to diversity goals to measure acute levels of both promotion and prevention foci as related to handling diversity-related challenges. Ten promotion focus items included "Right at this minute, in terms of my approach to diversity, I'm feeling ... free to pursue my goals/confident that I can go after my goals/focused on what I will achieve." Ten prevention focus items included "Right at this minute, in terms of my approach to diversity, I'm feeling ... as though I need to avoid risks/like I don't want to make any mistakes/like I want to make sure nothing bad happens." Responses were indicated on a 1 (strongly disagree) to 7 (strongly agree) scale. Items reflecting promotion and prevention focus were averaged separately to form the promotion ( $\alpha_{\text{Pretest}} = 0.81$ ,  $\alpha_{\text{Posttest}} = 0.88$ ) and prevention ( $\alpha_{\text{Pretest}} = 0.89$ ,  $\alpha_{\text{Posttest}} = 0.90$ ) focus scales, with higher numbers reflecting greater levels of promotion or prevention focus.

#### PROCEDURE

Participants were attendees at one of seven NASIs that were held nationwide between May and August 2012. These NASIs were designed to train life science instructors in the principles of scientific teaching (Handelsman *et al.*, 2004, 2007; Miller *et al.*, 2008; Couch *et al.*, 2015) and are described in detail elsewhere (Wood and Handelsman, 2004; Pfund *et al.*, 2009). Attendees from each of the seven 2012 NASIs participated in our study. Specifically, 2012 NASIs took place at the University

of Georgia–Athens (study participant  $N = 22$ ), West Virginia University (study participant  $N = 12$ ), Harvard University (study participant  $N = 25$ ), University of Minnesota–Twin Cities (study participant  $N = 20$ ), Louisiana State University (study participant  $N = 13$ ), Evergreen State College (study participant  $N = 19$ ), and University of Colorado–Boulder (study participant  $N = 15$ ). Each NASI included the “Scientific Diversity” workshop approximately two-thirds of the way through the weeklong curriculum.

Attendees were invited to participate in the current research after their NASI acceptance. During the online NASI registration process, NASI leaders informed attendees that they would be invited to participate in a research project conducted by an affiliated research team housed at Yale University that was not involved in delivering the NASI. To reduce self-selection concerns, we disguised the true nature of the study; participants were told that we were interested in learning about their general social attitudes, for the purposes of refining existing professional development programs. Assessments took place at two time points, and all measures were completed online using participants’ own computers. At pretest (2 weeks before attending the NASI), participants received an emailed invitation to participate in the study originating from the email address of one of the authors (M.J.G.). After granting informed consent, participants ensured that their responses would remain anonymous by creating a unique participant ID code, which enabled their responses at the two time points to be linked without being associated with their identity. They then completed an online questionnaire created using the survey administration program Qualtrics. This survey included all measures presented in a random order, and items within each measure were also randomized.

Participants who completed the pretest questionnaire received another emailed invitation to complete a nearly identical online questionnaire at posttest (2 weeks following the completion of the NASI). Although some prior work has investigated the immediate impact of bias-reduction techniques (e.g., Lai et al., 2014), comparatively few studies have assessed longer-term postintervention effects. Thus, we sought to measure the extent to which the intervention was associated with changes in participants’ awareness of diversity issues, subtle gender bias, and action readiness after the passage of a reasonable amount of time and once they had returned to their home environments. We also included an immediate postintervention assessment for a subset of participants and measures, as described in the *Supplemental Analyses* section below. The only difference between the pre- and posttest questionnaires was that participant demographics were not collected at the pretest and instead were measured only at the end of the posttest (to ensure that measuring demographics did not impact participants’ responses to other questions; Steele and Aronson, 1995).

On completion of the posttest measures, participants underwent a “funnel debriefing” (designed to probe for unusually high levels of suspicion that might render results invalid) before being fully informed as to the nature and goals of the current research (Bargh and Chartrand, 2000).<sup>2</sup> No participants

<sup>2</sup>A funnel debriefing is designed to detect any participants who accurately identified the true nature of the research, because this unusually high level of awareness could have distorted their responses (Bargh and Chartrand, 2000). However, it is not sufficient to simply inform participants of the purpose of the experiment and then ask them to report whether they had inferred it at any point, because

reported problematic suspicions. Participants who completed the pretest measures but elected not to complete the posttest measures underwent the debriefing after the posttest data collection period had ended. After being fully debriefed, participants were thanked for their time. No compensation was offered in exchange for participation. The study was reviewed and determined to be exempt by the Yale University Institutional Review Board.

## SUPPLEMENTAL DATA COLLECTION: IMMEDIATE POSTINTERVENTION EFFECTS

A subset of participants ( $N = 78$ ) also completed the action-readiness scales immediately after completing the “Scientific Diversity” workshop, to assess the acute relationship between attending the workshop and propensity to engage in behaviors addressing diversity challenges. Immediately following the intervention, all participants were given the option to complete paper-and-pencil versions of the promotion focus ( $\alpha_{\text{immediately postintervention}} = 0.86$ ) and prevention focus ( $\alpha_{\text{immediately postintervention}} = 0.84$ ) subscales. The scales were administered in the same room as the “Scientific Diversity” workshop, and scale items were presented in one of two random orders. The demographics of the subset of participants who chose to complete these immediate postintervention scales did not meaningfully differ from the full sample.<sup>3</sup>

## RESULTS

### Analytic Strategy

Because the data were collected at different NASIs held across the country during the Summer of 2012, there were many potential sources of variability that could have influenced participants’ responses to the surveys. For example, there might have been preexisting differences associated with participants’ choices to attend one NASI over another. Moreover, each NASI was run by different organizers, was held at a different academic institution, exposed participants to different coattendees, and so on. Finally, the “Scientific Diversity” workshop at each NASI was taught by different instructors. Thus, it was necessary to ensure that our statistical approach accounted for the potential variability across NASIs and successfully assessed change in

known “hindsight bias” effects could distort the accuracy of their reports (Hawkins and Hastie, 1990). Instead, a funnel debriefing involves posing increasingly specific questions to probe for participants’ levels of awareness before fully revealing the true nature of the research. In this case, we first broadly asked participants to report their perceptions of the aim of the study. Next, participants were asked to indicate whether anything had seemed suspicious or “off” at any point during the study. Finally, participants were told that the study was not solely intended to measure general social attitudes, for the purposes of refining existing professional development programs (i.e., the cover story they had been presented), and were asked to speculate as to possible true aims of the study.

<sup>3</sup>Seventy-eight white participants (62% of the full white sample) chose to complete the immediate postintervention measures (data from an additional 20 participants of color were not analyzed, consistent with the rationale described in the *Participants* section above). Of the subset of participants, 94% were life/biological scientists, 3% were chemists, 2% were physicists, and 1% was from another field; 72% of this subset was female and had an average age of 44 ( $SD = 10.65$ , range 30–64). Further, 13% were full professors, 18% were associate professors, 23% were assistant professors, 25% were non-tenure-track instructors, 12% were postdoctoral associates, and 9.0% reported another title; 65% were from R1 universities, 16% were from R2 universities, 13% were from liberal arts colleges, 4% were from community colleges, and 2% were from other types of institutions; 87% taught at public institutions.

the variables of interest over time beyond the potential variance associated with different “Scientific Diversity” workshops.

In other words, to take into account the fact that the time points were nested within participants and participants were nested within NASIs, which might have led to nonindependent observations (Kenny and Judd, 1986, 1996), we conducted linear mixed-effects modeling analyses in SPSS (version 20.0). We specified a random intercepts multilevel model (which allows the intercepts to vary) with a variance components (VCs) covariance structure. This structure models a different VC for each random effect, meaning that the errors of the different levels are independent from one another. We included time point (pretest = 0; posttest = 1) and gender (female = 0; male = 1) as fixed effects and participant and which NASI participants attended as random effects.

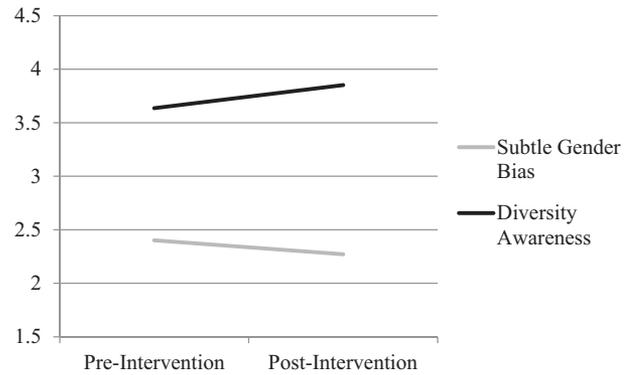
The random effect of NASI was not significant for each of the analyses.<sup>4</sup> This indicates that there was no greater correlation between the responses of participants within summer institute than there was between the participants in general, suggesting that responses did not vary systematically as a function of which NASI participants attended (i.e., were not non-independent at the level of NASI).

We also considered the role of participant gender in our analyses. Although the majority of previous research found that men and women are equally likely to express gender biases (Moss-Racusin *et al.*, 2012; Milkman *et al.*, 2015), some past research has suggested that men may exhibit slightly higher Modern Sexism Scale scores than women (Swim *et al.*, 1995, 2005). To account for any potential differences between male and female participants, we thus included gender as a covariate in all analyses.

### EVIDENCE OF INTERVENTION EFFECTIVENESS

In keeping with predictions 1–3, evidence emerged to support the efficacy of the “Scientific Diversity” workshop in that participants exhibited positive diversity-related changes 2 weeks after their exposure to the intervention relative to 2 weeks before the intervention (Figures 1 and 2). Supporting prediction 1, participants demonstrated increased awareness of diversity issues 2 weeks after the diversity intervention, indicated by heightened awareness of the relative demographic homogeneity of their academic departments,  $b = 0.22$ ,  $SE = 0.11$ ,  $t(123.82) = 1.96$ ,  $p = 0.04$ . Supporting prediction 2, participants also demonstrated reduced subtle gender bias, as shown by a decrease in scores on the Modern Sexism Scale after the diversity intervention,  $b = -0.13$ ,  $SE = 0.06$ ,  $t(124.53) = -2.14$ ,  $p = 0.03$ . Finally, supporting prediction 3, attendees showed evidence of readiness to take action on diversity issues: their approach orientation (promotion focus) increased after the diversity training,  $b = 0.30$ ,  $SE = 0.06$ ,  $t(123.62) = 4.80$ ,  $p < 0.001$ , while their

<sup>4</sup>The model for subtle gender bias did not converge, because the final Hessian matrix was not positive definite, suggesting that there was no variation in the data for the random effect of the summer institute. Hence, we ran the analysis excluding this random effect and report its results. Though the random effect for summer institute was not significant for awareness of diversity issues and action readiness, no convergence problems occurred for these dependent variables. As conducting the analyses excluding this random effect yielded nearly identical estimates, we report the results from the analyses that included both random effects for these variables.

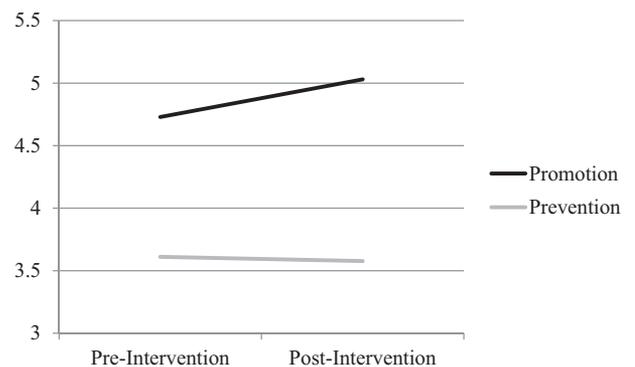


**FIGURE 1.** Means for subtle gender bias and diversity awareness variables measured at pretest (2 weeks before the “Scientific Diversity” workshop) and posttest (2 weeks after the “Scientific Diversity” workshop). Scales range from 1 to 7, with higher numbers reflecting a greater extent of each variable (a truncated scale range is presented for ease of interpretation). After the intervention, diversity awareness increased significantly ( $p = 0.04$ ), and subtle gender bias decreased significantly ( $p = 0.03$ ).  $N = 126$ .

avoidance orientation (prevention focus) remained stable,  $b = -0.03$ ,  $SE = 0.08$ ,  $t(124.06) = -0.45$ ,  $p = 0.66$ .

### THE ROLE OF PARTICIPANT GENDER

Participant gender did not significantly impact the awareness of diversity issues or action readiness outcomes,  $t < 0.95$  and  $p > 0.34$  for all analyses. That is, male and female faculty did not demonstrate different levels of awareness of diversity issues or action readiness. However, a significant main effect of participant gender on subtle gender bias was observed at pretest. This indicates that, before the intervention, male participants displayed higher levels of modern sexism ( $M = 2.79$ ,  $SD = 0.76$ )



**FIGURE 2.** Means for action-readiness variables measured at pretest (2 weeks before the “Scientific Diversity” workshop) and posttest (2 weeks after the “Scientific Diversity” workshop). Scales range from 1 to 7, with higher numbers reflecting a greater extent of each variable (a truncated scale range is presented for ease of interpretation). After the intervention, approach orientation (promotion focus) increased significantly ( $p < 0.001$ ), while avoidance orientation (prevention focus) was unaffected ( $p = 0.66$ ).  $N = 126$ .

than did female participants ( $M = 2.38$ ,  $SD = 0.89$ ),  $t(124) = 2.49$ ,  $p = 0.01$ . This indicates that our pretest data support some prior research suggesting that, under some circumstances, men may express higher levels of gender bias than women (Swim *et al.*, 1995, 2005).

Notably, this gender difference was no longer present at posttest,  $t(123) = 1.46$ ,  $p = 0.15$ , suggesting that the intervention may have been particularly effective at reducing male participants' subtle gender bias. However, the interaction between gender and time did not significantly predict subtle gender bias,  $b = -0.17$ ,  $SE = 0.13$ ,  $t(123.39) = -1.27$ ,  $p = 0.21$ . This suggests that, even though the differences between men and women were significant at pretest but not at posttest, the broader interaction effect of time and gender was not reliable. Thus, we caution against drawing strong conclusions from the pretest gender difference in subtle gender bias without further replication (see *Discussion* section).

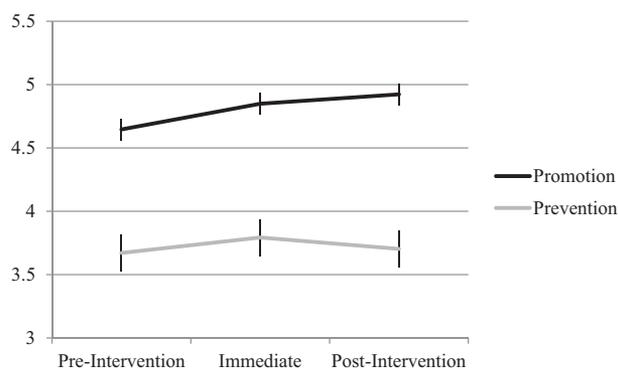
### SUPPLEMENTAL ANALYSES: IMMEDIATE POSTINTERVENTION EFFECTS

As discussed in the *Supplemental Data Collection* section above, a subsample of participants completed the action readiness scales immediately postintervention, in addition to the pre- and posttest. To compare action readiness results across three points for this subsample, we performed linear mixed-effects modeling analyses that included time point with three levels (pretest = -1; immediately postintervention = 0; 2-weeks posttest = 1) and gender (female = 0; male = 1) as fixed effects and participant and summer institute as random effects. To account for the possibility that the error terms within each subject were correlated, we also specified a first-order autoregressive covariance matrix for the residuals.

Consistent with prediction 3, immediate postintervention action readiness results for this subsample mirrored the 2-week posttest results obtained with the full sample, in that the main effect of time for avoidance orientation was not significant,  $F(2, 136.10) = 0.73$ ,  $p = 0.48$ , but the main effect of time for approach orientation was significant,  $F(2, 138.46) = 5.91$ ,  $p = 0.003$  (see Figure 3).<sup>5</sup> Pairwise comparisons indicated that participants' approach orientation ("promotion focus") not only increased directly after the diversity training ( $M_{\Delta t1-12} = -0.20$ ,  $p = 0.03$ ), but that, as discussed in *Results* above, this effect remained 2 weeks after the training ( $M_{\Delta t2-13} = -0.08$ ,  $p = 0.40$ ;  $M_{\Delta t1-13} = -0.28$ ,  $p = 0.001$ ).

### DISCUSSION

This research addressed a critical gap in the existing literature by testing the efficacy of a gender bias intervention targeting STEM instructors. Although calls for STEM diversity interventions have increased (Al-Gazali *et al.*, 2013; *Nature*, 2015), there are few existing interventions available for implementation, and of these interventions, almost none have been subject



**FIGURE 3.** Means for action-readiness variables measured at pretest (2 weeks before the “Scientific Diversity” workshop), immediately following the intervention, and posttest (2 weeks after the “Scientific Diversity” workshop). Scales range from 1 to 7, with higher numbers reflecting a greater extent of each variable (a truncated scale range is presented for ease of interpretation). After the intervention, approach orientation (promotion focus) increased significantly ( $p = 0.003$ ), and remained constant 2 weeks later, while avoidance orientation (prevention focus) did not change after the intervention ( $p = 0.48$ ), and was also unchanged 2 weeks later.  $N = 78$ .

to systematic evaluation (Moss-Racusin *et al.*, 2014). Our results reflect promising evidence suggesting that a scientific approach to the design, assessment, and implementation of diversity interventions can be linked to positive outcomes.

Specifically, these findings have critical implications for STEM educators, in that they suggest that life scientists' subtle or unintentional gender biases may be mitigated by effective interventions. The evidence-based “Scientific Diversity” workshop was associated with increased awareness of diversity issues (supporting prediction 1), reduced subtle gender bias (supporting prediction 2), and heightened propensity to take action to address diversity challenges (supporting prediction 3) on the part of life science instructors. Because these participants regularly interact with the next generation of scientists, they reflect an important group to target with diversity interventions.

While improvements in awareness of diversity issues and reduced gender bias are heartening outcomes, they are by themselves insufficient. Effective diversity interventions must also increase participants' readiness to engage in behaviors that promote gender parity. Thus, the current action readiness results are particularly salient: the “Scientific Diversity” workshop was associated with increased approach orientation (i.e., promotion focus) and was not related to heightened avoidance orientation (i.e., prevention focus). That is, participants exhibited optimized action readiness following the intervention, such that they were more focused on engaging with their diversity goals and were not more reactive or avoidant.

This pattern of results is promising, in that heightened approach orientation (rather than reduced avoidance orientation) has been identified as a critical predictor of positive outcomes across a wide variety of domains (Crowe and Higgins, 1997; Higgins *et al.*, 1997; Moss-Racusin and Rudman, 2010; Rudman *et al.*, 2012), including increased support for social equality (Does *et al.*, 2011) and preserved executive functioning after interracial interactions (Trawalter and Richeson, 2006). As

<sup>5</sup>The estimated variance for each repeated promotion measure was significant,  $\sigma^2 = 0.27$ ,  $SE = 0.04$ , Wald  $Z = 7.44$ ,  $p < 0.001$ ; The estimated correlation between consecutive promotion measurements was not significant,  $\rho = -0.14$ ,  $SE = 0.15$ , Wald  $Z = -0.92$ ,  $p = 0.357$ . The estimated variance for each repeated prevention measure was significant,  $\sigma^2 = 0.37$ ,  $SE = 0.05$ , Wald  $Z = 7.49$ ,  $p < 0.001$ ; the estimated correlation between consecutive promotion measurements was not significant,  $\rho = -0.11$ ,  $SE = 0.15$ , Wald  $Z = -0.75$ ,  $p = 0.454$ .

a result, approach orientation was the primary action readiness variable of interest in the current research. These action readiness results highlight the potential benefits of promoting parity by working *toward* positive diversity goals, rather than warning *against* the possibility of diversity setbacks, and further underscore the importance of actively engaging attendees in diversity goals rather than blaming them for existing challenges.

Positive changes in all variables were observed during the monthlong period from pretest to posttest (as well as immediately following the intervention for action readiness), highlighting the potential of the “Scientific Diversity” workshop to initiate long-term progress. Although the effects of time point may appear modest, it is important to note that they are statistically significant (where noted and predicted) after adjusting for all NASI site-specific differences. That is, these changes in participants’ diversity awareness, subtle bias, and action readiness were observed above and beyond any differences associated with each specific NASI (such as the different individuals leading the workshop, the various times over the summer when each workshop took place, the different people—and unique personalities—at each NASI, varying resources and facilities). Thus, the predicted results are robust to these site-specific fluctuations.

Consistent with some prior work (Swim *et al.*, 1995, 2005), results revealed that male participants expressed higher levels of subtle gender bias than did female participants at pretest. In contrast, there was no significant sex difference at posttest, raising the possibility that the “Scientific Diversity” workshop may have been particularly effective at targeting male participants’ gender biases. We caution against strong interpretation of these results without further replication, in part because the interaction of participant gender and time was not significant. Additionally, the majority of existing work has not revealed gender differences in participants’ gender biases (Moss-Racusin *et al.*, 2012; Milkman *et al.*, 2015). At a minimum, the current results suggest that additional work is needed to shed light on the ways in which diversity interventions can be tailored to most effectively address the various needs and perspectives of attendees from different demographic backgrounds.

### LIMITATIONS AND FUTURE DIRECTIONS

Both the findings and limitations of this work highlight important avenues for future research. The present study is not an RCT and thus cannot generate definitive information about causal relationships. That is, because participants were not randomly assigned to an intervention or control condition, we cannot conclusively determine that the positive effects observed on our outcome variables were due to the intervention itself. These positive changes could be attributable to other factors that occurred during the month between measurements (e.g., engaging in cooperative group work at the NASI, attending one of the other NASI workshops) or the mere passage of time itself. However, the confluence of positive results across our three different outcomes is promising, as is the fact that the immediate postintervention results for action readiness mirrored those obtained 2 weeks later at posttest. Additionally, this type of correlational evidence is necessary to provide preliminary support for the efficacy of an intervention, laying the groundwork for subsequent complex (and expensive) RCTs (Moss-Racusin *et al.*, 2014). Nonetheless, future research should expand upon these results by conducting an RCT to determine whether the

positive changes observed here are directly attributable to the “Scientific Diversity” workshop.

Future research should also assess the extent to which these results generalize to other participant populations. Participation in the current study was voluntary, and results may thus be subject to some self-selection biases. We sought to mitigate these concerns by disguising the true nature of the research to avoid the possibility that only individuals who were particularly interested in gender bias might choose to participate in the study. Additionally, the current sample was not necessarily representative of the underlying population of life science instructors. For example, women were overrepresented in the current sample, reflective of the demographic characteristics of NASI attendees. Future research should thus utilize a randomly selected, representative participant sample to assess the extent to which the current results generalize to other groups. Additionally, the current research relied on a single-item measure of awareness of diversity issues. Specifically, we assessed the extent to which participants were aware of the relative demographic homogeneity of their home departments. Future work should utilize other multi-item measures of this construct to assess the replicability of the current results.

In keeping with other studies evaluating diversity interventions (Devine *et al.*, 2012; Prime *et al.*, 2012), the current analyses did not include data from participants of color. Because people of color often respond differently than white people on measures pertaining to diversity issues (e.g., Purdie-Vaughns *et al.*, 2008), it is essential to statistically account for participant race in this type of research. The low numbers of participants of color in our sample did not afford sufficient statistical power to conduct these analyses. Thus, to avoid distorting results, we limited the current analyses to white participants (although, as noted in footnote 1, results did not change when participants of color were included in analyses). Future research should include a plan to recruit racially diverse participant samples in order to robustly examine any potential racial differences in the efficacy of diversity interventions. Moreover, moving beyond gender biases, it is essential to address potential racial, ethnic, and other types of biases in STEM fields. These biases should be the sole focus of future research and interventions.

The current research assessed the extent to which participants were poised to engage in behaviors designed to address gender bias by assessing their approach orientation (i.e., promotion focus) and avoidance orientation (i.e., prevention focus). It is worth noting that promotion and prevention goal-pursuit strategies are independent of actual goal content (Crowe and Higgins, 1997). For example, two people could pursue different diversity goals (e.g., making sure not to spend extra time with male mentees vs. fairly evaluating the competence of female students) while both employing a promotion focus. Of importance, prior work has suggested that promotion focus is associated with improved outcomes regardless of specific goal content (Higgins *et al.*, 1997; Moss-Racusin and Rudman, 2010; Rudman *et al.*, 2012). That is, it appears that goal strategy (promotion or prevention) is the critical construct relevant for predicting behavioral outcomes, rather than the precise nature of the goal itself.

Thus, in keeping with previous research, we chose to focus on participants’ levels of promotion and prevention foci regarding the pursuit of their diversity goals, rather than measuring

participants' idiosyncratic definitions of personal diversity-related behavioral goals. One potential shortcoming associated with this approach is that it does not rule out the possibility that participants developed somewhat neutral or weak diversity-related goals (e.g., letting issues resolve themselves) but pursued these goals using a promotion focus. In this case, participants would report high levels of promotion focus but may not actually be working vigorously to address diversity issues. However, because of the active, approach-oriented nature of a promotion focus, it is somewhat difficult to pursue neutral goals using this strategy; indeed, it is challenging to envision how one might actively pursue a goal that is, by definition, inactive. Future research should address this concern by including a measurement of the specific diversity-related goals identified by participants, as well as their promotion and prevention focus.

Finally, our results may be subject to some social desirability concerns. That is, if participants guessed the true purpose of the study, they may have attempted to provide socially desirable responses. Thus, results could reflect this desire to respond in a "politically correct" manner rather than genuine improvements in outcomes. This concern is mitigated by at least two factors. First, participants did not report guessing the true purpose of the study during the debriefing. Second, a large body of literature suggests that biases stem from a two-step process, involving both *activation* and *expression* (e.g., Bodenhausen and Macrae, 1998). Thus, even if the current intervention did not undermine the activation of biases, the fact that participants were aware of and motivated to express the socially desirable response suggests that the intervention may have positively affected the expression of gender biases. It may not be necessary (or possible) for diversity interventions to eradicate all traces of participants' biases. Instead, successful interventions could operate efficiently on participants' motivations and abilities not to act on biased responses even if they are activated. Again, future research utilizing an RCT design should address these issues and should include additional outcomes that are less subject to possible social desirability concerns (e.g., implicit and behavioral measures).

Relatedly, the subgroup of participants who completed the immediate postintervention measures encountered an additional exposure to the action readiness scales relative to the rest of the sample. This increased familiarity could have heightened their ability to identify the true purpose of the study, and thus adjust their responses accordingly (i.e., demand characteristics). This concern is alleviated by the fact that this subgroup was no more likely than the rest of the sample to guess the true purpose of the study during the debriefing. Moreover, results drawn from this subsample did not significantly differ from other participants at posttest, suggesting that any additional demand characteristics were unlikely to have meaningfully impacted the responses of participants who completed the postintervention measures.

## CONCLUSIONS

Lingering gender biases can undermine the representation of women in STEM, ultimately restricting access to talented professionals (Moss-Racusin *et al.*, 2012). By clearing obstacles facing talented female scientists, effective diversity interventions thus hold the potential not only to broaden women's participation but also to enhance the quality and pace of scientific

progress. Despite calls for their widespread implementation in the STEM community, few evidence-based, validated diversity interventions are presently available. The current research revealed promising evidence of the efficacy of one such intervention. We hope that these results will help to generate additional research investigating scientific approaches to the development, evaluation, and implementation of effective diversity interventions.

## ACKNOWLEDGMENTS

We are grateful to the 2012 NASI leaders and "Scientific Diversity" workshop instructors for their partnership and invaluable assistance with data collection and for the participation of NASI attendees. Key personnel who designed, revised, and facilitated the "Scientific Diversity" workshop include J. Frederick, J. Labov, R. Lue, S. Miller, C. Pfund, and J. Stith. We thank M. de Rooij for statistical consultation. This work was supported in part by Howard Hughes Medical Institute (HHMI) Professors Program 2010 grant 52007053 and HHMI grant 52007129 to J.H., Alfred P. Sloan Foundation grant 213-3-15 to J.H. and C.A.M.-R., and Alfred P. Sloan Foundation grant B2013-38 to C.A.M.-R.

## REFERENCES

- Al-Gazali L, Valian V, Barnes B, Wu LA, Andrei EY, Handelsman J, Moss-Racusin CA, Hsu L (2013). Scientists of the world speak up for equality. *Nature* 495, 35–38.
- Bargh JA, Chartrand TL (2000). The mind in the middle: a practical guide to priming and automaticity research. In: *Handbook of Research Methods in Social and Personality Psychology*, ed. HT Reis and CM Judd, New York: Cambridge University Press, 253–285.
- Bever L (2015). Nobel Prize-winning scientist Tim Hunt resigns after commenting on the "trouble with girls." *Washington Post*. [www.washingtonpost.com/news/morning-mix/wp/2015/06/11/nobel-prize-winning-scientist-tim-hunt-resigns-position-after-commenting-on-the-trouble-with-girls/](http://www.washingtonpost.com/news/morning-mix/wp/2015/06/11/nobel-prize-winning-scientist-tim-hunt-resigns-position-after-commenting-on-the-trouble-with-girls/) (accessed 23 July 2015).
- Bodenhausen GV, Macrae CN (1998). Stereotype activation and inhibition. In: *Stereotype Activation and Inhibition: Advances in Social Cognition*, ed. RS Wyer, Mahwah, NJ: Erlbaum, 1–52.
- Carnes M, Devine PG, Manwell LB, Byars-Winston A, Fine E, Ford CE, Forscher P, Isaac C, Kaatz A, Magua W, *et al.* (2015). The effect of an intervention to break the gender bias habit for faculty at one institution: a cluster randomized, controlled trial. *Acad Med* 90, 221–230.
- Carnes M, Devine PG, Manwell LB, Ford CE, Byars-Winston A, Fine E, Sheridan JT (2012). Promoting institutional change through bias literacy. *J Divers High Educ* 5, 63–77.
- Case KA (2007). Raising male privilege awareness and reducing sexism: an evaluation of diversity courses. *Psychol Women Q* 31, 426–435.
- Ceci SJ, Ginther DK, Kahn S, Williams WM (2014). Women in academic science: a changing landscape. *Psychol Sci Pub Int* 15, 75–141.
- Chamany K, Allen D, Tanner K (2008). Making biology learning relevant to students: integrating people, history, and context into college biology teaching. *CBE Life Sci Educ* 7, 267–278.
- Couch BA, Brown TL, Schelpat TJ, Graham MJ, Knight JK (2015). Scientific teaching: defining a taxonomy of observable practices. *CBE Life Sci Educ* 14, ar9.
- Crowe E, Higgins ET (1997). Regulatory focus and strategic inclinations: promotion and prevention in decision-making. *Organ Behav Hum* 69, 117–132.
- Dasgupta D (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: the stereotype inoculation model. *Psychol Inq* 22, 231–246.
- Dasgupta N, Scircle MM, Hunsinger M (2015). Female peers in small work groups enhance women's motivation, verbal participation, and career aspirations in engineering. *Proc Natl Acad Sci USA* 112, 4988–4993.

- Devine PG (1989). Stereotypes and prejudice: their automatic and controlled components. *J Pers Soc Psychol* 56, 5–18.
- Devine PG, Forscher PS, Austin AJ, Cox WTL (2012). Long-term reduction in implicit race bias: a prejudice habit-breaking intervention. *J Exp Soc Psychol* 48, 1267–1278.
- Dobbin F, Kalev A (2013). The origins and effects of corporate diversity programs. In: *The Oxford Handbook of Diversity and Work*, ed. QM Robertson, Oxford, UK: Oxford University Press, 253–281.
- Dobbin F, Schrage D, Kalev A (2015). Rage against the iron cage, the varied effects of bureaucratic personnel reforms on diversity. *Am Sociol Rev* 80, 1014–1044.
- Does S, Derks B, Ellemers N (2011). Thou shalt not discriminate: how emphasizing moral ideals rather than obligations increases whites' support for social equality. *J Exp Soc Psychol* 47, 562–571.
- Dovidio JF, Gaertner SL (2004). Aversive racism. In: *Advances in Experimental Social Psychology*, ed. MP Zanna, New York: Elsevier, 1–51.
- Eddy SL, Brownell SE, Wenderoth MP (2014). Gender gaps in achievement and participation in multiple introductory biology classrooms. *CBE Life Sci Educ* 13, 478–492.
- Finson KD, Beaver JB, Cramond BL (1995). Development and field test of a checklist for the draw-a-scientist test. *School Sci Math* 95, 195–205.
- Glick P, Fiske ST (1996). The ambivalent sexism inventory: differentiating hostile and benevolent sexism. *J Pers Soc Psychol* 70, 491–512.
- Greenwald AG, Banaji MR, Nosek BA (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *J Pers Soc Psychol* 108, 553–561.
- Handelsman J, Ebert-May D, Beichner R, Bruns P, Chang A, DeHaan R, Gentile J, Lauffer S, Stewart J, Tilghman SM, et al. (2004). Scientific teaching. *Science* 304, 521–522.
- Handelsman J, Miller S, Pfund C (2007). *Scientific Teaching*, New York: Freeman.
- Hawkins SA, Hastie R (1990). Hindsight: biased judgments of past events after the outcomes are known. *Psychol Bull* 107, 311–327.
- Higgins ET, Shah J, Friedman R (1997). Emotional responses to goal attainment: strength of regulatory focus as a moderator. *J Pers Soc Psychol* 72, 515–525.
- Jackson SM, Hillard AL, Schneider TR (2014). Using implicit bias training to improve attitudes toward women in STEM. *Soc Psychol Educ* 17, 419–438.
- Kenny DA, Judd CM (1986). Consequences of violating the independence assumption in analysis of variance. *Psychol Bull* 99, 422–431.
- Kenny DA, Judd CM (1996). A general procedure for the estimation of interdependence. *Psychol Bull* 119, 138–148.
- Lai CK, Marini M, Lehr SA, Cerruti C, Shin JEL, Joy-Gaba JA, Ho AK, Teachman BA, Wojcik SP, Koleva SP, et al. (2014). Reducing implicit racial preferences. I. A comparative investigation of 17 interventions. *J Exp Psychol Gen* 143, 1765–1785.
- Lariviere V, Ni C, Gingras Y, Cronin B, Sugimoto CR (2013). Bibliometrics: global gender disparities in science. *Nature* 504, 211–213.
- Legault L, Gutsell JN, Inzlicht M (2011). Ironic effects of antiprejudice messages: how motivational interventions can reduce (but also increase) prejudice. *Psychol Sci* 22, 1472–1477.
- Milkman KL, Akinola M, Chugh D (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *J Appl Psychol* 100, 1678–1712.
- Miller S, Pfund C, Pribbenow CM, Handelsman J (2008). Scientific teaching in practice. *Science* 322, 1329–1330.
- Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham JM, Handelsman J (2012). Science faculty's subtle gender biases favor male students. *Proc Natl Acad Sci USA* 109, 16474–16479.
- Moss-Racusin CA, Rudman LA (2010). Disruptions in women's self-promotion: the backlash avoidance model. *Psychol Women Quart* 34, 186–202.
- Moss-Racusin CA, van der Toorn J, Dovidio JF, Brescoll VL, Graham MJ, Handelsman J (2014). Scientific diversity interventions. *Science* 343, 615–616.
- National Science Foundation (NSF) (2008). Thirty-Three Years of Women in S&E Faculty Positions. [www.nsf.gov/statistics/infbrief/nsf08308](http://www.nsf.gov/statistics/infbrief/nsf08308) (accessed 23 July 2015).
- NSF (2012). Women, Minorities, and Persons with Disabilities in Science and Engineering. [www.nsf.gov/statistics/2015/nsf15311/digest/theme2.cfm](http://www.nsf.gov/statistics/2015/nsf15311/digest/theme2.cfm) (accessed 23 July 2015).
- Nature (2015). Sexism has no place in science. *Nature* 522, 255.
- Nobel Foundation (2015). Facts on the Nobel Prize in Physiology or Medicine. [www.nobelprize.org/nobel\\_prizes/facts/medicine](http://www.nobelprize.org/nobel_prizes/facts/medicine) (accessed 23 July 2015).
- Nosek BA, Banaji MR, Greenwald AG (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dyn Theor Res* 6, 101–115.
- Paluck EL (2006). Diversity training and intergroup contact: a call to action research. *J Soc Issues* 62, 577–595.
- Paluck EL, Green DP (2009). Prejudice reduction: what works? A review and assessment of research and practice. *Annu Rev Psychol* 60, 339–367.
- Pfund C, Miller S, Brenner K, Bruns P, Chang A, Ebert-May D, Fagen AP, Gentile J, Gossens S, Khan IM, et al. (2009). Summer institute to improve university science teaching. *Science* 324, 470–471.
- Pfund C, Spencer KC, Asquith P, House SC, Miller S, Sorkness CA (2015). Building national capacity for research mentor training: an evidence-based approach to training the trainers. *CBE Life Sci Educ* 14, ar24.
- Prime J, Foust-Cummings H, Salib ER, Moss-Racusin CA (2012). Calling All White Men: Can Training Help Create Inclusive Workplaces? New York: Catalyst.
- Purdie-Vaughns V, Steele CM, Davies PG, Dittmann R, Crosby JR (2008). Social identity contingencies: how diversity cues signal threat or safety for African Americans in mainstream institutions. *J Pers Soc Psychol* 94, 615–630.
- Reuben E, Sapienza P, Zingales L (2014). How stereotypes impair women's careers in science. *Proc Natl Acad Sci USA* 111, 4403–4408.
- Rudman LA, Moss-Racusin CA, Glick P, Phelan JE (2012). Reactions to vanguards: advances in backlash theory. In: *Advances in Experimental Social Psychology*, ed. PG Devine and EA Plant, New York: Elsevier, 167–227.
- Sheltzer JM, Smith JC (2014). Elite male faculty in the life sciences employ fewer women. *Proc Natl Acad Sci USA* 111, 10107–10221.
- Shields SA, Zawadzki MJ, Johnson RN (2011). The impact of a workshop activity for gender equity simulation in the academy (WAGES-Academic) in demonstrating cumulative effects of gender bias. *J Divers High Educ* 4, 120–129.
- Steele CM, Aronson J (1995). Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol* 69, 797–811.
- Steinpreis RE, Anders KA, Ritzke D (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: a national empirical study. *Sex Roles* 41, 509–528.
- Stout JG, Dasgupta N, Hunsinger M, McManus MA (2011). STEMing the tide: using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *J Pers Soc Psychol* 100, 255–270.
- Swim JK, Aikin K, Hall WS, Hunter BA (1995). Sexism and racism: old-fashioned and modern prejudices. *J Pers Soc Psychol* 68, 199–214.
- Swim JK, Malett R, Russo-Devosa Y, Stangor C (2005). Judgments of sexism: a comparison of the subtlety of sexism measures and sources of variability in judgments of sexism. *Psychol Women Q* 29, 406–411.
- Thomason A (2015). Nobel-winning biochemist apologizes for remarks about "trouble with girls." *Chronicle of Higher Education*. <http://chronicle.com/blogs/ticker/jp/nobel-prize-winning-biochemist-apologizes-for-remarks-about-trouble-with-girls> (accessed 23 July 2015).
- Trawalter S, Richeson JA (2006). Regulatory focus and executive function after interracial interactions. *J Exp Soc Psychol* 42, 406–412.
- Uhlmann EL, Cohen GL (2007). "I think it, therefore it's true": effects of self-perceived objectivity on hiring discrimination. *Organ Behav Hum Dec* 104, 207–223.
- Van der Lee R, Ellemers N (2015). Gender contributes to personal research funding success in the Netherlands. *Proc Natl Acad Sci USA* 112, 12349–12353.
- Wood WB, Handelsman J (2004). Meeting report: the 2004 National Academies summer institute on undergraduate education in biology. *Cell Biol Educ* 3, 215–217.
- Zawadzki MJ, Danube CL, Shields SA (2012). How to talk about gender inequity in the workplace: using WAGES as an experiential learning tool to reduce reactance and promote self-efficacy. *Sex Roles* 67, 605–616.