Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Experimental Social Psychology 39 (2003) 83–90
Journal of Experimental Social Psychology
www.elsevier.com/locate/jesp

Alleviating women's mathematics stereotype threat through salience of group achievements

Rusty B. McIntyre, René M. Paulson, and Charles G. Lord*

Department of Psychology, Texas Christian University, Fort Worth, TX 76129, USA

Received 28 September 2001; revised 28 December 2001

Abstract

Stereotype threat impairs performance in situations where a stereotype holds that one's group will perform poorly. Two experiments investigated whether reminding women of other women's achievements might alleviate women's mathematics stereotype threat. In Experiment 1, college women performed significantly better on a difficult mathematics test when they were first told that women in general make better participants than men in psychology experiments. In Experiment 2, college women performed significantly better on a difficult mathematics test when they first read about four individual women who had succeeded in architecture, law, medicine, and invention. The results are seen as having implications for theories of stereotype threat, self-evaluation, and performance expectations.

> Very learned women are to be found, in the same manner as female warriors; but they are seldom or never inventors.
> Voltaire, *Dictionnaire Philosophique*, 1764

Voltaire's disparaging remark about women inventors might be dismissed as merely one man's opinion, except that it conveys a commonly held stereotype and might pose a threat to women's aspirations and performance. Stereotype threat impairs performance in situations where a stereotype holds that one's group will perform poorly (Steele & Aronson, 1995). African-Americans might suffer from stereotype threat in standardized testing situations. Similarly, women might suffer from stereotype threat in situations that require mathematical ability. Stereotype threat adds to an already threatening or frustrating situation, such as taking the Scholastic Achievement Test (SAT) or Graduate Record Examination (GRE), the additional burden of not wanting to harm one's group by confirming a negative stereotype.

Several studies have found performance decrements attributable to stereotype threat. Other studies have investigated ways to alleviate stereotype threat. The present experiments examined one procedure that might be used to alleviate stereotype threat and improve performance. Specifically, women's mathematics stereotype threat might be alleviated by making women aware of other areas in which women have been successful as a group or as individuals.

Detrimental effects of stereotype threat

Steele and Aronson (1995, Study 4) gave Black and White students difficult GRE verbal questions, where the stereotype holds that Black students score poorly. Some students had to indicate their race at the top of the form; others did not. As the researchers predicted, Blacks scored significantly worse than Whites in the condition where their race was made salient, but not in the condition where their race was not made salient.

Stereotype threat is a general phenomenon that applies to any group about which there exists a negative stereotype, not just to one particular race, ethnic group, or other social category (Croizet & Claire, 1998). Stone, Lynch, Sjomeling, and Darley (1999, Study 1), for instance, had Black and White college students putt golf balls. When the experimenter described putting as diagnostic of "sports intelligence," Blacks performed

* Corresponding author.
E-mail address: c.lord@tcu.edu (C.G. Lord).

0022-1031/02/$ - see front matter © 2002 Elsevier Science (USA). All rights reserved.
PII: S0022-1031(02)00513-9

worse than Whites. When the experimenter described the same task as diagnostic of "natural athletic ability," Blacks performed better than Whites. Also, Black students who thought the task was a test of sports intelligence performed worse than black students who thought it was a test of natural athletic ability.

Stereotype threat also affects mathematics. Aronson et al. (1999) had White college students, who would not normally suffer from stereotype threat, take difficult items from the GRE quantitative section. Some of the students, but not others, were reminded just before they took the test of the stereotype that Whites are inferior to Asians at mathematics. As predicted, White students who were made aware of a relatively negative stereotype performed more poorly than did students in the control group, where the negative stereotype was not made salient.

In a related study, Shih, Pittinsky, and Ambady (1999) gave Asian-American women difficult mathematics questions. Some of the women were reminded that they were women; others were reminded that they were Asian. When the women were made aware of their identity as Asians, they scored better on the mathematics test than did women in a control group. When they were made aware of their identity as women, they scored worse than did women in the control group. Stereotype threat, then, can occur for whichever of an individual's group identities happens to be temporarily salient and applicable to the situation.

Spencer, Steele, and Quinn (1999, Studies 2 & 3) also investigated women's mathematics stereotype threat. They gave college students two difficult mathematics tests. Before one test, the researchers told students that many people believe women are inferior to men at mathematics, but the empirical evidence is mixed. Before the other test, the researchers explained that women scored as well as men on that specific test. The researchers thus made the stereotype salient for half of the tests. Relative to men, women had lower mean test scores (corrected for guessing) on the "women score worse" test than on the "no-gender differences" test.

Alleviating stereotype threat

Steele (1997); (Steele & Aronson, 1995) speculated about the mechanisms responsible for stereotype threat. People who are aware of a situation-relevant stereotype about their group's inferiority might become aroused and inappropriately narrow their focus of attention (Easterbrook, 1959), become self-conscious (Lord & Saenz, 1985; Saenz, 1994; Saenz & Lord, 1989), and/or lower their performance expectations (Stangor, Carr, & Kiang, 1998). Attempts to alleviate stereotype threat and improve performance, therefore, have addressed these possible mechanisms.

Some studies have attempted to alleviate stereotype threat by redefining the situation as less threatening (Croizet & Claire, 1998; Steele & Aronson, 1995). Steele and Aronson (1995, Studies 1 & 2), for example, gave Black and White college students GRE verbal questions, but with different introductions. They told some of the students that the test was diagnostic of verbal ability. They told other students that the test was not diagnostic of verbal ability, which presumably made the situation less arousing. Although Blacks scored worse than Whites with the standard introduction that the test is diagnostic, they scored as well as Whites with the non-diagnostic introduction. Telling Black students that the test was not diagnostic of verbal ability presumably removed the added pressure of representing their group, decreased arousal, and allowed them to perform better.

Other studies have attempted to alleviate stereotype threat by claiming that the stereotype does not apply to the situation in question. Spencer et al. (1999, Studies 2 & 3) gave women and men difficult GRE mathematics questions, but told some of them that women do as well as men on such tests, contradicting the stereotype. With these instructions, women scored as well as men, presumably because they no longer believed that the stereotype applied to the type of test they were taking. As Steele and Aronson (1995) explained, stereotype threat affects only situations where a stereotype holds that one's group will perform poorly. By telling participants explicitly that women perform as well as men, the researchers removed their specific test from the category of threatening situations.

Finally, some studies have alleviated stereotype threat through misattribution. Brown and Josephs (1999, Studies 2 & 3) had men and women try to answer difficult GRE mathematics questions, but provided some of them in advance with an excuse for possible failure. When they had an excuse, women scored as well as men. The misattribution manipulation might have removed the extra pressure that members of negatively stereotyped groups experience in stereotype-relevant situations. When they have an excuse, their performance cannot reflect poorly on their group.

The present research examined an additional way to alleviate the detrimental effects of stereotype threat. Rather than redefining the situation or claiming that the stereotype is incorrect, one might reassure people that their group could take care of itself regardless of their own performance, thus diffusing responsibility (Darley & Latané, 1968). Making the group's achievements salient might lessen the threat and increase performance expectations.

Performance expectations play a prominent role in research on the causes of stereotype threat. Steele and Aronson (1995) blamed task-related anxiety and lowered expectations. Stangor et al. (1998, Experiment 1) provided direct empirical evidence that stereotype threat lowers performance expectations. They gave college women bogus feedback that they had done well or

poorly on a word-finding task said to be diagnostic of originality and creativity. Then they introduced stereotype threat in half of the women in each group by telling them that the second task involved spatial abilities, at which women perform more poorly than men. They told the other women that men and women perform equally on spatial abilities. Then they asked all participants how well they expected to perform on the spatial abilities task.

Stereotype threat influenced performance expectations (Stangor et al., 1998). Women who were told that women have worse spatial abilities than men also expected to score worse on the spatial abilities test than did women who were told that women were as good as men. Furthermore, feedback on the initial word-finding task had no effect. Once stereotype threat had been aroused, women's confidence in their own originality and creativity was overwhelmed by the assertion that women as a group scored poorly on spatial abilities (Stangor et al., 1998). The women drew negative performance expectations from being reminded that they were part of an inferior group. Conversely, women might draw positive performance expectations from being reminded that they are part of a group that does well at even moderately related tasks.

Stangor et al. (1998), however, were primarily concerned with the effects of stereotype threat and group identification on performance expectations, rather than performance itself. The present two studies focused instead on performance—specifically, alleviating the effects of stereotype threat on women's mathematics test performance by reminding them of other domains in which women have succeeded.

Experiment 1

Experiment 1 assessed whether being reminded of women's achievements would improve women's performance on a difficult mathematics test even when the "achievements" were by women in general, in a conceptually unrelated domain. We attempted to alleviate stereotype threat by telling some students (and not others) that women in general perform better than men as participants in psychology experiments.

Method

Participants

One hundred and sixty-two college students (116 women and 46 men) participated for course credit, in mixed-sex groups of 12–30.

Procedure

A female experimenter told students that they would help standardize new items for the GRE quantitative section. To insure that all participants were aware of the stereotype, the experimenter explained that some research shows men outperform women in mathematics, but the empirical evidence is mixed. Some studies have shown an advantage for men, but others have not.

The experimenter distributed booklets that contained the experimental manipulation. The booklets all had the same cover page, so the experimenter did not know which participant received which set of instructions. Students in both conditions read that additional experiments would be offered for extra credit later in the semester. Students in the successful women condition also read that only women could participate in those experiments because "women produce more reliable and valid data, comprehend the task requirements better, and produce better results in all types of psychological experiments." Students in the no instructions condition read no such restriction.

Then the experimenter administered a mathematics test that included 34 very difficult quantitative items from sample GRE tests. Difficult questions were used because previous research indicated that women's mathematics stereotype threat interferes with performance primarily on difficult items (Spencer et al., 1999). The experimenter read aloud the usual test instructions that strongly advised against guessing or skipping items. She then gave participants 20 min to complete the test.

After she collected the tests, the experimenter distributed additional questions regarding students' perceptions of their performance. On 7-point Likert scales, students rated the test's difficulty, their own performance, their performance relative to the average freshman, sophomore, junior, senior, male, and female, how strongly they agreed with the stereotype that men outperform women on mathematics tests, and how much their sex affected their performance. Finally, students were asked to report their SAT verbal and quantitative scores and to list all math courses they had taken in college.[1] Then the experimenter conducted a thorough debriefing.

Results and discussion

Both SAT and GRE scores are usually adjusted for guessing. An adjusted score was calculated for each participant, therefore, by subtracting one-fifth the number of incorrect answers from the number of correct answers. Table 1 shows the mean adjusted scores, corrected for guessing. A 2 (Sex of Participant) × 2 (Condition) analysis of variance (ANOVA) of these adjusted scores yielded a marginal main effect of sex, in which men scored higher ($M = 10.08$, $SD = 4.94$) than women ($M = 8.63$, $SD = 4.16$), $F(1, 158) = 3.46$, $p = .065$. The

[1] Too few SAT scores were reported to use them in the analyses.

Table 1
Mean adjusted score on a mathematics test by men and women who were or were not told that women are more successful than men as experimental participants (Experiment 1)

Instructions	No instructions	Successful women
Women	7.61[a]	9.76[b]
	(3.69)	(4.43)
	$n = 61$	$n = 55$
Men	9.46[c]	10.70[c]
	(.90)	(1.15)
	$n = 23$	$n = 23$

Note. Standard deviations are in parentheses. Row means with different superscripts were significantly different by planned comparisons.

analysis also yielded a significant main effect of condition. Participants who read the successful women instructions scored higher ($M = 10.04$, $SD = 4.76$) than did participants who read no such instructions ($M = 8.12$, $SD = 3.90$), $F(1, 158) = 5.08$, $p = .026$. The sex × condition interaction was not significant, $F(1, 158) = 0.36$, $p = .55$. The overall ANOVA, however, tests all possible patterns of interaction among the means, whereas the central hypothesis called for one specific pattern in which the mean in one cell—women in the no instructions condition—would be significantly lower than the means in the other three cells. To test this hypothesis, therefore, a planned comparison was performed using contrast weights of −3, 1, 1, and 1. The planned comparison was significant, $F(1, 158) = 11.27$, $p < .001$. Women in the no instructions condition scored significantly worse than did the average participant in the other three conditions.

Other planned comparisons tested specific questions associated with the central hypothesis. The most important of these questions was whether the manipulation significantly improved women's test scores. A planned comparison of the means in the top row of Table 1 showed that women scored significantly better in the successful women condition than in the control or no instructions condition, $F(1, 158) = 7.99$, $p = .006$. Reading that women outperform men as participants in psychology experiments caused women to score better than they would otherwise on a difficult mathematics test. Also, contrary to speculations that these differences might have represented a self-fulfilling prophecy in which participants merely performed as they were led to believe they might, a planned comparison of the means in the bottom row of Table 1 showed that men scored no differently in the two conditions, $F(1, 158) = 1.62$, $p = .33$.

A planned comparison of means in the first column of Table 1 revealed that when participants had no alleviation instructions, women scored marginally worse than men, $F(1, 158) = 3.07$, $p = .082$. In contrast, a planned comparison of means in the second column of Table 1 showed that when both sexes read about women being successful as participants in psychology experiments, women scored as well as men, $F(1, 158) = 0.78$, $p = .377$.

These results were unchanged when the number of college mathematics courses participants had taken was used as a covariate. The results reported for adjusted scores were also similar to results for the percentage of attempted items correct. On these scores, the planned comparison for the central hypothesis was marginally significant, $F(1, 158) = 2.82$, $p = .095$. More importantly, a separate planned comparison showed that women in the successful women condition got a higher percentage of items correct ($M = 58.68$, $SD = 14.54$) than did women in the control condition ($M = 49.98$, $SD = 16.39$), $F(1, 158) = 8.28$, $p = .005$. None of the other planned comparisons was significant.

Finally, women in the successful women condition might have followed a more cautious strategy, attempting only items on which they felt confident. That "increased cautiousness" explanation, however, was not supported. In a planned comparison of the number of items attempted, women in the successful women condition attempted no fewer questions ($M = 19.72$, $SD = 7.01$) than did women in the control condition ($M = 20.06$, $SD = 7.28$), $F(1, 158) = .06$, $p = .80$. When women were told that women excel in psychology experiments they attempted just as many items, but correctly answered a larger percentage of the items. The manipulation also had no effect on the number of items that men attempted (Ms = 23.43, $SD = 7.30$ vs. 26.78, $SD = 7.34$), $F(1, 44) = 2.40$, $p = .117$.[2]

Experiment 2

Experiment 1 showed that being reminded of women's general achievements in an unrelated domain increased the participating women's performance on a difficult mathematics test. The manipulation, however, entailed a broad assertion by a female experimenter that women excel. Participants might have accepted this assertion uncritically because a female in authority delivered it. Experiment 2, in contrast, had a male experimenter who made no such assertions. Participants simply read about individual women who had succeeded in different professions. Furthermore, the manipulation in Experiment 1 conveyed information about women as a group. The generalizability of such manipulations would be enhanced if similar alleviation occurred following information about successful individual women.

[2] On the additional questions asked after the mathematics test, the only significant finding was that women in the successful women condition perceived that they had performed better ($M = 3.16$, $SD = 1.12$) than did women in the control condition ($M = 2.69$, $SD = 1.09$), $F(1, 158) = 5.27$, $p = .023$.

In everyday life, women might encounter media portrayals of individual women who have burst the glass ceiling and beat men in such fields as business, law, medicine, and invention. Experiment 2 tested whether such portrayals might make group achievements salient and increase women's performance even in the face of stereotype threat.

Method

Participants

One hundred and six college students (74 women and 32 men) participated for course credit, in mixed-sex groups of 10–30.

Procedure

The procedure was similar to that used in Experiment 1, except for the manipulation. A male experimenter told students they would participate in two separate studies. In the first study, they would help develop stimulus materials for future experiments. In the second study, they would help standardize items for the quantitative GRE. He mentioned in passing that some previous research shows men outperform women on math tests, but the empirical evidence is mixed. All students were thus made aware of the relevant stereotype prior to any manipulation intended to alleviate that stereotype.

Then all students read and critiqued four brief biographical essays, supposedly drawn from *Entrepeneur* and *Who's Who*, that might be used in future studies. In reality, students were randomly assigned to two conditions. In the successful women condition, all four biographical essays concerned women who had succeeded in various professions. One of the fictitious biographies, for instance, described Janet Haley, a successful architect who started working for Houghton and Associates. The "old boy network" at Houghton assigned her very few contracts, so she took a risk by starting her own consulting business. For two years, she had no contracts, but continued to stress quality over quantity. Then the Tate Museum hired Janet to design a modern sculpture facility in Bradford-on-Avon, which opened to enormous critical acclaim. The other three biographies described other women who had succeeded in law, medicine, and invention. In each case, the woman in question had overcome gender bias. In no case, however, did the description refer either explicitly or implicitly to mathematics or related skills.

Students in the successful corporations condition read almost identical biographical essays, except that the essays concerned four successful corporations rather than four successful women. The architectural essay, for example, described ART-itectural Associates, a highly successful architectural firm that was formed as a risky venture by former Houghton and Associates architects. ART-itectural Associates had to prove itself by emphasizing quality over quantity. For two years they had no contracts, but then they were hired by the Tate Museum to design a modern sculpture facility in Bradford-on-Avon. They received the same critical acclaim as had Janet Haley. Essays in the successful corporations condition were as similar as possible to those in the successful women condition, except for attributing success to business firms rather than to individual women.

To insure that students read the materials, participants in both conditions had to answer questions about the profession described, the philosophy highlighted, the success of that philosophy, and the obstacles overcome. When students had written their answers (which indicated that all participants had read and comprehended the materials), the male experimenter collected the materials and left the room.

A female experimenter, who had no knowledge of which participant had received which essays, entered to conduct her study, which was supposedly unrelated to the male experimenter's. She gave students the same mathematics test, with the same instructions and time limit, as was used in Experiment 1. After she had collected the tests, she distributed and collected the same additional questions as described for Experiment 1 and conducted a thorough debriefing.

Results and discussion

Manipulation checks

Fifty-three additional students read and rated either the four successful corporation essays or the four successful women essays. By one-way ANOVAs, students who read about successful women were more likely than those who read about successful corporations to think the essays meant that women can score well on math tests, $F(1,51) = 4.23$, $p < .05$, and that women can score better on math tests than they had previously believed, $F(1,51) = 4.64$, $p < .05$. The two sets of essays differed significantly in their effects on performance expectations, as defined by women's likelihood of performing well on mathematics tests.

Mathematics performance

As in Experiment 1, an adjusted score was calculated for each participant by subtracting one-fifth the number of incorrect items from the number of correct items. Table 2 shows the mean adjusted scores. A 2 (Sex of Participant) $\times$ 2 (Condition) ANOVA of these adjusted scores yielded a main effect of sex, in which men scored higher ($M = 10.32$, $SD = 4.19$) than women ($M = 8.21$, $SD = 4.46$), $F(1,102) = 5.85$, $p = .017$. The analysis also yielded a marginally significant main effect of condition. Participants who read the successful women essays scored higher ($M = 9.78$, $SD = 4.41$) than did participants who read the successful corporation essays ($M = 7.83$, $SD = 4.36$), $F(1,102) = 3.09$, $p = .082$.

Table 2
Mean adjusted score on a mathematics test by men and women who had recently read biographical essays about successful corporations or successful women (Experiment 2)

Biographical essays	Successful corporations	Successful women
Women	6.82[a]	9.45[b]
	(4.17)	(4.40)
	$n = 35$	$n = 39$
Men	10.05[c]	10.61[c]
	(4.07)	(4.44)
	$n = 16$	$n = 16$

Note. Standard deviations are in parentheses. Row means with different superscripts were significantly different by planned comparisons.

The sex × condition interaction was not significant, $F(1, 102) = 1.29,\ p = .257$. As in Experiment 1, however, the central hypothesis was tested with a planned comparison that used contrast weights of −3, 1, 1, and 1. The planned comparison was significant, $F(1, 102) = 12.42,\ p < .001$. Women in the control condition scored significantly worse than did the average participant in the other three conditions.

Other planned comparisons tested specific questions associated with the central hypothesis. The most important of these questions was whether the manipulation significantly improved women's test scores. A planned comparison of the means in the top row of Table 2 showed that women scored significantly better when they read about successful women than when they read about successful corporations, $F(1, 102) = 6.92,\ p = .01$. Also, the differences did not represent a self-fulfilling prophecy. A planned comparison of the means in the bottom row of Table 1 showed that men scored no differently in the two conditions, $F(1, 102) = .14,\ p = .71$.

A planned comparison of means in the first column of Table 2 revealed that when participants read about successful corporations, women scored significantly worse than men, $F(1, 102) = 6.23,\ p = .014$. In contrast, a planned comparison of means in the second column of Table 2 showed that when both sexes read about successful women, women scored as well as men, $F(1, 102) = 0.83,\ p = .363$.

These results were unchanged when the number of college mathematics courses participants had taken was used as a covariate. The results reported for adjusted scores were also similar to results for the percentage of attempted items correct. On these percent correct scores, the planned comparison for the central hypothesis was significant, $F(1, 102) = 8.78,\ p = .004$. More importantly, a separate planned comparison showed that women in the successful women condition got a higher percentage of items correct ($M = 58.68,\ SD = 14.54$) than did women in the control condition ($M = 49.98,\ SD = 16.39$), $F(1, 102) = 7.34,\ p = .008$, whereas men did not ($M$s $= 58.75,\ SD = 12.14$ vs. $66.22,\ SD = 17.93$), $F(1, 102) = 1.59,\ p = .209$. Finally, women scored significantly worse than men in the control condition, $F(1, 102) = 7.98,\ p = .006$, but not in the alleviation condition, $F(1, 102) = .58,\ p = .448$.

As in Experiment 1, the results were not affected by cautious versus risky test-taking strategies. In a planned comparison of the number of items attempted, women in the successful women condition attempted no fewer of the questions ($M = 18.87,\ SD = 7.68$) than did women in the successful corporations condition ($M = 18.29,\ SD = 7.04$), $F(1, 102) = .12,\ p = .734$. When women read about other women (rather than corporations) who had succeeded in various professions, they attempted just as many items, but correctly answered a larger percentage of those items. The manipulation also had no effect on the number of items that men attempted (Ms $= 22.25,\ SD = 6.96$ vs. $19.19,\ SD = 8.08$), $F(1, 102) = 1.32,\ p = .26$.[3]

General discussion

The results of both experiments supported the central hypothesis that women's performance under mathematics stereotype threat might be improved by reminding them of other women's achievements. The "other women's achievements" were attributed to the entire group in Experiment 1, where women who were under stereotype threat were told that women in general make more conscientious participants in psychology experiments. The "other women's achievements" were more specific in Experiment 2, where women who were under stereotype threat read about four individual women who had succeeded in law, medicine, architecture, and invention. In both experiments, though, women who were reminded of other women's achievements scored better on difficult mathematics questions than did women in a control group.

These results suggest that being reminded of any type of group achievement might alleviate the performance of people who belong to negatively stereotyped groups. However, such a prospect, although desirable, seems unlikely. First, the manipulation in both studies emphasized not just other women's achievements, but other women's achievements relative to men. In Experiment 1,

[3] On the additional questions asked after the math test, two comparisons were significant for women. First, women in the successful women condition perceived themselves as having performed better compared to the average man ($M = 3.59,\ SD = 1.41$) than did women in the successful corporations condition ($M = 2.97,\ SD = 1.10$), $F(1, 102) = 4.61,\ p = .034$. Second, women who read about successful women were less likely than women who read about successful corporations to believe that their sex affected their performance (Ms $= 3.41,\ SD = 1.33$ vs. $4.06,\ SD = 1.41$), $F(1, 102) = 3.97,\ p = .049$.

women learned that men would no longer be allowed to participate in psychology experiments, because women were better at it. In Experiment 2, women read about four other women who had overcome gender bias to beat men in male-dominated professions. Future research on this method of alleviating women's mathematics stereotype threat should include information about other women's achievements that does not mention beating men, to assess whether that element of the information is essential to alleviation effects.

Future research might also examine whether this method of alleviation depends on information about success in stereotype-relevant domains. Intuitively, law, medicine, architecture, and invention differ in their relevance to mathematics. Because participants in Experiment 2 read all four scenarios, it is impossible to tell whether the essays most relevant to mathematics alleviated stereotype threat and the others had no effect. Reminders of group achievement might even exacerbate the threat. Imagine, for instance, reminding Black students, just before taking the GRE-Verbal test, that Blacks are superior to Whites at basketball. Such a reminder might activate a different stereotype—that athletes make poor students—and introduce a new source of stereotype threat.

Another interesting topic for future research involves the importance of a successful role model. In both experiments, the mathematics test was administered by a female experimenter. Steele (1997, p. 625) cited successful role models as one way of alleviating the detrimental effects of stereotype threat. It is possible, then, that the information about successful women might have had no effect if its underlying message—that women are competent—had been undermined by having a man administer the test. A male test administrator might have inadvertently reinforced the stereotype that only men are capable of taking charge in situations that involve mathematics.

One could argue that the successful lawyer, brain surgeon, architect, and inventor in Experiment 2 served as positive role models for our female participants. In a relevant experiment, participants learned of one other person who was similar to them in a coincidental way, but that other person's success on a task led participants to expect to do well themselves on the same task (Stotland & Hillmer, 1962). In our experiments, women learned of four other women, and these other women's success in four professions led participants to perform better on a different task. If the present alleviation effects depended on generalizing from the achievements of group role models to one's own likely performance, then "success by association," as predicted by earlier ideas about generalization of similarities (e.g., Tolman, 1949; Heider, 1958), might have even broader theoretical and practical implications than previously identified in studies of "basking in reflected glory" (Cialdini et al., 1976; Cialdini & de-Nicholas, 1989).

In the present experiments, however, a new procedure for alleviating stereotype threat showed initial promise. The procedure was different from those used in previous studies, which involved either downplaying the diagnostic importance of the test (Croizet & Claire, 1998; Steele & Aronson, 1995), claiming that the specific test is impervious to the stereotype (Spencer et al., 1999), or providing an opportunity to misattribute arousal (Brown & Josephs, 1999; Stone et al., 1999). The present results constitute only a first step, however, in exploring the impact of information about group achievements on alleviating stereotype threat. Before drawing strong conclusions about alleviation and its underlying mechanisms, it seems necessary to replicate the finding with groups other than women, stereotypes other than mathematics test performance, achievements of differing relevance to the stereotype, and different types of role models (Steele, 1997). For now, however, the results suggest preliminary optimism that information about group achievements might be added to the list of manipulations that have been found to decrease stereotype threat and its pernicious consequences. Perhaps Voltaire's disparaging remark that began this article needs to be counterbalanced by a woman's perspective:

> Women have served all these centuries as looking-glasses possessing the magic and delicious power of reflecting the figure of man at twice its natural size.
>
> Virginia Woolf, *A Room of One's Own*, 1929

Acknowledgments

We thank Christina Blodgett for assistance with Experiment 1 and Charles F. Bond, Jr. for comments on an earlier draft. Reprint requests: C. Lord, Department of Psychology, TCU 298920, Ford Worth, TX 76129, USA.

References

Aronson, J., Lustina, M. J., Good, C., Keogh, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29–46.

Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76*, 246–257.

Cialdini, R. B., Borden, R. J., Thorne, A., Walker, M. R., Freeman, S., & Sloan, L. R. (1976). Basking in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology, 34*, 366–375.

Cialdini, R. B., & de-Nicholas, M. E. (1989). Self-presentation by association. *Journal of Personality and Social Psychology, 57*, 626–631.

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588–594.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8*, 377–383.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*, 183–201.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Lord, C. G., & Saenz, D. S. (1985). Memory deficits and memory surfeits: Differential cognitive consequences of tokenism for tokens and observers. *Journal of Personality and Social Psychology, 49*, 918–926.

Saenz, D. S. (1994). Token status and problem-solving deficits: Detrimental effects of distinctiveness and performance monitoring. *Social Cognition, 12*, 61–74.

Saenz, D. S., & Lord, C. G. (1989). Reversing roles: A strategy for undoing memory deficits associated with tokenism. *Journal of Personality and Social Psychology, 56*, 698–708.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28.

Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75*, 1191–1197.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613–629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology, 77*, 1213–1227.

Stotland, E., & Hillmer, M. L. (1962). Identification, authoritarian defensiveness, and self-esteem. *Journal of Abnormal and Social Psychology, 64*, 334–342.

Tolman, E. C. (1949). There is more than one kind of learning. *Psychological Review, 56*, 144–155.