ARTICLE IN PRESS

Journal of Experimental Social Psychology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Case Report

The impact of an "aha" moment on gender biases: Limited evidence for the efficacy of a game intervention that challenges gender assumptions

Gili Freedman[a,*], Max Seidman[a], Mary Flanagan[a], Geoff Kaufman[b], Melanie C. Green[c]

[a] Dartmouth College, USA
[b] Carnegie Mellon University, USA
[c] University of Buffalo, USA

ARTICLE INFO

Handling editor: Elizabeth Page-Gould

Keywords:
Gender
Biases
STEM
Intervention
Game

ABSTRACT

In two studies, the present research tested whether a paper-based game intervention that guides participants into understanding and questioning their assumptions about gender can decrease biases. Participants in Study 1 ($N = 143$ college students) and Study 2 ($N = 341$ high school students) played a game in which they either had to realize that a scientist character was a woman (Intervention condition) or a professor (Control condition) to solve the mystery. Across both studies, in a game with a storyline that included both male and female scientists, the vast majority of students who used gendered pronouns assumed that non-gendered scientist characters were men. In Study 1, playing the Intervention version of the game had no effect on college students' explicit or implicit attitudes toward women in science. In Study 2, there was a positive effect of the Intervention condition on implicit attitudes: participants in the Intervention condition were less likely to describe a female professor as a man than were participants in the Control condition. However, there was a negative effect of the Intervention condition on explicit attitudes toward women in science. Taken together, the present research points to the continued need for research on raising awareness of bias and developing interventions that can decrease biases while avoiding defensiveness.

The impact of biases against women in science is clear: women are underrepresented in STEM (science, technology, engineering, and math) fields (National Science Foundation, 2017) and often experience hostile working environments (Beasley & Fischer, 2012; Clancy, Nelson, Rutherford, & Hinde, 2014; Reuben, Sapienza, Zingales, & Greenwald, 2014). But what happens when people are made aware that they may exhibit their own biased assumptions? When individuals are confronted with evidence of their biases, they sometimes respond defensively (Frantz, Cuddy, Burnett, Ray, & Hart, 2004; Hillard, Ryan, & Gervais, 2013; Howell et al., 2013; Howell, Gaither, & Ratliff, 2015; Howell & Ratliff, 2017). For example, when participants expect feedback from an Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998) to indicate biased intergroup preferences (i.e., negative feedback), they opt to not learn the results and regret learning them if they do find out (Howell et al., 2013). Even the knowledge that a test may potentially reveal biases can lead to defensiveness and, paradoxically, produce higher bias scores (Frantz et al., 2004). Furthermore, individuals are prone to interpret published research about stereotypes in a biased fashion if the research implicates their identity has one that engages in prejudicial acts (Handley, Brown, Moss-Racusin, & Smith, 2015).

There have been a few interventions that have effectively utilized confrontational strategies, particularly for confronting individuals about racial bias. These racial bias confrontations have been shown to reduce prejudiced attitudes and to induce negative self-directed emotions (Czopp & Monteith, 2003; Czopp, Monteith, & Mark, 2006; Gulker, Mark, & Monteith, 2013). The impact of confronting individuals about gender biases, however, has been mixed with participants feeling dismissive about gender bias related confrontations (Gulker et al., 2013). For example, although a series of confrontations about gender biases were successful in inducing negative self-directed emotions and increasing concern about being prejudiced in the future, they also increased participants' likelihood of responding defensively by devaluing women in science and thinking that the researchers who mentioned gender biases were overly sensitive (Parker, Monteith, Moss-Racusin, & Van Camp, 2018).

Thus, an important open question is whether there are ways to create interventions for combating gender biases that both reduce prejudicial thoughts and avoid increasing defensiveness. In other

* Corresponding author at: Dartmouth College, 22 Lebanon St, 246 Black Family VAC, Hanover, NH 03755, USA.
E-mail address: gili.freedman@gmail.com (G. Freedman).

https://doi.org/10.1016/j.jesp.2018.03.014
Received 7 July 2017; Received in revised form 26 February 2018; Accepted 30 March 2018
0022-1031/

words, how can evidence of personal biases be presented to individuals in a way to would invite greater receptivity and less rejection? One promising strategy has emerged in the domain of persuasion. Dispelling participants' illusion of invulnerability to illegitimate appeals (i.e., making people aware that they are vulnerable to persuasive attempts regardless of their legitimacy) reduces susceptibility to such appeals – but information alone is not enough. Individuals have to be fooled and have the mistake pointed out to them (Sagarin, Cialdini, Rice, & Serna, 2002).

The present work applies this strategy (i.e., producing an "aha" moment) to illuminate and dispel biased assumptions about gender in STEM domains. Despite attempts to increase the representation of women across STEM fields (Moss-Racusin et al., 2014), a large gender gap still exists (National Science Foundation, 2017), and this imbalance is perpetuated by stereotypes about the ability of women to excel in STEM domains (Hill, Corbett, & St. Rose, 2010). The present work employs a game-based intervention as a less-threatening context in which to confront biases. Games are a promising avenue for persuasive intervention because they can provide a less explicit—and thus, psychologically safer—means of dealing with difficult issues (Bessarabova et al., 2016; Dunbar et al., 2014), particularly if they are designed with the possibility of defensiveness in mind (Kaufman, Flanagan, & Seidman, 2015). Furthermore, interventions to combat biases are particularly effective when they involve active participation rather than passive learning, as demonstrated by the WAGES intervention—a game intervention in which participants learn about biases against women by attempting to become a Distinguished Professor (Shields, Zawadzki, & Johnson, 2011). Games by definition involve active participation and may therefore be a useful context for creating an intervention combating biases against women in science. The present research introduces a novel intervention in which participants play a logic-puzzle game that they can only win by realizing that one of the scientist characters is a woman. Across the two studies, the hypotheses were that playing the intervention version of the game compared to the control version of the game would increase positive attitudes toward women in science and reduce sexism. In Study 1, we also hypothesized that playing the intervention version compared to the control version would increase monetary allocations to women in STEM organizations. We report all measures, manipulations, and exclusions in these studies.

1. Study 1

Study 1 examined the impact of experiencing an "aha" moment about assumptions about women in science on subsequent attitudes toward women in science. Undergraduate students played a logic mystery game in which the solution hinged on recognizing and correcting gender assumptions.

1.1. Method

1.1.1. Participants

Based on the results of a power analysis with 80% power and an expected effect size of $d = 0.47$ (Sagarin et al., 2002), we aimed to recruit 146 participants from undergraduate housing communities. One hundred forty-four college students participated; one participant was excluded for knowing the study's purpose, leaving a final sample of 143 participants (77 men, 65 women, 1 did not report gender; $M_{age} = 20.29$, $SD_{age} = 1.04$).

1.1.2. Procedure

After providing consent, participants were randomly assigned to pairs and to conditions. In sessions in which there were an uneven number of participants, one group played in a group of three—there were a total of three groups of three and the rest of the groups were pairs. During development of the game, undergraduate students were asked to playtest the game to test its length and how enjoyable it was. No formal data were collected, but playtest observations revealed that the players who did the game on their own took longer and seemed less interested than players who were allowed to work in groups or pairs. Therefore, the research team made the decision to have participants in Study 1 and 2 play in pairs or groups. The study was conducted at the students' housing locations. Therefore, all of the students knew each other within each session. Participants were randomly handed an index card with an ID number on it. Based on the last two digits of the ID, participants found their pair. Thus, pairs were randomly assigned. Most of the housing locations were single sex, which resulted in 34 male pairs/groups, 29 female pairs/groups, and 7 mixed pairs/groups.

In the study, participants were told that they had 35 min to solve a mystery in which a dangerous disease sample went missing from a lab. Participants were given maps of the building, logs of who entered and left each room, and notes from the FBI director (see Supplementary materials). In the Control condition, participants could solve the mystery by realizing that one of the characters was a professor, and that the sample had been hidden in a faculty-only bathroom. In the Intervention condition, participants could solve the mystery by realizing that one of the characters (a scientist) was a woman, and that the sample had been hidden in a women's bathroom. In both conditions, participants within pairs worked together to solve the mystery; the experimenters did not dictate how the pairs should engage in the task. From informal observation, pairs differed in their strategies. Some pairs read the documents separately before speaking with each other, whereas others began talking earlier in the process.

All characters were assigned gender ambiguous names. The Officer and the Rival Researcher were described using masculine pronouns, the Assistant Professor was described using feminine pronouns, and the Lab Head was not gendered. The Second in Command was gendered as a woman in the Control condition and not gendered in the Intervention condition. Participants individually completed an answer sheet in which they recorded the two-part answer—the sample's location and the thief's identity—as well as a statement about each character's guilt, which was used to assess participants' use of gendered pronouns as an implicit measure of gender bias (Fazio & Olson, 2003). After completing this answer sheet, an experimenter guided each pair to the correct answer and the reasoning behind it to ensure that all participants had the "aha" moment (see Supplementary Material). Participants then completed the questionnaires.

1.1.3. Measures

The questionnaires included a monetary allocation task in which participants had to allocate $500 among fourteen college organizations, two of which supported women in STEM; a shortened version of the Attitudes Toward Women in Science Scale (ATWSS; Erb & Smith, 1984); the Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996); and demographics.

1.2. Results

1.2.1. Solving the game

Overall, 23.1% of the participants solved the game on their own prior to the explanation from the experimenter. Solving the game was defined as correctly identifying the thief and correctly identifying the location of the stolen sample. A binary logistic regression found no main effects of gender ($b = 0.42$, $SE = 0.56$, $p = .456$), condition ($b = 0.18$, $SE = 0.55$, $p = .737$), or their interaction ($b = -0.56$, $SE = 0.80$, $p = .485$).

1.2.2. Character pronouns

Responses using gendered pronouns were analyzed for frequency of masculine versus feminine pronouns (see Table 1). Fifty-eight of the 143 participants provided a gendered pronoun for the Lab Head and the Assistant Professor. For the Lab Head (who was not gendered), 96.6% of participants who used a gendered pronoun used a masculine pronoun.

Table 1
Gendered pronouns in Study 1.

Character	Gender in game	% Gendered with male pronoun		% Gendered with female pronoun		% Did not use gender pronoun	
		Control	Intervention	Control	Intervention	Control	Intervention
Officer	Male	54.2	52.1	0	0	45.8	47.9
Assistant professor	Female	6.9	5.6	31.9	36.6	61.1	57.7
Lab head	None	44.4	33.8	0	2.8	63.4	62.9
2nd in command	Female in control; none in intervention	30.6	23.9	1.4	4.2	68.1	71.8
Rival researcher	Male	30.6	45.1	1.4	0	68.1	54.9

In addition, 15.5% of participants mistakenly described the Assistant Professor (who was labeled as female in the game documents) using masculine pronouns. Chi-square tests found no gender differences in assigning pronouns to the Lab Head ($\chi^2(1) = 1.93$, $p = .16$) or the Assistant Professor ($\chi^2(1) = 0.13$, $p = .72$).

1.2.3. Gender equality and sexism

Separate random intercept hierarchical linear models (HLMs) using the lme4 (Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) packages in R were conducted on each of the dependent variables (ATWSS and ASI) with game Condition (control coded as "−1", intervention coded as "1"), participant Gender (male coded as "−1", female coded as "1"), their interaction entered as fixed factors, Group entered as the second order random factor, and random slopes and intercepts for Condition and Gender. The HLM on the ATWSS found a main effect of gender ($b = 0.20$, $SE = 0.09$, $t(65.26) = 2.25$, $p = .028$) such that women had more positive views about women in science than men, but there was no main effect of Condition ($b = 0.08$, $SE = 0.09$, $t(25.51) = 0.88$, $p = .388$) or interaction of Gender and Condition ($b = -0.19$, $SE = 0.12$, $t(1, 64.95) = -1.57, p = .121$). The overall effect size of the model was $R_m^2 = 0.04$ (Nakagawa & Schielzeth, 2013). The HLM on the ASI found no main effect of Condition ($b = -0.13$, $SE = 0.16$, $t(39.26) = -0.82$, $p = .418$), no main effect of Gender ($b = -0.12$, $SE = 0.17$, $t(36.77) = -0.67$, $p = .504$), and no interaction of Gender and Condition ($b = 0.28$, $SE = 0.24$, $t(78.22) = 1.19$, $p = .239$). The overall effect size of the model was $R_m^2 = 0.01$ (Nakagawa & Schielzeth, 2013).

1.2.4. Money allocation

A random intercept hierarchical linear model (HLM) using the lme4 and (Bates et al., 2015) lmerTest (Kuznetsova et al., 2017) packages in R was conducted on the percent of money participants allocated toward women in STEM organizations with game Condition (control coded as "−1", intervention coded as "1"), participant Gender (male coded as "−1", female coded as "1"), their interaction entered as fixed factors, Group entered as the second order random factor, and random slopes and intercepts for Condition and Gender. Women were more likely to allocate money to women in STEM organizations: ($b = 12.44$, $SE = 4.35$, $t(39.89) = 2.86$, $p = .007$), but there was no main effect of Condition ($b = -1.92$, $SE = 3.60$, $t(41.43) = -0.53$, $p = .597$) or interaction of Gender and Condition: ($b = -7.36$, $SE = 5.62$, $t(66.48) = -1.31$, $p = .195$). The overall effect size of the model was $R_m^2 = 0.11$ (Nakagawa & Schielzeth, 2013).

1.3. Discussion

The college student participants were much more likely to assume a non-gendered scientist would be a man rather than a woman. In addition, contrary to hypotheses, the intervention condition did not increase positive attitudes toward women in science, decrease sexism, or increase donations to women in STEM organizations.

2. Study 2

To test the effects of the gender bias game among high school students, in Study 2, students at a boarding school played either the gender bias or control game from Study 1 as part of a diversity event. In addition, measures about game enjoyment were added to see if enjoying the game was associated with less bias. Based on the results from Study 1, we hypothesized that game difficulty may have frustrated participants and interfered with the potentially positive effects of the intervention. Therefore, in Study 2 we measured game enjoyment to see if that was correlated with responses to the measures after playing the game. As the current research was concerned with reducing defensive reactions to learning that one has engaged in biased assumptions, game enjoyment may provide a buffer for that potentially defensive reaction. That is, if participants are enjoying the game, they may feel less threatened by what the game is teaching them.

2.1. Method

2.1.1. Participants

Study 2 was conducted with 374 high school students at a boarding school. The sample size was determined based on the school enrollment: all students were allowed to participate in the study. As the sample size was based on a convenience sample, a power analysis was not run prior to data collection; however, the sample size was over twice as large as the sample in Study 1. A parental information sheet was emailed to parents and assent was obtained from all participants. Of the 374 participants, 17 were excluded for not providing gender, 4 were excluded for choosing "other" for gender, and 12 were excluded because a teacher informed them of the purpose of the study prior to the questionnaires, leaving a final sample of 341 (168 men, 173 women, $M_{age} = 16.33$, $SD_{age} = 1.3$).

2.1.2. Procedure

Participants were randomly assigned to groups of four to seven players in different rooms (fifty-five groups were mixed-gender, 7 groups were all male, 7 groups were all female). Participants were in larger groups in Study 2 than in Study 1 due to space constraints (i.e., participants in pairs would not have been able to spread out enough from each other). Each room was randomly assigned to one of the two conditions from Study 1: control or gender bias. The game documents were slightly shortened and clarified for easier comprehension (see Supplementary materials), and playtime was increased to 45 min. As in Study 1, after participants recorded their answers, an experimenter went through the solution logic with them. Participants then completed the questionnaire from Study 1 without the money allocation task and with the addition of five gameplay questions answered on a 1 (not at all true) to 5 (very true) scale: I enjoyed playing the game; I disliked playing the game; I learned something new about myself from playing this game; I learned something new about my peers from playing this game; The game was challenging; The game has opened my mind.

2.2. Results

2.2.1. Solving the game

Overall, 38.7% of the high school participants arrived at the correct solution. It is likely that more high school students were able to solve the game than the college students due to the changes made to the game that were designed to make it easier (e.g., less reading, simplified maps). A binary logistic regression examining the impact of Condition and Gender on whether participants arrived at the correct solution found a significant main effect of Gender ($b = -0.33$, $SE = 0.12$, $\chi^2(1) = 7.47$, $p = .006$) and a significant main effect of Condition ($b = 0.54$, SE = 0.12, $\chi^2(1) = 20.67$, $p < 0.001$). Women and participants in the Intervention condition were more likely to solve the game than men and participants in the Control condition, respectively. The two main effects were qualified by an interaction ($b = 0.39$, $SE = 0.12$, $\chi^2(1) = 10.52$, $p = .001$), such that men (odds ratio = 0.42) were more likely than women (odds ratio = 0.15) to solve the game in the Control condition but men (odds ratio = 0.49) were not more likely than women (odds ratio = 0.52) to solve the game in the Intervention condition.

2.2.2. Character pronouns

The same character descriptions were used from Study 1 including the gender assignments for the characters (see Table 2). For the Lab Head role, which was not explicitly gendered, 60.1% of participants used a gendered pronoun and of those, 93.7% used a masculine pronoun. A chi-square test found no significant difference in pronoun usage for the Lab Head between the Intervention condition (93.5%) and the Control condition (93.8%; $\chi^2(1) = 0.003$, $p = .95$). In addition, for the Assistant Professor (an explicitly labeled woman) 50.4% of participants used a gendered pronoun. Of the participants who gendered the Assistant Professor, 15.7% of participants mistakenly used masculine pronouns. Furthermore, a chi-square test found that participants in the Control condition were more likely to use a masculine pronoun to describe the Assistant Professor (23.6%, 21 out of 89) than participants in the Intervention condition (7.2%, 6 out of 83; $\chi^2(1) = 8.70$, $p = .003$).

2.2.3. Game enjoyment

Game enjoyment was positively correlated with positive attitudes toward women in science and negatively correlated with sexism (see Table 3).

2.2.4. Gender equality and sexism

As in Study 1, separate random intercept hierarchical linear models (HLMs) using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages in R were conducted on each of the dependent variables (ATWSS and ASI) with game Condition (control coded as "−1", intervention coded as "1"), participant Gender (male coded as "−1", female coded as "1"), their interaction entered as fixed factors, Group entered as the second order random factor, and random slopes and intercepts for Condition and Gender. For the ATWSS, there was a main effect of condition ($b = -0.07$, $SE = 0.03$, $t(77.02) = -2.37$, $p = .020$) such that participants in the Control condition scored higher on positive attitudes toward women in science than participants in the Intervention condition, and a main effect of gender ($b = 0.15$, $SE = 0.03$, $t(84.56) = 4.83$, $p < .001$) such that women scored higher than men. The interaction was not significant: $b = 0.05$, $SE = 0.03$, $t(84.56) = 2.51$, $p = .117$. The overall effect size of the model was $R^2_m = 0.26$ (Nakagawa & Schielzeth, 2013). Similarly, for the ASI, there was a main effect of condition ($b = 0.07$, $SE = 0.04$, $t(88.97) = 2.06$, $p = .042$) such that participants in the Control condition scored lower on sexism than participants in the Intervention condition. There was also a main effect of gender ($b = -0.32$, $SE = 0.03$, $t(163.89) = -9.27$, $p < .001$) such that women had lower sexism scores than men. The interaction was not significant: $b = 0.004$, $SE = 0.03$, $t(163.89) = 0.12$, $p = .908$. The overall effect size of the model was $R^2_m = 0.21$ (Nakagawa & Schielzeth, 2013).

Table 2
Gendered pronouns in Study 2.

Character	Gender in game	% Gendered with male pronoun		% Gendered with female pronoun		% Did not use gender pronoun	
		Control	Bias	Control	Bias	Control	Bias
Officer	Male	62.2	65.1	0	1.8	37.8	33.1
Assistant professor	Female	12.2	3.6	39.5	45.6	51.7	50.9
Lab head	None	50.6	62.1	3.5	4.1	45.9	33.7
2nd in command	Female in control; none in bias	33.1	21.9	3.5	9.5	63.4	68.6
Rival researcher	Male	36.0	66.9	1.7	0	62.2	33.1

2.3. Discussion

As in Study 1, most participants in Study 2 assumed that a non-gendered scientist was male, but participants in the Intervention condition were less likely to describe a female professor as male than were participants in the Control condition. In Study 2, participants in the Intervention condition scored lower on positive attitudes toward women in science and higher on levels of sexism than participants in the Control condition. Furthermore, in Study 2, the Intervention condition was easier to solve than the Control condition, particularly for women, which is potentially problematic and emphasizes the difficulty in keeping difficulty levels equivalent across intervention conditions.

3. General discussion

Across two studies, over 90% of students who provided a gender pronoun described a scientist as a man. In Study 1, the intervention did not impact college students' perceptions of women, but in Study 2, participants in the intervention condition reported less positive views of women than participants in the control condition. However, in Study 2 there was evidence that the intervention may have been positively affecting implicit attitudes toward women in science: participants in the Intervention condition were less likely to assign masculine pronouns to a female assistant professor. These results simultaneously point to (1) the continued prevalence of unconscious gender assumptions in the sciences, (2) the difficulty of decreasing biases, (3) the potential promise of game methods, (4) the need to consider both explicit and implicit measure of bias, and (5) the need to consider how age impacts responses to bias interventions.

Key to future research on interventions to reduce biases will be to identify how to maneuver around, and overcome, participant defensiveness. Although the present studies did not explicitly measure defensiveness due to concern about alerting participants to the true purpose of the studies, the intention of the intervention seemed to backfire, particularly with high school students who viewed women less positively after playing the Intervention condition. It is possible that college students, compared to high school students, are more comfortable with confronting their biases. For high school students, the "aha" moment might need to be more subtle. Prior game interventions to decrease bias have had success with the idea of carefully embedding the bias themes in a more covert way (Bessarabova et al., 2016; Dunbar et al., 2014; Flanagan & Kaufman, 2016; Kaufman & Flanagan, 2015). However, the results from Study 2 point to the idea that a game intervention that allows individuals to question their assumptions may have a positive impact on more implicit measures. One question for future research will

Table 3
Correlations of attitudes toward women and game enjoyment.

	Control condition		Intervention condition		Attitudes toward women in science	ASI scores
	Men	Women	Men	Women		
I enjoyed playing the game	$M = 3.54$; $SD = 1.21$	$M = 3.82$; $SD = 0.98$	$M = 3.96$; $SD = 1.13$	$M = 3.90$; $SD = 1.06$	**$r(340) = 0.19, p < .001$**	**$r(341) = -0.23, p < .001$**
I disliked playing the game	$M = 2.08$; $SD = 1.14$	$M = 1.60$; $SD = 0.82$	$M = 1.85$; $SD = 1.04$	$M = 1.82$; $SD = 1.03$	**$r(337) = -0.19, p < .001$**	**$r(338) = 0.27, p < .001$**
I learned something new about myself	$M = 1.84$; $SD = 1.14$	$M = 1.80$; $SD = 1.00$	$M = 1.93$; $SD = 1.02$	$M = 2.16$; $SD = 1.17$	$r(338) = 0.02, p = .69$	$r(339) = 0.04, p = .49$
I learned something new about my peers	$M = 2.48$; $SD = 1.22$	$M = 2.35$; $SD = 1.24$	$M = 2.62$; $SD = 1.22$	$M = 2.84$; $SD = 1.29$	$r(337) = 0.09, p = .09$	$r(338) = -0.04, p = .50$
The game was challenging	$M = 4.06$; $SD = 0.85$	$M = 3.98$; $SD = 0.80$	$M = 3.59$; $SD = 1.07$	$M = 3.73$; $SD = 1.02$	$r(338) = 0.11, p = .05$	$r(339) = -0.03, p = .56$
The game opened my mind	$M = 2.45$; $SD = 1.28$	$M = 2.57$; $SD = 1.20$	$M = 2.42$; $SD = 1.39$	$M = 2.80$; $SD = 1.32$	$r(339) = 0.08, p = .16$	$r(340) = -0.06, p = .30$

Results significant at $p < .001$ are bolded.

be how individual differences may impact receptiveness to a game intervention about gender biases. For example, one's motivation to control prejudice and one's proneness to bias blind spot may influence how well this type of intervention works. One possibility is that individuals who are motivated to avoid prejudice and who are more prone to bias blind spot may be particularly good candidates for this type of intervention. Future research should continue to consider how to create interventions that can avoid a defensive backfiring, the individual differences that may impact responses to interventions, and which measures are most likely to capture potential change—whether positive or negative.

3.1. Limitations

An important consideration for the present two studies was that in Study 1 most of the groups were single-gender whereas in Study 2 most of the groups were mixed-gender and, as such, there was not adequate power within study to run analyses on gender composition. As gender composition can affect the way that a task is completed (Apesteguia, Azmat, & Iriberri, 2012) and the perception of the team (West, Heilman, Gullett, Moss-Racusin, & Magee, 2012), it will be important for future research to examine how group gender composition influences the way that team members solve a bias-related puzzle and how the puzzle-solving among groups with different gender compositions may impact assumptions and attitudes about women in science.

A second limitation in the present research is that the experimenters were not able to be blind to condition. Particularly in Study 2 in which each room of participants contained a single condition, it was not possible for the experimenters to be unaware of what was happening. In addition, the debriefing was different for the two conditions. Thus, it is possible that there may have been experimenter effects. However, the results did not show the expected pattern in terms of the explicit measures, indicating that experimenter effects are less likely. In the future, it would be useful to create versions of these types of interventions that can be conducted without an in-person experimenter to reduce the chances for experimenter bias.

Finally, in Study 1, the power analysis was based on a medium effect size from the prior literature on persuasion (Sagarin et al., 2002); however, the effect sizes found in the present research were smaller than predicted: R_m^2 ranged from 0.01 to 0.26. Although Study 2 did not involve an a priori power analysis, the sample size from Study 1 to Study 2 was more than doubled and we were able to recruit a participant sample that is less frequently studied within social psychology. It will be important in future research examining similar interventions to consider the smaller effects found in the present studies and to take those smaller effect sizes into account.

4. Conclusion

The present research provides a new look at an old topic: how can social psychologists apply psychological theory (e.g., illusion of invulnerability; Sagarin et al., 2002) to create interventions that will decrease biases? The present two studies provide evidence that creating an “aha” moment may increase explicit biases but could help reduce implicit biases. In other words, researchers creating an “aha” moment intervention should be aware that the intervention could backfire, but perhaps if designed with that in mind, the intervention could improve implicit attitudes toward women in science. The current research also reinforces the need for future work to elucidate the individual difference and contextual factors that predict whether that “aha” moment will be met with receptivity or reactance.

Open practices

All materials and data have been made publicly available and are linked to the manuscript via Mendeley Data.

Acknowledgments

This work was supported by the National Science Foundation [DRL-1420036 and 1462063].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2018.03.014.

References

Apesteguia, J., Azmat, G., & Iriberri, N. (2012). The impact of gender composition on team performance and decision making: Evidence from the field. *Management Science, 58*, 78–93. http://dx.doi.org/10.1287/mnsc.1110.1348.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education, 15*, 427–448. http://dx.doi.org/10.1007/s11218-012-9185-3.

Bessarabova, E., Piercy, C. W., King, S., Vincent, C., Dunbar, N. E., Burgoon, J. K., ... Lee, Y.-H. (2016). Mitigating bias blind spot via a serious video game. *Computers in Human Behavior, 62*, 452–466. http://dx.doi.org/10.1016/j.chb.2016.03.089.

Clancy, K. B. H., Nelson, R. G., Rutherford, J. N., & Hinde, K. (2014). Survey of academic field experiences (SAFE): Trainees report harassment and assault. *PLoS One, 9*(7), 1–9.

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin, 29*, 532–544. http://dx.doi.org/10.1177/0146167202250923.

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology, 90*, 784–803. http://dx.doi.org/10.1037/0022-3514.90.5.784.

Dunbar, N. E., Miller, C. H., Adame, B. J., Elizondo, J., Wilson, S. N., Lane, B. L., ... Zhang, J. (2014). Implicit and explicit training in the mitigation of cognitive bias through the use of a serious game. *Computers in Human Behavior, 37*, 307–318. http://dx.doi.org/10.1016/j.chb.2014.04.053.

Erb, T. O., & Smith, W. S. (1984). Validation of the attitude toward women in science scale for early adolescents. *Journal of Research in Science Teaching, 21*(4), 391–397.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297–327. http://dx.doi.org/10.1146/annurev.psych.54.101601.145225.

Flanagan, M., & Kaufman, G. (2016). Shifting implicit biases with games using psychology: The embedded design approach. In Y. Kafai, G. Coleman, & B. Tynes (Eds.). *Diversifying Barbie and Mortal Kombat: Intersectional perspectives and inclusive designs in gaming.* Pittsburgh: CMU/ETC Press.

Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin, 30*(12), 1611–1624. http://dx.doi.org/10.1177/0146167204266650.

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*(3), 491–512.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480. http://dx.doi.org/10.1037/0022-3514.74.6.1464.

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

Handley, I. M., Brown, E. R., Moss-Racusin, C. A., & Smith, J. L. (2015). Quality of evidence revealing subtle gender biases in science is in the eye of the beholder. *Proceedings of the National Academy of Sciences, 112*(43), 13201–13206. http://dx.doi.org/10.1073/pnas.1510649112.

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics.* Washington, D.C.: American Association of University of Women.

Hillard, A. L., Ryan, C. S., & Gervais, S. J. (2013). Reactions to the implicit association test as an educational tool: A mixed methods study. *Social Psychology of Education, 16*(3), 495–516. http://dx.doi.org/10.1007/s11218-013-9219-5.

Howell, J. L., Collisson, B., Crysel, L., Garrido, C. O., Newell, S. M., Cottrell, C. A., ... Shepperd, J. A. (2013). Managing the threat of impending implicit attitude feedback. *Social Psychological and Personality Science, 4*(6), 714–720. http://dx.doi.org/10.1177/1948550613479803.

Howell, J. L., Gaither, S. E., & Ratliff, K. A. (2015). Caught in the middle: Defensive responses to IAT feedback among whites, blacks, and biracial black/whites. *Social Psychological and Personality Science, 6*(4), 373–381. http://dx.doi.org/10.1177/1948550614561127.

Howell, J. L., & Ratliff, K. A. (2017). Not your average bigot: The better-than-average effect and defensive responding to implicit association test feedback. *British Journal of Social Psychology, 56*, 125–145. http://dx.doi.org/10.1111/bjso.12168.

Kaufman, G., & Flanagan, M. (2015). A psychologically "embedded" approach to designing games for prosocial causes. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 9*(3), http://dx.doi.org/10.5817/CP2015-3-5.

Kaufman, G., Flanagan, M., & Seidman, M. (2015). Creating stealth game interventions for attitude and behavior change: An "embedded design" model. *Proceedings of DiGRA*.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26. http://dx.doi.org/10.18637/jss.v082.i13.

Moss-Racusin, C. A., Van Der Toorn, J., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2014). Scientific diversity interventions. *Science, 343*, 615–616. http://dx.doi.org/10.1126/science.1245936.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*, 133–142. http://dx.doi.org/10.1111/j.2041-210x.2012.00261.x.

National Science Foundation, N. C. for S. and E. S (2017). Women, minorities, and persons with disabilities in science and engineering: 2017. Arlington, VA. Retrieved from: www.nsf.gov/statistics/wmpd/.

Parker, L. R., Monteith, M. J., Moss-Racusin, C. A., & Van Camp, A. R. (2018). Promoting concern about gender bias with evidence-based confrontation. *Journal of Experimental Social Psychology, 74*, 8–23. http://dx.doi.org/10.1016/j.jesp.2017.07.009.

Reuben, E., Sapienza, P., Zingales, L., & Greenwald, A. G. (2014). How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences, 111*(12), 4403–4408. http://dx.doi.org/10.1073/pnas.1314788111.

Sagarin, B. J., Cialdini, R. B., Rice, W. E., & Serna, S. B. (2002). Dispelling the illusion of invulnerability: The motivations and mechanisms of resistance to persuasion. *Journal of Personality and Social Psychology, 83*(3), 526–541. http://dx.doi.org/10.1037//0022-3514.83.3.526.

Shields, S. A., Zawadzki, M. J., & Johnson, N. R. (2011). The impact of the workshop activity for gender equity simulation in the academy (WAGES–academic) in demonstrating cumulative effects of gender bias. *Journal of Diversity in Higher Education, 4*, 120–129. http://dx.doi.org/10.1037/a0022953.

West, T. V., Heilman, M. E., Gullett, L., Moss-Racusin, C. A., & Magee, J. C. (2012). Building blocks of bias: Gender composition predicts male and female group members' evaluations of each other and the group. *Journal of Experimental Social Psychology, 48*, 1209–1212. http://dx.doi.org/10.1016/j.jesp.2012.04.012.