

The Effect of Negative Performance Stereotypes on Learning

Robert J. Rydell, Michael T. Rydell, and Kathryn L. Boucher
Indiana University

Stereotype threat (ST) research has focused exclusively on how negative group stereotypes reduce performance. The present work examines if pejorative stereotypes about women in math inhibit their ability to learn the mathematical rules and operations necessary to solve math problems. In Experiment 1, women experiencing ST had difficulty encoding math-related information into memory and, therefore, learned fewer mathematical rules and showed poorer math performance than did controls. In Experiment 2, women experiencing ST while learning modular arithmetic (MA) performed more poorly than did controls on easy MA problems; this effect was due to reduced learning of the mathematical operations underlying MA. In Experiment 3, ST reduced women's, but not men's, ability to learn abstract mathematical rules and to transfer these rules to a second, isomorphic task. This work provides the first evidence that negative stereotypes about women in math reduce their level of mathematical learning and demonstrates that reduced learning due to stereotype threat can lead to poorer performance in negatively stereotyped domains.

Keywords: stereotype threat, learning, performance, mathematics, gender differences

Women, especially those in the male-dominated fields of science and mathematics, are routinely subjected to the negative stereotype that "women are bad at math." When women have this pejorative stereotype accessible in memory, they experience *stereotype threat*: the arousal, worrying thoughts, and temporary cognitive deficits evoked in situations where a group member's performance can confirm the negative stereotype about their group's ability in that domain (e.g., Schmader, Johns, & Forbes, 2008; Steele, 1997; Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002). Consistent with this definition, research in social psychology has documented that stereotype threat leads to increased arousal (e.g., Ben-Zeev, Fein, & Inzlicht, 2005; Murphy, Steele, & Gross, 2007; O'Brien & Crandall, 2003), reduced working memory capacity (see Schmader et al., 2008, for a review), and increased emotional suppression (e.g., Johns, Inzlicht, & Schmader, 2008) that harm performance in the negatively stereotyped domain (e.g., Beilock, Rydell, & McConnell, 2007; Schmader et al., 2008;

Spencer, Steele, & Quinn, 1999). The experience of stereotype threat for women in math and its detrimental impact on performance is obviously unwanted, leading many researchers to examine how to reduce the impact of stereotype threat on women's math performance, a strategy that has been extremely fruitful and important (Schmader et al., 2008; Steele et al., 2002). In the current work, we examine the broader and heretofore untested question of whether there is another process through which stereotype threat can reduce performance, namely, by impairing learning in the negatively stereotyped domain (cf. Rydell, Shiffrin, Boucher, Van Loo, & Rydell, 2010).

Stereotype threat is thought of as a performance-based phenomenon that reduces women's ability to solve math problems when their grade is on the line, even though they could solve the same math problems when the threat is absent (e.g., Beilock, 2008; Schmader et al., 2008; Steele, 1997; Steele & Aronson, 1995). This is undoubtedly true when men and women have equivalent levels of mathematical knowledge. Yet, are such assumptions about equal learning warranted? What would be the implications, theoretical and practical, of assuming they are not? Past work has demonstrated that stereotype threat harms women's math performance (e.g., Spencer et al., 1999), but can it also rob women of the ability to learn? If stereotype threat harms learning, then this threat, in addition to any in-the-moment pressure to perform well that a woman experiences, will take a toll on subsequent achievement.

To address these issues, we examined whether stereotype threat undercuts women's ability to learn the mathematical concepts, principles, and operations necessary to excel in math.

This article was published Online First October 04, 2010.

Robert J. Rydell, Michael T. Rydell, and Kathryn L. Boucher, Department of Psychological and Brain Sciences, Indiana University.

We thank Jim Sherman, Allen McConnell, Rich Shiffrin, Rob Goldstone, John Skowronski, and the members of the 21st Annual Duck Conference on Social Cognition for providing helpful feedback about this work. We also thank Alan Boucher for his help with the abstract symbolic logic task in Experiment 3.

Correspondence concerning this article should be addressed to Robert J. Rydell, 1101 East 10th Street, Bloomington, IN 47405. E-mail: rjrydell@indiana.edu

One problem for research that assesses learning is defining what learning is (e.g., Rumelhart & Todd, 1990). In general, during a learning task, people can either not acquire any information, acquire information that is irrelevant to the learning domain, or acquire information that is relevant to the learning domain. In the current work, we define learning as the ability to encode into memory information that is necessary for successful skill completion (e.g., concepts, principles, procedures, rules, relations). Therefore, our definition of learning involves identifying and acquiring relevant information (while ignoring irrelevant information), storing the relevant information into memory, and integrating the relevant information with previously acquired knowledge. Understanding how stereotype threat affects these processes will broaden the domain of stereotype threat to situations where people acquire the skills necessary to perform. For women in math, we propose that stereotype threat reduces women's ability (a) to encode or acquire mathematical concepts, principles, and operations; (b) to select the optimal problem-solving strategy to complete the problem (Quinn & Spencer, 2001); and (c) to correctly execute mathematical operations when the correct concept or principle is applied to a problem. Here, the first process is considered learning and the latter two processes are considered performance.

Although the failure to execute a learned skill is perhaps the most blatant and discouraging aspect of stereotype threat, stereotype threat's impact on women in math may also operate through a more subtle and indirect route: hampering women's ability to acquire or learn the mathematical concepts necessary to perform. It is proposed that reduced learning due to stereotype threat starts by activating the negative stereotype that "women are bad at math" in women's minds (e.g., Wheeler & Petty, 2001). Once this negative stereotype is activated, we predict that women will become worried about confirming the stereotype (e.g., Beilock et al., 2007; Marx & Stapel, 2006), feel aroused (e.g., Ben-Zeev et al., 2005; Murphy et al., 2007; O'Brien & Crandall, 2003), have reduced working memory capacity (e.g., Beilock et al., 2007; Croizet, Despres, Gauzins, Hugeot, & Leyens, 2004; Rydell, McConnell, & Beilock, 2009; Schmader & Johns, 2003), and experience decreased levels of interest in learning math-related material (e.g., Major, Spencer, Schmader, Wolfe, & Crocker, 1998; Steele, 1997).

Because of increased worry, increased arousal, and reduced working memory capacity, women should have difficulty attending to, encoding, and storing the mathematical rules or operations being learned into memory for later use (e.g., Eysenck, 1976; Unsworth & Engle, 2005) when women are interested in learning math. However, stereotype threat may also reduce learning by decreasing women's motivation to learn math (e.g., Major et al., 1998; Steele, 1997). Experiencing stereotype threat may lead women to disengage from information related to the domain of mathematics, reducing the extent to which they attend to and therefore encode information presented about mathematical rules and operations. Given these effects of stereotype threat, we predict that stereotype threat will reduce women's ability to encode mathematical information and to generalize any information learned to similar tasks or problems. In situations where the unlearned information is necessary for skill execution, reduced learning will ultimately lead to reduced math performance.

The detrimental effect of stereotype threat on learning has not been directly tested. But a handful of findings in the stereotype threat literature may be explained, at least in part, by stereotype threat having a detrimental effect on learning. For example, Grimm, Markman, Maddox, and Baldwin (2009) showed that stereotype threat reduced performance on a perceptual categorization task that involved learning which category a line belonged to, but only when the task involved avoiding losses. Because this research involved category learning and it took women longer than men to reach a specified criterion under stereotype threat, it could be inferred that stereotype threat impaired learning. In other work consistent with our predictions, Adams, Garcia, Purdie-Vaughns, and Steele (2006) had a male instructor, whom a female confederate indicated was or was not sexist, teach participants how to solve logic problems. Women told that the instructor was sexist performed more poorly on a subsequent logic test than did women who were told the instructor was not sexist. Given that learning from a supposedly sexist instructor led to lower performance than did learning from a nonsexist instructor, women with a sexist instructor could have learned less during the instruction session.

Furthermore, Cohen and colleagues showed that presenting a self-affirmation manipulation—a manipulation shown to eliminate stereotype threat (Martens, Johns, Greenberg, & Schimel, 2006)—reduced the achievement gap between Caucasian and African American middle school students over a 2-year period (Cohen, Garcia, Apfel, & Master, 2006; Cohen, Garcia, Purdie-Vaughns, Apfel, & Brzustoski, 2009). This reduction in the achievement gap was likely due to reductions in stereotype threat during performance, but the long-lasting nature of these effects may be partially due to increased learning after stereotype threat was reduced. It is important to note that Adams et al. (2006), Cohen et al. (2006, 2009), and Grimm et al. (2009) did not explain their results in terms of stereotype threat-based learning effects. Even if learning effects had been discussed, this past work does not provide clear evidence of stereotype threat-based learning effects and did not attempt to delineate learning from performance in response to threatening situations. For our purposes, the main point is that past work on stereotype threat, even research that on the surface appears to be examining learning, has not directly examined if stereotype threat reduces learning.

There are at least two reasons why this hypothesis has not been directly tested. The first reason is that there are generally no performance differences between the stereotyped group (e.g., women) and the nonstereotyped group (e.g., men) when stereotype threat is not evoked (e.g., Steele et al., 2002). For example, men and women perform equivalently on problems taken from standardized mathematics tests when not under threat (e.g., Spencer et al., 1999). From such a finding, it may seem reasonable to assume equivalent learning of mathematical concepts for women and men. This reasoning, however, fails to take into account several mitigating factors. The majority of stereotype threat research uses college populations that have performed well enough on the mathematics portion of college entrance exams to be admitted. Also, many stereotype threat experiments have selected members of the stereotyped group (e.g., women) who are highly identified with the performance domain (e.g., math) as participants (e.g., Beilock et al., 2007;

Spencer et al., 1999), perhaps obscuring learning differences. Therefore, it could be the case that women in past stereotype threat research may have already learned the mathematical skills necessary to solve the questions asked in those experiments (e.g., word problems that utilize algebra) and therefore performed at the same level as men when stereotype threat was absent. In addition, any deficits in learning between young men and young women could be obscured by increased preparation for math performance events (e.g., increased studying of mathematical concepts, tutoring, prep courses, practice tests, one-on-one student–teacher interaction) by young women who are aware of the negative stereotype for their performance, and increased preparation has not been accounted for in past stereotype threat research. Given these mitigating factors, a better test of our hypothesis examines if stereotype threat inhibits women’s ability to learn new mathematical information.

The second reason why research has not examined whether stereotype threat reduces learning is that learning is difficult to distinguish from performance. One way that learning paradigms assess learning is by measuring people’s ability to perform the learned task. Because stereotype threat reduces performance when learning is equivalent or statistically controlled (e.g., Steele & Aronson, 1995), stereotype threat’s proposed effect on learning cannot be determined from performance measures alone. The research herein is designed to differentiate learning and performance for women in math by using several different methods to assess learning that are designed to show that stereotype threat reduces learning and to rule out performance-based explanations for our results (e.g., an implicit measure,¹ performance on math problems that are vs. are not susceptible to stereotype threat–based performance effects, transfer tasks).

Overview of the Current Work

We examined whether stereotype threat has a detrimental impact on women’s ability to learn math across three experiments. In each experiment, we manipulated stereotype threat and taught women (and men in Experiment 3) a novel math-related task. Experiment 1 was designed to show that stereotype threat reduces learning by inhibiting women’s ability to encode mathematical information into memory. Experiments 2 and 3 were designed to differentiate stereotype threat–based reductions in learning from stereotype threat–based reductions in performance. In Experiment 2, we manipulated stereotype threat before or after women learned how to complete modular arithmetic (MA) problems. Then, women completed MA problems that were susceptible (difficult MA problems) or were not susceptible (easy MA problems) to stereotype threat–based performance decrements (Beilock et al., 2007). We predicted that when stereotype threat was presented before learning, participants would have more difficulty encoding the mathematical operations used to solve MA problems and show reduced performance on easy MA problems.

In Experiment 3, we manipulated stereotype threat before men and women learned an abstract symbolic logic task. We predicted that women under stereotype threat, as compared with men and women not under stereotype threat, would have more difficulty learning this task and transferring the mathematical principles learned on one task to a second task that used the

same underlying principles. Further, we expected an implicit measure of the associations formed during learning, which is not vulnerable to math-related performance effects, to show weaker associations for women under stereotype threat than for men and women not under stereotype threat. By providing converging evidence across three experiments and three different math tasks, we aimed to show that stereotype threat leads to reduced learning and that reduced learning due to stereotype threat can affect performance.

Experiment 1

The first experiment was designed to be an initial examination of our prediction that stereotype threat can reduce learning by interfering with encoding processes. We proposed that stereotype threat reduces women’s ability to learn mathematical rules and operations by reducing their ability to encode mathematical information into memory, not by inhibiting the ability to retrieve mathematical information from memory. In Experiment 1, women learned half of the mathematical rules necessary to successfully solve math problems before either control instructions or stereotype threat instructions were presented. After reading these instructions, they learned the remaining mathematical rules; then mathematical learning and performance were assessed. We expected three sets of findings to obtain, consistent with the hypothesis that stereotype threat reduces learning by inhibiting encoding. First, we expected that women in the control and stereotype threat conditions would learn the same number of mathematical rules presented before the manipulation of stereotype threat and correctly answer the same number of math problems when tested later. Second, we expected that women who received stereotype threat instructions would learn the mathematical rules presented after the instructions more poorly and solve fewer math problems based on those rules than would controls. Third, we expected that when learning was controlled, those who received stereotype threat instructions would solve a smaller proportion of problems than would those in the control condition, regardless of whether the problems were based on rules learned before or after the manipulation.

Because women in the stereotype threat condition will be under threat when answering the mathematical learning and performance questions (i.e., stereotype threat is always present at retrieval), the roles of encoding and retrieval in stereotype threat-based learning effects can be determined. If stereotype threat reduces learning through inhibiting retrieval, mathematical rules presented before and after the induction of stereotype threat should show universally poor mathematical learning and performance. If stereotype threat inhibits the encoding of mathematical information, women should show poorer recall of the mathematical rules presented after the induction of stereotype threat than the rules presented before the induction of stereotype threat. Also, women should solve fewer of the problems

¹ We use the term *implicit* in reference to this measure because of its similarity in structure to other implicit measures. However, it is not entirely clear that the measure we used is implicit, and we do not want to overstate what can be inferred from it.

based on the mathematical rules presented after stereotype threat induction than those presented before stereotype threat induction. However, because the execution of learned skills is affected by threat (e.g., Steele & Aronson, 1995), when learning was controlled, we expected women receiving stereotype threat instructions would perform worse than women receiving control instructions regardless of whether the problems were based on rules presented before or after the stereotype threat manipulation.

Method

Participants. Female undergraduates ($N = 59$) participated for course credit. They were randomly assigned to receive control or stereotype threat instructions.

Procedure.

Learning the novel math task. Participants were informed that they would be learning a “new” form of math that they would later use to solve math problems.² In this new form of math, they had to use certain equations to solve problems that gave them information about what math equation and coefficient to use to solve the equation (e.g., $>5Xb$). Specifically, participants used the coefficient presented in the problem (e.g., 5) as the variable x to solve the equation presented in the row marked with a $>$ and the column marked Xb : in this case, $3(x) + 2$; which is solved $3(5) + 2 = 17$ (see Figure 1). Participants were taught this new form of math by receiving a detailed tutorial of how the system of math worked, during which they were shown how to solve an example problem for each of the eight unique row and column combinations. However, this tutorial came in two parts, separated by the manipulation of stereotype threat (see below). Before the manipulation, participants learned how to complete problems on the basis of the first two rows of Figure 1 while the third and fourth rows of Figure 1 were not visible. After the manipulation, participants learned the remaining four mathematical rules (the third and fourth rows of Figure 1) while the first and second rows were not visible.³ The full table, with all of the equations visible, was presented for 20 s at the end of the learning session.

Stereotype threat manipulation. Participants received one of two instruction sets to manipulate stereotype threat. After completing half of the tutorial, participants were told that the purpose of the experiment was to investigate why some people are better at certain types of math problems than others (see Beilock et al., 2007, p. 276). In the *control condition*, no reference was made to gender. In the *stereotype threat condition*, participants were informed that the research was investigating why women are generally worse at math than men.

	Xa	Xb
*	$(x)/4$	$5(x)-60$
+	$(x)+15$	$10(x)-15$
-	$3(x)+5$	$2(x)$
$>$	$(x)/2$	$3(x)+2$

Figure 1. The table of the mathematical rules presented to participants in Experiment 1. Each of the eight unique row–column combinations was explained to participants in depth and with an example problem.

Math test. Next, participants completed 16 math problems (e.g., $+9Xa$) by typing their answer into a box presented on the screen under the problem. Two problems were randomly presented from each of eight unique Row \times Column combinations. The coefficients in these problem were always positive integers between 1 and 9. The number of correct answers given (out of 16) served as our measure of math performance.

Learning test. After the math test, participants were presented with a version of the table they used for learning that contained all of the row labels (e.g., +, *) and all of the column labels (Xa and Xb) but that was otherwise blank. They were asked to recall the equations they had learned during the learning session for each of the eight unique row and column combinations by typing the appropriate equation into a box presented at the bottom of the screen. The number of correct equations recalled (out of eight) served as our measure of mathematical learning. Mathematical learning and performance were highly correlated for problems based on rules presented before and the rules presented after the introduction of the instructions (before the instructions, $r = .74$, $p < .001$; after the instructions, $r = .91$, $p < .001$).

Results

The impact of stereotype threat on mathematical learning and math performance was assessed with separate 2 (stereotype threat: control, stereotype threat) \times 2 (learning time: before instructions, after instructions) mixed-model analyses of variance (ANOVAs), with the latter factor within subjects. The results for mathematical learning showed the expected two-way interaction, $F(1, 57) = 7.25$, $p < .01$, $\eta_p^2 = .11$ (see Table 1). As predicted, the stereotype threat manipulation did not affect women’s learning of mathematical rules presented before the instructions, $F(1, 57) = 0.68$, $p = .41$, $\eta_p^2 = .01$; however, women in the stereotype threat condition learned fewer mathematical rules presented after the instructions than did women in the control condition, $F(1, 57) = 3.96$, $p = .05$, $\eta_p^2 = .07$. Also as predicted, learning time did not impact the number of mathematical rules women learned in the control condition, $F(1, 28) = 0.61$, $p = .44$, $\eta_p^2 = .02$, but women in the stereotype threat condition learned more of the mathematical rules presented before the instructions than the mathematical rules presented after the instructions, $F(1, 29) = 15.83$, $p < .001$, $\eta_p^2 = .35$.

The results for math performance also showed a two-way interaction, $F(1, 57) = 4.02$, $p = .05$, $\eta_p^2 = .07$ (see Table 1). The stereotype threat manipulation did not affect women’s performance on problems that used mathematical rules presented before the instructions, $F(1, 57) = 0.16$, $p = .69$, $\eta_p^2 = .00$; however, women in the stereotype threat condition solved fewer mathematical problems based on the rules presented after the instructions

² In all three of the experiments presented in this work, two men and six women served as the experimenters and between one and seven participants completed the experiment in each session. The contact between experimenters and participants was minimal. Experimenters simply directed each participant to a private room. After being escorted into a room where they could complete the experiment alone, participants received the instructions, the manipulations, and the dependent variables via computer.

³ In Experiment 1, the equations presented in the first two rows versus the third and fourth rows were counterbalanced. This manipulation had no effects and is not discussed further.

Table 1
Mathematical Learning and Math Performance as a Function of Instruction and Learning Time in Experiment 1

Measure	Control		Stereotype threat	
	Before instructions	After instructions	Before instructions	After instructions
Mathematical learning	0.90 _a	0.76 _a	1.20 _a	0.27 _b
Math performance	1.59 _a	1.55 _a	1.37 _a	0.27 _b

Note. Means in a row with different subscripts were significantly different at the $p \leq .05$ level. The range of scores for mathematical learning was 0–4 both before and after the instructions, whereas the range of scores for math performance was 0–8 both before and after the instructions. For math performance, performance refers to work on math problems based on mathematical rules learned before or after the instructions; however, the math problems were completed after all learning commenced.

than did women in the control condition, $F(1, 57) = 8.12, p = .01, \eta_p^2 = .13$. In addition, learning time did not affect math performance for women in the control condition, $F(1, 28) = .01, p = .93, \eta_p^2 = .00$, but women in the stereotype threat condition solved more of the problems based on the rules presented before the instructions than the problems based on rules presented after the instructions, $F(1, 29) = 8.18, p = .01, \eta_p^2 = .22$.

To examine the impact of our manipulation on performance devoid of its impact on learning, we examined performance on problems in which participants correctly remembered the relevant equation. This analyses showed that the percentage of math problems answered correctly when the mathematical equation was correctly recalled was better in the control condition than in the stereotype threat condition at both Time 1, $t(25) = 2.76, p = .01, d = 1.07$ (control $M = 72\%$, stereotype threat $M = 35\%$), and Time 2, $t(14) = 4.89, p < .001, d = 2.51$ (control $M = 98\%$, stereotype threat $M = 45\%$). Thus, stereotype threat exerted an impact on performance when learning was controlled, consistent with past stereotype threat research (e.g., Steele et al., 2002).

Discussion

Experiment 1 showed that stereotype threat reduced mathematical learning by inhibiting the encoding of mathematical rules into memory. Because the measure of mathematical learning was given after stereotype threat was induced, if retrieval processes were inhibited by stereotype threat, then women in the stereotype threat condition would have shown less mathematical learning and poorer performance regardless of whether the mathematical rules were encoded before or after the introduction of stereotype threat. This was clearly not the case, as learning was greater for mathematical rules encoded before the introduction of stereotype threat than for mathematical rules encoded after the introduction of stereotype threat. Additionally, women in the control condition showed equivalent learning of the mathematical rules regardless of when the mathematical rules were learned (i.e., it was not the case that mathematical rules presented later and after an interruption were universally learned more poorly). This pattern of results on the learning measure and corresponding results on the performance

measure support our prediction that stereotype threat inhibits the encoding of math-related information. Further, we replicated past stereotype threat–based performance results by showing that when learning was controlled, women under threat performed more poorly than did women not under threat.

Experiment 2

In Experiment 1, we showed that stereotype threat reduces women's ability to encode mathematical rules. Despite obtaining the predicted results, we sought to address several issues related to the learning task used in Experiment 1 in Experiments 2 and 3. First, because solving the equation was extremely difficult without having successfully encoded the equation, our learning and performance measures were highly correlated in Experiment 1. Admittedly, these high correlations diminish our ability to delineate between learning and performance. Second, the system of math that we used in Experiment 1 involved memorization (as do many forms learning), but unlike the underlying logic and internal coherence of established mathematical principles, this system comprised arbitrary rules that we created. Therefore, in Experiments 2 and 3, we examined if stereotype threat reduces learning on mathematical tasks that are internally consistent, less reliant on memorization, and more reliant on the use of mathematical principles.

By demonstrating that stereotype threat reduces learning on an established mathematical task (Experiment 2) and on an abstract symbolic logic task that uses mathematical principles (Experiment 3), we sought to address the limitations of our initial novel math task. Most important, we aimed to convincingly delineate the effect of stereotype threat on learning from the effect of stereotype threat on performance. To provide a compelling demonstration that stereotype threat reduces mathematical learning, we must supply more than just a demonstration of stereotype threat's impact on the encoding of math-related information; we must also show that stereotype threat reduces learning and that these learning reductions cannot be explained by reductions in performance.

In Experiment 2, we turned to a mathematical task that allows us to more clearly delineate learning from performance in response to stereotype threat: MA. In past work with stereotype threat and MA, Beilock et al. (2007) manipulated stereotype threat after women had learned how to complete MA problems (i.e., learning could not have been affected by stereotype threat) and varied the difficulty of MA problems. Beilock and colleagues' research showed that stereotype threat only reduced women's performance on difficult problems (i.e., those that involved large numbers and a borrow operation), whereas easy MA problems (i.e., those that involved smaller numbers and did not necessitate a borrow operation) were immune to performance-related stereotype threat effects. We therefore reasoned that any performance decrements on easy MA problems would have to be due to difficulty encoding the proper mathematical operations necessary to complete MA problems and not stereotype threat–based performance effects.

On the basis of this reasoning, in Experiment 2, we experimentally manipulated stereotype threat either before or after participants learned how to solve MA problems. Presenting stereotype threat before learning should reduce women's ability to encode the mathematical rules and operations necessary to solve MA problems, whereas presenting stereotype threat after learning should not affect encoding of the operations necessary to complete MA.

We took advantage of the finding that easy MA problems are immune to stereotype threat–based performance decrements to demonstrate stereotype threat–based learning decrements. If women have learned how to solve MA problems (i.e., stereotype threat is presented after learning), they should be able to solve easy MA problems regardless of whether they are experiencing stereotype threat because the math involved in these problems is so rudimentary that they can be solved even when women are under stereotype threat (e.g., Beilock et al., 2007; O’Brien & Crandall, 2003; Spencer et al., 1999). If women have not learned how to solve MA problems (i.e., stereotype threat is presented before learning), they should not be able to solve easy MA problems because they would not have properly encoded how to complete these problems (even though the math is rudimentary). Because performance on difficult MA problems should be poor under stereotype threat, we expected performance on difficult MA problems to be lower in the stereotype threat condition than in the control condition, regardless of when stereotype threat is presented.

In this experiment, we also assessed learning by having participants describe how to solve a MA problem and define the meaning of the term *mod* (i.e., an abbreviation of the term *modular* used in the problems). We predicted that women who received stereotype threat before learning would (compared with the women in the three remaining experimental conditions) give a poorer explanation of MA and would be less likely to correctly define the term *mod*. Consistent with our prediction that reductions in learning would cause reduced performance on easy MA problems, we expected that the quality of the explanation of MA would account for the effect of the experimental manipulations on easy MA problems.

Method

Participants. Female undergraduates ($N = 104$) participated for course credit. They were randomly assigned to condition in a 2 (stereotype threat: control, stereotype threat) \times 2 (placement of manipulation: before learning, after learning) between-subjects factorial. Because it was crucial that participants were learning a novel math task in this experiment, data from 12 participants who reported previously learning how to solve MA problems were excluded from the analyses. This left 92 participants in the sample.

Procedure. Participants were informed that they would be taking part in an experiment on cognitive skills. Half of the participants received the manipulation of stereotype threat at the beginning of the session; the other half received the manipulation after they learned MA. The stereotype threat instructions were the same as those provided in Experiment 1, but the control instructions were simplified, explaining that our lab was examining “how people solve a number of different types of problems” and “what strategies people use to solve [problems].” In this experiment, we used MA, a relatively simple but little-known math task that involves determining if a math equation with three whole numbers is true (i.e., the answer is a whole number) or false (i.e., the answer is not a whole number).

In MA, an equation that takes the form of $a \equiv b \pmod{c}$ —for example, $9 \equiv 3 \pmod{2}$ —is solved by subtracting b (3) from a (9) and then dividing the solution of $a - b$ (6) by c (2). In our example problem, the solution of $a - b$ (6) divided by c (2) is 3 (which is

a whole number); therefore, the correct answer is true. If one solved the equation $9 \equiv 3 \pmod{4}$, the correct answer would be false because $9 - 3 = 6$ and $6 \div 4 = 1.5$, which is not a whole number. We taught MA by showing participants three sequential computer screens that outlined how to solve this type of problem. The instructions were the exact same instructions used by Beilock et al. (2007). As participants progressed through the tutorial, we measured the amount of time they spent studying the instructions (*study time*) by averaging the amount of time they spent examining each screen ($M = 17.51$ s, $SD = 5.17$).

Learning tests. Next, two measures were used to assess learning of MA. For the equation problem, participants were asked to provide the steps necessary to solve MA problems using the equation $a \equiv b \pmod{c}$ and to indicate when an equation would be true and when an equation would be false. Participants were given unlimited space to describe how MA works; their responses were coded by two raters, who were unaware of the hypotheses or the participants’ experimental condition, on a scale ranging from 1 (*very poor explanation*) to 5 (*excellent explanation*). The raters’ scores were highly interrelated, $r = .82$, $p < .001$, so their scores were averaged to create an *explanation of MA* score. For the *mod* definition question, participants were asked to define the meaning of the mathematical operation *mod*. As there was little ambiguity in correct responses for the *mod* definition question, participants’ answers were scored as correct or incorrect by Robert J. Rydell.

Math test. Finally, participants completed 54 MA problems (18 easy, 18 moderate, and 18 difficult) by indicating whether the equations presented were true (by pressing the *t* key) or false (by pressing the *f* key). The problems were presented in a random order and every participant completed all 54 problems. Easy problems used numbers between 1 and 9 and did not involve a borrow operation when subtracting b from a . Moderate problems used numbers between 2 and 19 (with the mod number always between 2 and 9) and did not involve a borrow operation. The difficult problems used larger numbers (i.e., numbers between 19 and 99) and required a borrow operation. Because the easy and moderate problems were highly correlated, $r = .80$, $p < .001$, and showed the exact same pattern of results in all of the analyses, they were combined and labeled easy problems.⁴ Accuracy scores were tabulated separately for easy and difficult MA problems by dividing the number of problems answered correctly by the total number of problems, with greater scores indicating greater accuracy on these problems.

Results

Learning. A 2 (stereotype threat) \times 2 (placement of manipulation) ANOVA was conducted on the explanation of MA scores. The results showed a significant main effect of stereotype threat, $F(1, 88) = 7.18$, $p = .01$, $\eta_p^2 = .08$; however, this effect was qualified by a significant two-way interaction, $F(1, 88) = 4.15$,

⁴ All of the significant effects with this measure that combine easy and moderate problems are significant when the easy and moderate problems are examined separately. Further, the correlation between the easy problems and the difficult problems, $r = .35$, $p = .001$, and the correlation between the moderate problems and the difficult problems, $r = .39$, $p < .001$, were much lower than the correlation between the easy and moderate problems, $r = .80$, $p < .001$.

$p = .045$, $\eta_p^2 = .05$. There was a significant simple effect of stereotype threat when it was presented before learning, $F(1, 88) = 14.68$, $p < .001$, $\eta_p^2 = .16$, showing that women in the control condition provided better explanations of MA ($M = 4.25$) than did women in the stereotype threat condition ($M = 3.26$). There was no simple effect of the stereotype threat manipulation on explanations of MA when it was presented after learning, $F(1, 88) = 0.17$, $p = .68$, $\eta_p^2 = .01$ (control $M = 4.19$, stereotype threat $M = 4.05$).

Binary logistic regression was used to examine the effect of stereotype threat (with the control condition coded -1 and the stereotype threat condition coded 1), the placement of the manipulation (with placement of the manipulation after learning coded -1 and placement before learning coded 1), and their interaction (multiplicative function) on the percentage of participants who were able to correctly define *mod*. The regression showed a tendency toward the expected two-way interaction, $B = -0.37$, $SE = 0.22$, $Wald(1) = 2.86$, $p = .09$. When the manipulation was presented before learning, there was a significant effect of the stereotype threat manipulation on participants' ability to define *mod*, $B = -0.62$, $SE = 0.28$, $Wald(1) = 4.90$, $p = .03$ (control = 61%, stereotype threat = 31%), but there was no effect of stereotype threat on the ability to define *mod* when the manipulation was presented after learning, $B = 0.05$, $SE = 0.34$, $Wald(1) = 0.02$, $p = .88$ (control = 50%, stereotype threat = 53%).⁵

Math performance. Performance on the MA problems was examined by conducting a 2 (stereotype threat) \times 2 (placement of manipulation) \times 2 (problem difficulty: easy, difficult) mixed-model ANOVA. The results showed a main effect of problem difficulty, $F(1, 88) = 64.35$, $p < .001$, $\eta_p^2 = .42$; a main effect of stereotype threat, $F(1, 88) = 19.63$, $p < .001$, $\eta_p^2 = .18$; a Stereotype Threat \times Problem Difficulty interaction, $F(1, 88) = 5.54$, $p = .02$, $\eta_p^2 = .06$; and the predicted three-way interaction, $F(1, 88) = 4.38$, $p = .04$, $\eta_p^2 = .05$ (see Figure 2). To inspect the three-way interaction, we examined the Stereotype Threat \times Problem Difficulty interaction separately for the easy and the difficult MA problems. For easy MA problems, the predicted two-way interaction was significant, $F(1, 88) = 6.50$, $p = .01$, $\eta_p^2 = .07$. This interaction showed that participants who received the manipulation of stereotype threat before learning showed better performance on easy problems in the control condition than in the stereotype threat condition, $F(1, 88) = 11.83$, $p = .001$, $\eta_p^2 = .18$,

whereas those who received the manipulation of stereotype threat after learning did not show an effect of stereotype threat, $F(1, 88) = 0.15$, $p = .70$, $\eta_p^2 = .01$. For difficult problems, only the predicted simple effect of stereotype threat was significant, $F(1, 88) = 18.32$, $p < .001$, $\eta_p^2 = .17$ (control $M = .87$, stereotype threat $M = .74$).

We also examined the amount of time taken to provide an answer on the MA problems by conducting a Stereotype Threat \times Placement of Manipulation \times Problem Difficulty mixed-model ANOVA. There was a significant main effect of problem difficulty showing that easy problems were solved more quickly ($M = 5,218$ ms) than were difficult problems ($M = 7,902$ ms), $F(1, 88) = 267.11$, $p < .001$, $\eta_p^2 = .75$. There was also a significant Stereotype Threat \times Problem Difficulty interaction, $F(1, 88) = 5.26$, $p = .02$, $\eta_p^2 = .06$. The interaction was decomposed by looking at the effect of stereotype threat on easy and difficult problems separately. These analyses showed that stereotype threat did not impact response latencies for easy problems, $F(1, 88) = 0.87$, $p = .35$, $\eta_p^2 = .01$, or difficult problems, $F(1, 88) = 2.36$, $p = .13$, $\eta_p^2 = .03$. The interaction of stereotype threat and problem difficulty was due to stronger effects of problem difficulty on response latencies in the control condition, $F(1, 88) = 204.12$, $p < .001$, $\eta_p^2 = .83$ (easy $M = 5,410$ ms, difficult $M = 8,471$ ms), than in the stereotype threat condition, $F(1, 88) = 103.73$, $p < .001$, $\eta_p^2 = .69$ (easy $M = 5,026$ ms, difficult $M = 7,333$ ms). It is important to note that the three-way interaction was not significant, $F(1, 88) = 0.93$, $p = .34$, $\eta_p^2 = .01$, showing that the accuracy results were not due to a speed-accuracy trade-off.

Amount of time studying the instructions. We also conducted a Stereotype Threat \times Placement of Manipulation ANOVA on study time. There were no significant effects on this measure, $F_s < 1$. However, we wanted to assess if women who are under stereotype threat could reduce differences in learning by spending more time studying. To do this, we conducted two multiple regression analyses. Namely, we regressed (a) accuracy on the easy problems or (b) accuracy on the difficult problems onto the stereotype threat manipulation (with the control condition coded -1 and the stereotype threat condition coded 1), the placement of the manipulation (with placement of the manipulation after learning coded -1 and placement before learning coded 1), study time, all possible two-way interactions, and the three-way interaction.

For easy problems, there were several effects, but they were qualified by the significant three-way interaction, $\beta = .71$, $p = .02$. To examine this interaction, we regressed accuracy on the easy MA problems onto the stereotype threat manipulation, study time, and their interaction separately in the manipulation before learning and the manipulation after learning conditions. The two-way interaction was significant in the manipulation before learning condition, $\beta = 1.20$, $p = .01$ (see Figure 3A), but not in the manipulation after learning condition, $\beta = .08$, $p = .89$ (see Figure 3B). In the manipulation before learning condition, the simple slope

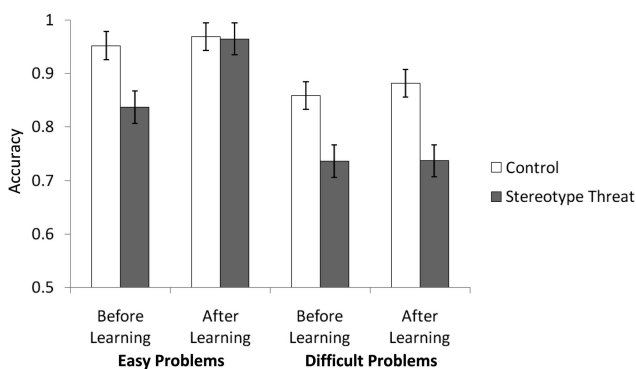


Figure 2. Accuracy on the modular arithmetic task as a function of stereotype threat manipulation, placement of manipulation, and problem difficulty in Experiment 2.

⁵ These numbers likely underestimate the number of participants that knew the correct definition of *mod*. This is because over 20% of participants in all of the experimental conditions answered this question with “modular” or “modulo,” as these were described in the instructions as the words from which *mod* was abbreviated, and these answers were scored as incorrect.

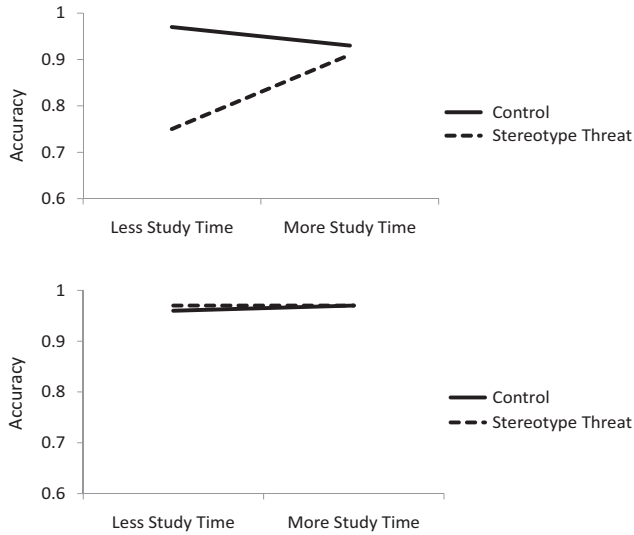


Figure 3. Accuracy on easy problems as a function of study time, stereotype threat, and placement of the manipulation in Experiment 2. Study time was plotted on the x-axis as less study time (1 standard deviation below the mean) or more study time (1 standard deviation above the mean). The top panel shows the results when the manipulation was presented before learning, and the bottom panel shows the results when the manipulation was presented after learning.

between study time and accuracy on easy MA problems was significant in the stereotype threat condition, $\beta = .46, p = .01$; however, the simple slope between study time and accuracy on easy MA problems was not significant in the control condition, $\beta = -.30, p = .12$. The results from the regression analyses show that when stereotype threat was presented before learning, increased study time led to increased performance. This relation was not seen in any of the other conditions ($ps > .12$). For the difficult MA problems, the only effects to obtain were the main effects of stereotype threat, $\beta = -.25, p < .001$, and study time, $\beta = .20, p = .03$: Performance on difficult MA problems was better in the control condition than in the stereotype threat condition, and as participants spent more time studying, their performance on difficult MA problems improved.

Mediational analyses. To examine the hypothesis that reduced performance on easy MA problems was due to reduced knowledge about how to complete them, we conducted mediational analyses. Specifically, we assessed the extent to which explanation of MA scores mediated the relation between the interaction of stereotype threat and placement of the manipulation and accuracy on easy MA problems.⁶ As reported above, the Stereotype Threat \times Placement of the Manipulation interaction was significant for both explanation of MA scores, $\beta = -.20, p = .04$, and accuracy on easy MA problems, $\beta = -.24, p = .01$. Further, the relation between explanation of MA scores and accuracy on the easy MA problems was significant, $\beta = .72, p < .001$. However, when accuracy on easy MA problems was regressed on the interaction term and explanation of MA scores, the interaction term no longer predicted accuracy on easy MA problems, $\beta = -.12, p = .12$, and a Sobel test obtained, $z = -2.00, p = .046$ (Baron & Kenny, 1986).

Discussion

Experiment 2 provided additional evidence that stereotype threat can reduce mathematical learning when a different, well-established mathematical task is learned: MA. First, when the stereotype threat manipulation was presented before learning, women receiving the stereotype threat instructions, compared with women in a control condition, showed reduced ability to explain how to solve MA problems and define *mod*. When the stereotype threat manipulation was presented after learning, women who received stereotype threat instructions and those that received control instructions were equally proficient at explaining MA and defining *mod*. Second, performance on easy MA problems was only reduced in the stereotype threat condition, relative to the control condition, when the stereotype threat manipulation was given before learning. When the stereotype threat manipulation was presented after learning, there was no difference in performance on easy MA problems between the stereotype threat and the control conditions (Beilock et al., 2007). The results for the easy MA problems are important because they show that even in cases where the skills necessary for performance are not harmed by stereotype threat (i.e., the math is rudimentary, using few online cognitive resources), performance deficits can occur when people attempt to learn these skills while experiencing stereotype threat.

Third, we demonstrated that performance on easy MA problems could be improved by increasing the amount of time spent studying how to solve MA problems when stereotype threat instructions were presented before learning. This suggests that stereotype threat-based learning deficits can be overcome by increased motivation to study and understand the mathematical rules and operations necessary to complete the math task. This also suggests that stereotype threat may reduce encoding by decreasing some women's attention to math-related materials (e.g., Major et al., 1998). Fourth, for easy MA problems, the mediation of the Stereotype Threat \times Placement of Manipulation interaction by the ability to explain how to complete MA problems strongly suggests that stereotype threat reduces performance on easy MA problems by inhibiting learning. When considered together, these four pieces of evidence provide strong support for our prediction that stereotype threat impairs learning.

In this experiment, we also replicated the findings of Beilock et al. (2007) by showing that women who received the stereotype threat manipulation (before or after learning) performed more poorly than the women in the control condition on difficult MA problems. Thus, the impact of stereotype threat on difficult MA problems was primarily due to stereotype threat-based performance effects (Beilock et al., 2007). This disconnect between the processes underlying performance on easy and difficult MA problems suggests that learning and performance can be dissociated using this task. Therefore, the findings from this experiment pro-

⁶ The same contrast coding used in the previous regression analyses was used for the mediational analyses. For all of the mediational analyses in Experiment 2 that examined the interaction between stereotype threat and placement of the manipulation, the main effects of these manipulations were included as variables in the regression analyses. Additionally, because the two-way interaction of stereotype threat and placement of the manipulation was not significant for difficult MA problems, we did not test for mediation on difficult MA problems.

vide compelling evidence that stereotype threat effects can be due to reduced learning of mathematical rules (i.e., easy MA problems when stereotype threat was induced before learning) or reduced online performance of the mathematical operations (i.e., difficult MA problems).

Experiment 3

In Experiment 3, we moved away from examining the detrimental effect of stereotype threat on learning on relatively concrete mathematical tasks to explore if stereotype threat-based learning effects can be captured on an abstract symbolic logic task that utilizes several mathematical principles (e.g., Kaminski, Sloutsky, & Heckler, 2008). Not only is this task internally consistent and based on logic, it uses more complicated mathematical principles. Instead of having participants solve equations involving numbers, this task only relies on participants using mathematical reasoning. The use of an abstract symbolic logic task is advantageous for examining if stereotype threat reduces learning because the mathematical principles that underlie this task can be applied to novel tasks. Therefore, we can assess people's learning differently, by measuring their ability to apply the abstract knowledge from the focal task (i.e., the learned task) to a transfer task (i.e., a task using the same mathematic principles as the focal task but for which there was no task-specific learning). This allows us not only to examine if stereotype threat reduces women's ability to learn the focal task but also to assess if the impact of stereotype threat is exacerbated when abstract knowledge from the focal task must be generalized to the transfer task.

There is substantial evidence that learning abstract concepts results in a richer and more elaborate schematic representation of the knowledge learned (e.g., Gick & Holyoak, 1983; Novick & Holyoak, 1991) and this more elaborate schematic representation increases people's ability to apply or transfer this knowledge to novel domains (e.g., Catrambone & Holyoak, 1989; Goldstone & Sakamoto, 2003; Kaminski et al., 2008). In mathematics, Kaminski and colleagues showed that learning mathematical principles based on abstract examples led to high levels of transfer from one task to another, with essentially no transfer deficits (i.e., the same level of learning was displayed on the focal and the transfer tasks). We examined whether stereotype threat, in addition to reducing initial mathematical learning, inhibited transfer of this learning to novel domains using a learning paradigm that involved abstract examples. Given the results of Kaminski et al. (2008), we did not expect women in the control condition to show reduced learning on the transfer task relative to the focal task. However, consistent with Experiments 1 and 2, we expected that stereotype threat would reduce women's ability to learn the mathematical principles in the focal task. Moreover, we expected that women under stereotype threat would show additional deficits in generalizing the information learned in the focal task to the transfer task.

Deficits in transferring information from the focal task to the transfer task should lead women under threat to show poorer learning on the transfer task than the focal task. We predict this poorer transfer task learning for women under threat because the transfer additionally requires participants to take the schematic knowledge of mathematical principles garnered from the focal

task, apply those rules to the transfer task, and map these schematic representations onto the novel symbols or elements of the transfer task despite superficial differences between the tasks. Put differently, women under threat should have greater difficulty applying knowledge gained in an initial domain to a second domain, indicating lower levels of learning about and understanding of the mathematical principles underlying both tasks (e.g., Goldstone & Sakamoto, 2003). Women under threat, relative to those not under threat, should have more poorly elaborated schemas of the mathematical principles learned in the focal task and have difficulty generalizing the schematic knowledge of the principles they obtained from learning the focal task to the transfer task, leading to more pronounced learning deficits on the transfer task.

In this experiment, we also assessed men's mathematical learning and ability to transfer mathematical learning to a new domain. Because we did not expect men to feel threatened by our stereotype threat manipulation, we did not expect stereotype threat to affect men's ability to learn or transfer the mathematical principles on the abstract symbolic logic task. If men do not show any effects of our manipulation, this would demonstrate that the impact of stereotype threat on mathematical learning is restricted to members of the stereotyped group. Given our predictions regarding men and women, we predicted an interaction between stereotype threat and gender for both focal task learning and transfer task learning. Women in the stereotype threat condition, compared with women in the control condition, should have lower levels of focal task and transfer task learning, but men's focal task and transfer task learning should not be influenced by the stereotype threat manipulation. However, because we expect the transfer task to be learned more poorly by women under stereotype threat than by women with the focal task, we also predicted an interaction of stereotype threat, gender, and task (focal vs. transfer). Specifically, women in the stereotype threat condition should show a more pronounced deficit in transfer task learning than in focal task learning, whereas women in the control condition and men in both the control and the stereotype threat conditions should show no differences in learning between the focal and transfer tasks.

In addition to including men in Experiment 3, we introduced an implicit measure of learning to assess learning without relying on mathematical performance in any way. We expected an association in memory between the symbols that were linked together by the rules of the focal task not only because participants learned the relationships between these symbols during the description of the task but also because they repeatedly used these pairings when completing measures of learning for both the focal task and the transfer task (see Gawronski & Bodenhausen, 2007; Strack & Deutsch, 2004). Our implicit learning measure, therefore, compared the extent to which people associated abstract symbols that were linked together by the rules of the focal task with the extent to which people associated abstract symbols that were used in the focal task but not linked together by the rules of the task (see below). If our main hypothesis that stereotype threat reduces women's ability to learn mathematical principles is correct, we would expect a weaker association between symbols that are part of the same rule (i.e., less elaborated schematic knowledge about the focal

task) for women in the stereotype threat condition than women in the control condition. Also, we would expect strong and equivalent associations between symbols that are part of the same rule (i.e., more elaborate and richer schematic knowledge about the focal task) for men (regardless of condition) and women in the control condition. If these results obtain for our implicit learning measure, it strongly supports our main hypothesis that stereotype threat reduces women's level of mathematical learning, especially because these findings are difficult to explain with a performance-based hypothesis.

Method

Participants. Undergraduates ($N = 92$; 59 women and 33 men) participated for course credit. They were randomly assigned to either a control condition or a stereotype threat condition. Following Kaminski et al. (2008), we excluded the data from 11 participants who scored at or below chance (37.5%) on the focal learning task.⁷ This left us with a final sample of 81 participants (50 women and 31 men).

Procedure. Participants were informed that they would be taking part in an experiment on cognitive skills. They were assigned to either the control condition or the stereotype threat condition using the same instructions as were provided in Experiment 2. However, in the stereotype threat condition, participants were also informed that these tasks were "problem-solving tasks that involve using mathematical principles" in order to connect the abstract tasks to the stereotyped domain of math.

Focal task learning. The focal task was described as a task about how to decipher hieroglyphics from a language that used three symbols (see Kaminski et al., 2008). The ability to correctly utilize this language depended on participants understanding the concept of the commutative mathematical group of order 3 using the addition operation. This group, which consists of three distinct elements (e.g., x , y , z), is defined by several principles: adding any combination of these elements must equal one of the elements of the group (closure law), the order in which elements are added does not impact the result (commutative and associative laws), one of the elements acts as an identity (identity law; if z is the identity, $x + z = x$ and $y + z = y$), and every element has an inverse (inverse law; $x + y = z$, where z is the identity). These mathematical principles were implied by the specific rules for the focal task.

More specifically, participants were told (a) the language involved only three symbols (closure law), (b) the ordering of the symbols did not matter (commutative and associative laws), (c) a diamond paired with a flag equaled a diamond and a circle paired with a flag equaled a circle (identity law, with the flag as the identity), (d) a circle and a diamond equaled a flag (inverse law), (e) two circles equaled a diamond, and (f) two diamonds equaled a circle.⁸ After reading the rules, participants completed a set of 24 questions to test how well they learned to decipher the hieroglyphics (see the supplemental materials from Kaminski et al., 2008).

Transfer task learning. Next, participants completed a transfer task that used exactly the same underlying mathematical principles as the focal task but included different objects to represent the three elements and included a new cover story. In this task, participants were asked to predict the outcomes of a pointing

game. In this fictional game, one child pointed to certain objects, and, given the objects that the child had already pointed to, the other children guessed the final object that the pointer would point to, using certain rules to make their predictions. Participants were asked to play along with the children, who were guessing by predicting the final object after being presented with the already chosen objects on the computer screen. The three objects (i.e., the three elements of the task) were directly equivalent to the elements from the focal task: a broach (diamond), a bottle (circle), or a ring (flag). The participants learned that the rules of this pointing game were similar to the rules involved in the first task: (a) a bottle paired with a ring equaled a bottle, (b) a bottle and a broach equaled a ring, (c) two bottles equaled a broach, and (d) two broaches equaled a bottle. These rules underlying the transfer task implied the same mathematical principles as the rules of the focal task, just with different objects serving as elements. The instructions for the transfer task were extremely impoverished (in addition to reading the cover story about the pointing game and seeing the four rules presented above, participants were only told that "the rules of the last system you learned are like the rules of this game"), making participants reliant on past learning. Participants completed 24 questions that assessed their ability to learn the transfer task. These questions were mathematically equivalent to those used for the focal task (see the supplementary materials from Kaminski et al., 2008).

Implicit learning task. Participants completed an implicit learning task after the transfer task. This task was a sequential priming task that was based on the rules of the focal task. It was described as a reaction time task where participants indicated whether a target symbol was old or new by pressing the appropriate response key as quickly as possible. The old symbols were the symbols that were used in the focal task (i.e., diamond, circle, flag). The new symbols were novel symbols that were not used in the focal task or transfer tasks (i.e., star, rectangle, triangle). The

⁷ Nine of the participants excluded were women (six in the control condition and three in the stereotype threat condition), and two of the participants excluded were men (one in each condition).

⁸ Rules e and f are compatible. The closure law states that circle + circle has to equal one of the three elements (i.e., circle, diamond, or flag). If circle + circle = circle, then circle + circle = circle + flag because of the identity law. Subtracting a circle from each side, circle = flag is left, which is untrue because the elements are distinct from one another (i.e., they are not the same element). If circle + circle = flag, then circle + circle = circle + diamond because of the inverse law. Subtracting a circle from each side, circle = diamond is left, which is also untrue. Therefore, circle + circle must equal diamond. The same logic can be used to show that diamond + diamond = circle (Papantonopoulou, 2002). More concretely, let a flag equal 0, a circle equal 1, and a diamond equal 2. These numbers are the corresponding numerical values for elements in an additive group of order 3, with the identity being 0. To determine the resulting value of a combination of elements, the values are added and then divided by 3 because there are three elements in this group. The resulting remainder identifies what value (and therefore what symbol) is produced by the certain combination of symbols. For example, circle + circle becomes $1 + 1 = 2$, and 3 divides into 2 zero times with a remainder of 2. Because 2 equals diamond in our task, circle + circle = diamond. For another example, diamond + diamond becomes $2 + 2 = 4$, and 3 divides into 4 one time with a remainder of 1. Because 1 equals circle in our task, diamond + diamond = circle.

target symbols were preceded by a prime that consisted of two symbols, both of which were symbols used in the focal task (flag–diamond, flag–circle, circle–circle, diamond–diamond, or diamond–circle).

The implicit learning task consisted of 80 trials presented on the computer. Each trial consisted of a fixation cross presented for 1,000 ms, a prime presented for 200 ms, a blank screen for 25 ms, and a target symbol to identify as old or new that was presented until the participant responded. The target symbol was old on 40 of the trials and new on 40 trials. Because the new symbols were not connected to the primes by the rules of the focal task, we did not expect the primes and the new symbols to be associated in memory. Therefore, we did not expect any effects of the manipulations on the trials where the target was new because the prime and target were not related to one another.⁹ Instead, we expected the manipulations to impact responses on the trials where the target was old. Among the old trials, 20 trials were rule consistent—the target symbol was the learned outcome, based on learning for the focal task, of the two symbols presented as the prime (e.g., diamond–diamond followed by a circle, circle–circle followed by a diamond)—and 20 trials were rule inconsistent—the target symbol was not the learned outcome of the two symbols presented as the prime (e.g., diamond–diamond followed by a flag, circle–flag followed by a diamond).

On rule-consistent old trials, participants who learned the rules of the focal task were expected to make the old–new judgment more quickly than those who did not learn the rules as well because, after participants both initially learn the rules and complete 48 problems that involve these rules (24 focal problems and 24 transfer problems), the target symbol should be more strongly related to the prime symbol in memory as a result of stronger and more elaborated schematic knowledge about the rules of the focal task. On rule-inconsistent old trials, participants who learned the rules of the focal task were expected to make old–new judgments relatively more slowly than those who did not learn the rules as well because over the course of the experimental session, the target symbol should not be related to the prime symbol in memory and should not be part of the schematic knowledge acquired about the focal task. Thus, on old trials, people who have learned the rules of the focal task, as compared with those who did not learn the rules as well, should respond more quickly on rule-consistent trials than on rule-inconsistent trials. Implicit learning was indexed by subtracting the response latencies for the rule-consistent trials from the response latencies for the rule-inconsistent trials, with greater scores indicating a stronger association between the prime symbols and their learned outcome.

As a manipulation check, participants then completed Marx and Stapel's (2006) three-item measure of threat-based concern ($\alpha = .83$) on scales ranging from 1 (*strongly disagree*) to 7 (*strongly agree*); greater scores indicated greater concern about confirming negative gender stereotypes. Finally, participants indicated their gender.

Results

Threat-based concern. A 2 (stereotype threat) \times 2 (gender) ANOVA was conducted on participant's threat-based concern scores. The expected two-way interaction was obtained, $F(1, 77) = 4.34, p = .04, \eta_p^2 = .05$. There was no difference between

the control ($M = 1.88$) and stereotype threat conditions ($M = 1.79$) on threat-based concern for men, $F(1, 29) = 0.47, p = .83, \eta_p^2 = .002$; however, women in the control condition showed less threat-based concern ($M = 2.07$) than did women in the stereotype threat condition ($M = 3.24$), $F(1, 48) = 8.82, p = .005, \eta_p^2 = .16$.

Focal task learning and transfer task learning. The impact of the manipulations on focal task learning and the transfer of this learning was examined by conducting a 2 (stereotype threat) \times 2 (gender) \times 2 (task: focal task learning, transfer task learning) mixed-model ANOVA. This analysis yielded a significant main effect of gender, $F(1, 77) = 8.45, p = .005, \eta_p^2 = .10$; a significant interaction of stereotype threat and gender, $F(1, 77) = 9.45, p = .003, \eta_p^2 = .11$; a significant interaction of gender and task, $F(1, 77) = 11.97, p = .001, \eta_p^2 = .14$; and the predicted three-way interaction, $F(1, 77) = 5.53, p = .02, \eta_p^2 = .07$ (see Figure 4). To examine the three-way interaction, we conducted separate Gender \times Stereotype Threat ANOVAs for learning on the focal task and learning on the transfer task.

Focal task learning. The interaction between gender and stereotype threat was significant for the focal learning task, $F(1, 77) = 4.35, p = .04, \eta_p^2 = .053$. There was no difference between the control and stereotype threat conditions on focal task learning for men, $F(1, 29) = 0.45, p = .51, \eta_p^2 = .02$; however, women in the control condition showed greater learning than did women in the stereotype threat condition, $F(1, 48) = 6.79, p = .01, \eta_p^2 = .12$.

Transfer task learning. The interaction between gender and stereotype threat was significant for the transfer learning task, $F(1, 77) = 12.70, p = .001, \eta_p^2 = .14$. There was no difference between the control and stereotype threat conditions on transfer task learning for men, $F(1, 29) = 1.33, p = .26, \eta_p^2 = .04$; however, women in the control condition showed much greater learning than did women in the stereotype threat condition, $F(1, 48) = 19.16, p < .001, \eta_p^2 = .29$.

The three-way interaction of gender, stereotype threat, and task was due to the significant two-way interaction of stereotype threat and task for women, $F(1, 48) = 6.17, p = .02, \eta_p^2 = .11$, but not for men, $F(1, 29) = 1.20, p = .28, \eta_p^2 = .04$. This interaction demonstrates that women showed a stronger effect of stereotype threat on transfer task learning, $F(1, 48) = 19.16, p < .001, \eta_p^2 = .29$, than focal task learning, $F(1, 48) = 6.79, p = .01, \eta_p^2 = .12$. Indeed, there is a much larger effect of the stereotype threat manipulation on women for the transfer task relative to the focal task, $t(26) = 4.26, p < .001, d = 0.82$. There was no difference in accuracy scores for the focal and transfer tasks in any of the other conditions ($ps > .10$).

Implicit learning task. A Gender \times Stereotype Threat ANOVA was conducted on scores from the implicit learning task. The only effect to obtain was the predicted two-way interaction, $F(1, 77) = 7.14, p = .01, \eta_p^2 = .09$. There was not a significant difference between men in the control condition ($M = 108$ ms) and the stereotype threat condition ($M = 175$ ms) on this measure, $F(1, 29) = 0.59, p = .45, \eta_p^2 = .02$. Women in the control condition showed much greater implicit learning ($M = 173$ ms) than did

⁹ Consistent with this expectation, a Gender \times Stereotype Threat ANOVA was conducted on response latencies from the new trials of the implicit learning task and showed no significant effects, $F_s < 1$. The new trials are, therefore, not discussed further.

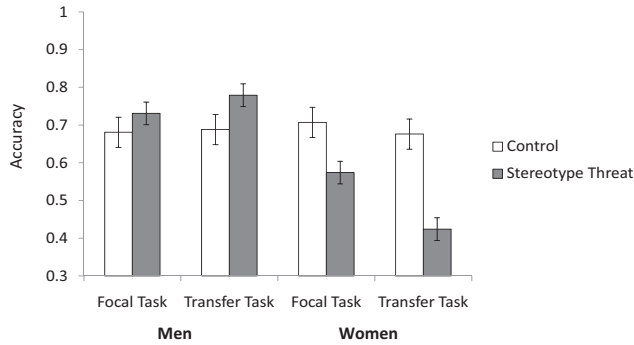


Figure 4. Accuracy on the focal task and transfer task as a function of the stereotype threat manipulation and participant gender in Experiment 3. Chance performance is an accuracy of score of .375.

women in the stereotype threat condition ($M = -10$ ms), $F(1, 48) = 13.43$, $p = .001$, $\eta_p^2 = .22$.

Amount of time studying the instructions. We also conducted a Stereotype Threat \times Gender ANOVA on the average amount of time spent studying the eight instruction screens for the focal task ($M = 18.15$ s, $SD = 6.46$). There was only a tendency toward a main effect of gender, $F(1, 77) = 3.19$, $p = .08$, $\eta_p^2 = .04$, showing that women ($M = 19.19$ s) spent somewhat more time than men ($M = 16.49$ s) studying the instruction screens. There were no other effects on this measure, $F_s < 1.5$, $p_s > .22$. We also examined the extent to which the average amount of time spent studying predicted focal task learning, transfer task learning, and implicit learning. On the basis of the study time results from Experiment 2, we expected to see a positive and significant correlation between study time and focal task learning, transfer task learning, and implicit learning for women in the stereotype threat condition. Indeed, for women in the stereotype threat condition, there was a significant positive correlation between study time and focal task learning, $r = .43$, $p = .03$, and between study time and transfer task learning, $r = .58$, $p = .002$. The relation between study time and implicit learning was not significant in this condition, $r = -.05$, $p = .82$. Additionally, study time was uncorrelated to focal task learning, transfer task learning, and implicit learning in any of the other experimental conditions, $r_s < .24$, $p_s > .33$. The three-way interaction (multiplicative function) between study time, gender (coded -1 for men and coded 1 for women), and stereotype threat (coded -1 for control and coded 1 for stereotype threat) was significant for focal task learning, $\beta = .25$, $p = .05$, but not for transfer task learning, $\beta = .13$, $p = .31$, or implicit learning, $\beta = .05$, $p = .72$.

Discussion

The results of Experiment 3 conformed exactly to our predictions. We showed that stereotype threat reduced women's, and not men's, ability to learn an abstract task that was based on mathematical principles (i.e., the focal task). Stereotype threat also inhibited women's ability to transfer knowledge they had gained about the focal task to the transfer task, producing even greater stereotype threat effects for women on the transfer task than on the focal task. In addition, women showed less implicit learning in the stereotype threat condition than in the control condition, whereas

men showed no differences in implicit learning as a function of the stereotype threat manipulation. The results on the implicit learning measure suggest that reduced learning of the transfer task relative to the focal task for women under threat was due, in large part, to reduced encoding of the mathematical principles necessary to complete the transfer task. The results of this experiment strongly support our main hypothesis that stereotype threat reduces mathematical learning because these findings cannot be easily explained by a performance-based account of stereotype threat effects.

General Discussion

This work shows the effect of stereotype threat for a new domain for women in math: learning. More specifically, this research shows that stereotype threat harms more than just the execution of skills in the stereotyped domain by demonstrating that stereotype threat reduces learning, these reductions in learning cannot be accounted for by stereotype threat-based performance decrements, and reduced learning can cause reduced performance. In short, this work reveals that stereotype threat also harms the acquisition of the skills necessary for successful performance.

Past research on stereotype threat provides compelling evidence that stereotype threat leads to poor performance by inhibiting the execution of a learned skill (e.g., Schmader et al., 2008). Stereotype threat triggers worries about confirming the stereotype, increased arousal, and reduced working memory that lead to poor performance. In-the-moment pressures caused by stereotype threat lead to choking under pressure (Beilock, 2008). This is one process through which stereotype threat can reduce performance. We provide evidence for another process: by inhibiting learning in the stereotyped domain. Stereotype threat also harms performance when it reduces the encoding of novel information that is necessary for skill execution. Because stereotype threat inhibits learning and performance on a task is unlikely to be successful without sufficient learning, this constitutes another way that stereotype threat can hurt performance.

The three experiments presented in this work provide considerable evidence of the latter process through which stereotype threat harms performance (i.e., through reductions in learning). We predicted that activating the negative stereotype that "women are bad at math" would induce worries about confirming the stereotype (e.g., Marx & Stapel, 2006). This worry should reduce working memory capacity (e.g., Schmader et al., 2008), increase arousal (e.g., Murphy et al., 2007), and decrease motivation to learn (e.g., Steele, 1997), which should then inhibit the encoding of mathematical information (e.g., mathematical principles and operations). Reduced encoding of mathematical information should lead to reduced performance. However, future researchers should examine the exact mechanism through which stereotype threat inhibits the encoding of information from the stereotyped domain. Reduced working memory and reduced motivation may, at some times, work in conjunction to produce stereotype threat-induced learning deficits, but at other times one of these process may be more important for understanding the learning deficits demonstrated in the current work. Future researchers should also examine when specific mechanisms account for reduced learning under stereotype threat.

In this work, we showed that women in a stereotype threat condition had greater threat-based concern compared with women in a control condition (Experiment 3). In addition, we showed that stereotype threat inhibited encoding of mathematical information. Specifically, we revealed the relation between stereotype threat and reduced encoding by showing that information learned before stereotype threat was induced was better recalled than information learned under stereotype threat (Experiment 1), mathematical operations were better learned when stereotype threat was induced after learning (Experiment 2), women's learning when under stereotype threat improved to the extent that they spent a longer time studying the mathematical information (Experiments 2 and 3), and an associative measure of learning showed little learning when women were under stereotype threat (Experiment 3). Further, reduced learning of mathematical information caused by stereotype threat led to poorer math performance (Experiments 1 and 2).

Given our results, it seems useful to briefly consider (a) the impact that negative performance stereotypes can have on learning and thus on performance and (b) the implications of stereotype threat-based learning effects for stereotype threat research. If, as this research suggests, stereotype threat reduces learning for stigmatized individuals in negatively stereotyped domains, it could mean that stereotype threat need only be present during learning and not during performance to impact performance. Indeed, these results provide initial evidence for the idea that negatively stereotyped people do not need to be under stereotype threat during performance for performance to be affected by it. This could mean that even in testing situations that are free of stereotype threat (i.e., women taking a math test in a room full of women; e.g., Inzlicht & Ben-Zeev, 2000), performance could suffer if learning occurred in a situation that evoked stereotype threat (e.g., a math class with more men than women). Thus, stereotype threat may not always be directly responsible for performance failure in the sense that stereotype threat is interfering with online task execution via reduced working memory capacity (e.g., Schmader et al., 2008). Instead, stereotype threat may lead to reduced performance by hindering learning of the skills necessary for task execution.

As mentioned earlier, models of stereotype threat generally assume equivalent levels of learning and attribute performance failures to the impact of negative stereotypes on skill execution (e.g., Beilock, 2008; Schmader et al., 2008). Given this assumption, the impact of this work for stereotype threat research could potentially be far reaching. The idea that learning and performance can both be reduced by stereotype threat (perhaps partially independent of one another) identifies many more areas and potential issues that underrepresented group members may have to deal with to succeed in domains in which they are negatively stereotyped. This conclusion has important social implications, most notably in the domain of education. Indeed, it may be that any situation involving skill acquisition is vulnerable to the negative effects of stereotype threat. If true, it is logical to assume that improvements in learning environments and teaching techniques would enhance performance in the stereotyped domain. Understanding the impact of stereotype threat on learning and structuring educational interventions to combat its impact could lead to greater success for stigmatized individuals in domains for which they are historically underrepresented (e.g., women in math, science, and engineering).

In addition, research on the impact of stereotype threat on learning could reveal why projects and interventions that exclu-

sively focus on performance under threat might not always work. This work shows that there may be more to the stereotype threat-performance phenomena than was originally thought. It also suggests that researchers should examine and more fully account for reduced learning caused by negative stereotypes in future research.

Conclusion

Although the impact of stereotype threat on both learning and performance makes the prospect of intervening to reduce its impact seem difficult, creating environments that can reduce the impact of stereotype threat during the acquisition of skills is critical for facilitating the entrance of group members into domains in which they have been historically absent. Because the impact of stereotype threat on performance can be lessened by a plethora of different manipulations, it is conceivable that learning environments that are explicitly structured to eradicate stereotype threat could be created. Stereotype threat research has advanced quickly by identifying and explaining the impact that this threat has on performance; future research in which the mechanisms underlying the impact of stereotype threat on learning are examined may help researchers more fully understand the phenomenon of stereotype threat and devise new and successful strategies to combat it. Knowing that stereotype threat reduces learning makes it more pervasive and insidious, indicating that there is much left to learn about stereotype threat and how to eradicate its influence.

References

- Adams, G., Garcia, G. M., Purdie-Vaughns, V., & Steele, C. M. (2006). The detrimental effects of a suggestion of sexism in an instruction situation. *Journal of Experimental Social Psychology, 42*, 602–615. doi:10.1016/j.jesp.2005.10.004
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182. doi:10.1037/0022-3514.51.6.1173
- Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science, 17*, 339–343. doi:10.1111/j.1467-8721.2008.00602.x
- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*, 256–276. doi:10.1037/0096-3445.136.2.256
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*, 174–181. doi:10.1016/j.jesp.2003.11.007
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1147–1156. doi:10.1037/0278-7393.15.6.1147
- Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006, September 1). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*, 1307–1310. doi:10.1126/science.1128317
- Cohen, G. L., Garcia, J., Purdie-Vaughns, V., Apfel, N., & Brzustoski, P. (2009, April 17). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science, 324*, 400–403. doi:10.1126/science.1170769
- Croizet, J. C., Despres, G., Gauzins, M., Hugeot, P., & Leyens, J. (2004). Stereotype threat undermines performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30*, 721–731. doi:10.1177/0146167204263961

- Eysenck, M. W. (1976). Arousal, learning, and memory. *Psychological Bulletin*, 83, 389–404. doi:10.1037/0033-2909.83.3.389
- Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition*, 25, 687–717. doi:10.1521/soco.2007.25.5.687
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1–38. doi:10.1016/0010-0285(83)90002-6
- Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, 46, 414–466. doi:10.1016/S0010-0285(02)00519-4
- Grimm, L. R., Markman, A. B., Maddox, W. T., & Baldwin, G. C. (2009). Stereotype threat interpreted as regulatory mismatch. *Journal of Personality and Social Psychology*, 96, 288–304. doi:10.1037/a0013463
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, 11, 365–371. doi:10.1111/1467-9280.00272
- Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, 137, 691–705. doi:10.1037/a0013834
- Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2008, April 25). The advantages of abstract examples in learning math. *Science*, 320, 454–455. doi:10.1126/science.1154659
- Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin*, 24, 34–50. doi:10.1177/0146167298241003
- Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology*, 42, 236–243. doi:10.1016/j.jesp.2005.04.010
- Marx, D. M., & Stapel, D. (2006). Distinguishing stereotype threat from priming effects: The role of the social self and threat-based concerns. *Journal of Personality and Social Psychology*, 91, 243–254. doi:10.1037/0022-3514.91.2.243
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, 18, 879–885. doi:10.1111/j.1467-9280.2007.01995.x
- Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 398–415. doi:10.1037/0278-7393.17.3.398
- O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, 29, 782–789. doi:10.1177/0146167203029006010
- Papantonopoulou, A. (2002). *Algebra: Pure and applied*. Upper Saddle River, NJ: Prentice Hall.
- Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, 57, 55–71. doi:10.1111/0022-4537.00201
- Rumelhart, D. E., & Todd, P. M. (1990). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 3–30). Cambridge, MA: MIT Press.
- Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology*, 96, 949–966. doi:10.1037/a0014846
- Rydell, R. J., Shiffrin, R., Boucher, K. L., Van Loo, K., & Rydell, M. T. (2010). Stereotype threat prevents perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 14042–14047.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440–452. doi:10.1037/0022-3514.85.3.440
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat on performance. *Psychological Review*, 115, 336–356. doi:10.1037/0033-295X.115.2.336
- 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. doi:10.1006/jesp.1998.1373
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613–629. doi:10.1037/0003-066X.52.6.613
- 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. doi:10.1037/0022-3514.69.5.797
- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). New York, NY: Academic Press.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220–247. doi:10.1207/s15327957pspr0803_1
- Unsworth, N., & Engle, R. W. (2005). Individual differences in working memory capacity and learning: Evidence from a serial reaction time task. *Memory & Cognition*, 33, 213–220.
- Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, 127, 797–826. doi:10.1037/0033-2909.127.6.797

Received August 25, 2009
 Revision received July 8, 2010
 Accepted July 12, 2010 ■

Instructions to Authors

For Instructions to Authors, please consult the November 2010 issue of the volume or visit www.apa.org/pubs/journals/psp and click on the "Instructions to authors" tab in the Journal Info box.