A gender bias habit-breaking intervention led to increased hiring of female faculty in STEMM departments⁎,⁎⁎

Patricia G. Devinea⁎, Patrick S. Forscherb, William T.L. Coxa, Anna Kaatzc, Jennifer Sheridan,d, Molly Carnesc

⁎ Department of Psychology, University of Wisconsin-Madison, United States
b Department of Psychology, University of Arkansas, United States
c Center for Women’s Health Research, University of Wisconsin-Madison, United States
d Women in Science and Engineering Leadership Institute, University of Wisconsin-Madison, United States

ABSTRACT

Addressing the underrepresentation of women in science is a top priority for many institutions, but the majority of efforts to increase representation of women are neither evidence-based nor rigorously assessed. One exception is the gender bias habit-breaking intervention (Carnes et al., 2015), which, in a cluster-randomized trial involving all but two departmental clusters (N = 92) in the 6 STEMM focused schools/colleges at the University of Wisconsin-Madison, led to increases in gender bias awareness and self-efficacy to promote gender equity in academic science departments and perceptions of a more positive departmental climate. Following this initial success, the present study compares, in a preregistered analysis, hiring rates of new female faculty pre- and post-intervention. Whereas the proportion of women hired by control departments remained stable over time, the proportion of women hired by intervention departments increased by an estimated 18 percentage points (OR = 2.23, dOR = 0.34). Though the preregistered analysis did not achieve conventional levels of statistical significance (p < 0.07), the study has a hard upper limit on statistical power, as the cluster-randomized trial has a maximum sample size of 92 departmental clusters. These findings, however, have undeniable practical significance for the advancement of women in science, and provide promising evidence that psychological interventions can facilitate gender equity and diversity.

Women remain underrepresented in doctoral-level careers in science, technology, engineering, math, and medical (STEMM) fields (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012; NSF, 2007). This gender inequity, paired with concurrent under-representation of racial minorities, has led numerous organizations to call for efforts to increase participation of women and minorities in STEMM (e.g., NSF, 2014; National Academy of Sciences, National Academy of Engineering, Institute of Medicine of the National Academies, 2006; NIH: Valantine & Collins, 2015; see also Corrce, 2009; Hill, Corbett, & St. Rose, 2010; Mitchell, Smith, & Latimer, 2016; Sevo & Chubin, 2008). Many existing efforts to address these issues, however, are neither evidence-based nor rigorously assessed in experimental trials (Moss-Racusin et al., 2014; Paluck & Green, 2009). When systematically assessed, these non-evidence-based efforts either do not work or make problems worse (Apfelbaum, Norton, & Sommers, 2012; Dobbin & Kalev, 2013; Legault, Gutsell, & Inzlicht, 2011).

Interventions designed to reduce intergroup biases should be rooted in well-supported theory about the nature of prejudice and bias reduction. One such theory is the prejudice habit model (Devine, 1989; Devine, Forscher, Austin, & Cox, 2012), which conceptualizes bias as a mental habit and lays out the steps needed to “break the bias habit.” Specifically, once a person is motivated to act in less biased ways, breaking the bias habit involves 1) becoming aware of when one is vulnerable to unintentional bias, 2) understanding the consequences of unintentional bias, and 3) learning and practicing strategies to reduce the impact of unintentional bias.

Devine et al. (2012) operationalized the components of the habit-breaking model into the prejudice habit-breaking intervention, which is...
thus far the only intervention experimentally shown to produce long-term changes in bias (Devine et al., 2012), with effects lasting at least 2 years post-manipulation (Forscher et al., 2017). One iteration of this intervention approach is the gender bias habit-breaking intervention (Carnes et al., 2015), which focused specifically on gender bias in STEMM fields and was implemented in a 2.5 h workshop to individual departments.

The workshop (see Fig. 1 and Carnes et al., 2012) reviews the key components of the habit model (awareness, consequences, and strategies). To increase awareness, prior to the workshop participants completed and received feedback on a gender/leadership Implicit Association Test (IAT). The workshop opened with evidence of continuing gender bias in STEMM, including the underrepresentation of women in faculty and leadership positions and the potential adverse impact such biases for the overarching goals of advancing science, national health, and economic vitality. Attendees learned how unintentional bias function like habits, leading people to often respond in ways that contradict egalitarian values. They then learned about six “bias constructs” that represent common manifestations of gender bias generally and in STEMM more specifically (i.e., expectancy bias, prescriptive gender norms, role congruity/incongruity, stereotype priming, reconstructing credentials, and stereotype threat). To allow attendees to actively engage with the constructs and foster learning of new material, attendees next read and discussed case studies to practice identifying and examining the bias-promoting impact of the constructs. To promote efficacy to reduce bias, attendees learned five evidence-based strategies (i.e., stereotype replacement, counterstereotypic imaging, individualization, perspective taking, and increasing opportunities for intergroup contact) that have been shown to counteract unintentional bias (Devine et al., 2012); attendees were told that practicing the strategies would help them to break the gender bias habit. Attendees also wrote statements of commitment to action to address gender bias and their written commitment to address gender bias should reduce the effects of unintentional bias on hiring, yielding more new women faculty hires. Fourth, hiring decisions are made by departments, not individuals, which is well-matched to the cluster-randomized design, in which departments were assigned to receive the intervention or serve as controls.1 In prior tests of the impact of the habit-breaking intervention, outcomes were assessed at the individual level even when evaluated as the cluster level (Carnes et al., 2015; Devine et al., 2012; Forscher et al., 2017). In the present context, we explore the potential for the intervention to affect individuals in ways that may promote change in institutional level outcomes. Finally, to our knowledge, no past work has investigated the impact of a real-world intergroup bias intervention on this type of highly consequential outcome. We anticipated that, compared to control departments and intervention departments in the pre-manipulation period, only intervention departments in the post-intervention period would show greater gender balance in their new hires.

1. Method

The pre-registered analytic plan, dataset, and supplemental analyses are available at https://osf.io/9yr23/. All measures, manipulations, and exclusions are disclosed here and in Carnes et al. (2015). At the study...
P.G. Devine et al.  

made 101 hires (47% women).2 To protect against potential spurious

defects related to under-represented minorities (URMs), which are also reported in Table 1. The overall number of

URMs was small, and there was no evidence of changes in hiring or attrition of URMs.

2The overall number of hires by cluster in the pre-intervention period was marginally
different, \( \chi^2(1, n = 92) = 2.95, p = 0.09 \), but intervention and control departments

hired approximately the same number of new faculty in the post-intervention period, \( \chi^2(1, n = 92) = 0.67, p = 0.41 \). Although the control departments hired somewhat more

faculty overall in the pre-manipulation period, the greater number of hires did not yield a

greater gender balance in hiring. Because our analyses either operate on proportions or

are weighted by the total number of hires, we do not think this pre-manipulation dif-

ference affects the interpretation of the hiring proportions.

in 2015, it is reasonable that men in intervention departments may have

been less likely to leave. We tested the “revolving door” account by separately estimating the

change in the numbers, rather than the proportion, of hires and

attrition for men and women using GLMEMs with log links from the

Poisson family. As shown in Table 1, the increase in intervention
departments’ proportions of female attrition was driven by decreases in the

number of men who left intervention departments relative to control
departments, \( RR = 0.64, \chi^2(1, n = 89) = 3.62, p = 0.06, 95% CI = [0.40, 1.01], \) not increases in the number of women who left, \( RR = 1.68, \chi^2(1, n = 89) = 1.28, p = 0.26, 95% CI = [0.68, 4.18], \). Given

previously-reported patterns showing the gender habit-breaking

intervention improved climate for both women and men (Carnes et al.,

2015), it is reasonable that men in intervention departments may have

been less likely to leave.

2.3. Exploratory “revolving door” analyses

Taken together, the hiring and attrition patterns raise the possibility that any apparent increase in new women hires may reflect a “revolving door” whereby female faculty leave and departments merely replace

them. We tested the “revolving door” account by separately estimating the

change in the numbers, rather than the proportion, of hires and

attrition for men and women using GLMEMs with log links from the

Poisson family. As shown in Table 1, the increase in intervention
departments’ proportions of female attrition was driven by decreases in the

number of men who left intervention departments relative to control
departments, \( RR = 0.64, \chi^2(1, n = 89) = 3.62, p = 0.06, 95% CI = [0.40, 1.01], \) not increases in the number of women who left, \( RR = 1.68, \chi^2(1, n = 89) = 1.28, p = 0.26, 95% CI = [0.68, 4.18], \). Given

previously-reported patterns showing the gender habit-breaking

intervention improved climate for both women and men (Carnes et al.,

2015), it is reasonable that men in intervention departments may have

been less likely to leave.

3. Discussion

Our findings are promising with regard to improving the representa-
tion of women in STEM disciplines. According to our pre-
registered GLMEMs estimate, intervention departments hired 18 per-
centage points more women in the post- than pre-intervention period.

Control departments did not vary in their hiring of women over time.

Pre-intervention, hires in control and intervention departments sub-
stantially favored men, but after the manipulation, new hires in inter-

vention departments were gender balanced. This gender-balanced

hiring is what one would expect if there are equal numbers of qualified

men and women applicants. Though increased hiring of women did not

achieve conventional levels of statistical significance, our study has a

hard upper limit on statistical power. We hasten to add, however, that

statistical certainty is only one criterion against which to judge the
Table 1

<table>
<thead>
<tr>
<th>Estimates</th>
<th>Time × condition</th>
<th>Control vs. intervention</th>
<th>Pre vs. post</th>
<th>Control</th>
<th>Intervention</th>
<th>Post</th>
<th>Control</th>
<th>Intervention</th>
<th>Post</th>
<th>Control</th>
<th>Intervention</th>
<th>Post</th>
<th>Control</th>
<th>Intervention</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proportion of women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hires</td>
<td>0.34</td>
<td>0.29</td>
<td>0.47</td>
<td>0.22</td>
<td>2.33</td>
<td>1.02</td>
<td>1.02</td>
<td>1.03</td>
<td>0.37</td>
<td>0.95</td>
<td>0.50</td>
<td>0.67</td>
<td>0.17</td>
<td>0.81</td>
<td>0.31</td>
</tr>
<tr>
<td>Attrition</td>
<td>0.20</td>
<td>0.19</td>
<td>0.27</td>
<td>0.27</td>
<td>3.03</td>
<td>0.33</td>
<td>0.53</td>
<td>0.22</td>
<td>0.04</td>
<td>0.83</td>
<td>0.71</td>
<td>0.67</td>
<td>0.18</td>
<td>0.83</td>
<td>0.17</td>
</tr>
<tr>
<td>Faculty</td>
<td>0.08</td>
<td>0.06</td>
<td>0.10</td>
<td>0.10</td>
<td>1.96</td>
<td>0.76</td>
<td>0.93</td>
<td>0.71</td>
<td>0.08</td>
<td>0.93</td>
<td>0.71</td>
<td>0.67</td>
<td>0.17</td>
<td>0.83</td>
<td>0.17</td>
</tr>
<tr>
<td>Attrition</td>
<td>0.03</td>
<td>0.05</td>
<td>0.02</td>
<td>0.04</td>
<td>1.62</td>
<td>0.80</td>
<td>0.93</td>
<td>0.71</td>
<td>0.08</td>
<td>0.93</td>
<td>0.71</td>
<td>0.67</td>
<td>0.17</td>
<td>0.83</td>
<td>0.17</td>
</tr>
<tr>
<td>Numbers of women</td>
<td>0.69</td>
<td>0.65</td>
<td>0.53</td>
<td>0.91</td>
<td>0.80</td>
<td>1.80</td>
<td>0.93</td>
<td>0.71</td>
<td>0.08</td>
<td>0.93</td>
<td>0.71</td>
<td>0.67</td>
<td>0.17</td>
<td>0.83</td>
<td>0.17</td>
</tr>
<tr>
<td>Faculty</td>
<td>0.29</td>
<td>0.21</td>
<td>0.27</td>
<td>0.27</td>
<td>1.03</td>
<td>3.03</td>
<td>0.53</td>
<td>0.22</td>
<td>0.04</td>
<td>0.83</td>
<td>0.71</td>
<td>0.67</td>
<td>0.18</td>
<td>0.83</td>
<td>0.17</td>
</tr>
<tr>
<td>Attrition</td>
<td>0.19</td>
<td>0.18</td>
<td>0.20</td>
<td>0.20</td>
<td>1.00</td>
<td>0.94</td>
<td>0.94</td>
<td>0.71</td>
<td>0.08</td>
<td>0.93</td>
<td>0.71</td>
<td>0.67</td>
<td>0.17</td>
<td>0.83</td>
<td>0.17</td>
</tr>
<tr>
<td>Numbers of men</td>
<td>1.03</td>
<td>0.93</td>
<td>0.51</td>
<td>0.92</td>
<td>0.80</td>
<td>1.80</td>
<td>0.93</td>
<td>0.71</td>
<td>0.08</td>
<td>0.93</td>
<td>0.71</td>
<td>0.67</td>
<td>0.17</td>
<td>0.83</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Importance of findings (Ross & Nisbett, 2011). Ross and Nisbett also highlight the importance of pragmatic criteria, and we contend that the shift observed in hiring of women has undeniable practical significance for the long-term goal of achieving gender equity in STEM.

We can only speculate about the processes that may have produced gender-balanced hiring. It is possible, for example, that faculty who attended the workshop became more concerned about gender discrimination (Forscher et al., 2017), which may have led them to be more active in hiring committees and more proactive in considering and advocating for female applicants (Bardi & Schwartz, 2003; Krosnick, 1988). The faculty may also have implemented strategies to circumvent their own gender biases (Devine et al., 2012) or identified and labeled common manifestations of gender bias in others, setting the stage for constructive conversations with colleagues about gender equity (Ashburn-Nardo, Morris, & Goodwin, 2008; Forscher et al., 2017; Mitamura, Erickson, & Devine, 2017; Nonaka, 1994). These processes may have caused intervention departments to seek out and make more offers to women. Alternatively, given that intervention departments appear to have better climates for women (and men), perhaps they were more effective at successfully recruiting women once an offer was extended. To the extent that these or other processes demonstrate an institutional commitment to the professional success of female faculty, they could have recursive and synergistic effects on future hiring and retention, with the potential to expand beyond institutional change.

Investigating these possibilities is a high priority for future work.

Though we are cautiously optimistic about our findings, we acknowledge that the marginal statistical significance of the hiring effect does not permit a high degree of certainty in our intervention’s influence on hiring. Moreover, hiring is but the first step on the long journey to achieving gender equity in STEM. As yet, there is no evidence that the intervention caused a change in the overall gender composition of experimental departments. Such a change would likely require hiring changes that endure much longer than two years. After hiring has occurred, gender equity can only truly be achieved if women thrive in their departments, achieve tenure, and ascend to leadership positions.

Our confidence in these findings, however, is bolstered by the strength of the cluster-randomized controlled experimental design and the long-term, longitudinal assessment of the intervention’s impact (Moss-Racusin et al., 2014). Though conducting comprehensive, longitudinal, theoretically-derived intervention work is an enormous undertaking, our study reveals the potential payoffs of such large-scale efforts. Moreover, we believe this type of approach is necessary to stem the tide of ineffective and sometimes harmful bias-reducing approaches based on intuition or wishful thinking. Translating psychological research into application in the form of evidence-based interventions is essential to fulfill the promise of psychological science as a force to improve people’s lives and society.

Author contributions

P. G. Devine and P. S. Forscher conceived the study concept. J. Sheridan retrieved the data from the University of Wisconsin-Madison Human Resources records. P. S. Forscher and W. T. L. Cox analyzed the data under the supervision of P. G. Devine. W. T. L. Cox, P. G. Devine, and P. S. Forscher planned and drafted the paper. M. Carnes was PI of NIH R01 GM088477 that funded the study on which the current research is based. All authors provided edits and approved the final version of the manuscript for submission.

Open practices

The experiment in this article earned Open Materials, Open Data, and Preregistered badges for transparent practices. Materials, data, and preregistration information for the experiment are available at https://osf.io/9yf23/.
References


