Stereotype threat prevents perceptual learning

Robert J. Rydell, Richard M. Shiffrin, Kathryn L. Boucher, Katie Van Loo, and Michael T. Rydell[1]

Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 45407

Edited* by Claude M. Steele, Columbia University, New York, NY, and approved June 28, 2010 (received for review March 4, 2010)

Stereotype threat (ST) refers to a situation in which a member of a group fears that her or his performance will validate an existing negative performance stereotype, causing a decrease in performance. For example, reminding women of the stereotype "women are bad at math" causes them to perform more poorly on math questions from the SAT and GRE. Performance deficits can be of several types and be produced by several mechanisms. We show that ST prevents perceptual learning, defined in our task as an increasing rate of search for a target Chinese character in a display of such characters. Displays contained two or four characters and half of these contained a target. Search rate increased across a session of training for a control group of women, but not women under ST. Speeding of search is typically explained in terms of learned "popout" (automatic attraction of attention to a target). Did women under ST learn popout but fail to express it? Following training, the women were shown two colored squares and asked to choose the one with the greater color saturation. Superimposed on the squares were task-irrelevant Chinese characters. For women not trained under ST, the presence of a trained target on one square slowed responding, indicating that training had caused the learning of an attention response to targets. Women trained under ST showed no slowing, indicating that they had not learned such an attention response.

gender | math | visual search

Across all cultures that have been examined there is a persistent stereotype held by both men and women that "women are bad at math." This stereotype is consequential: Women's math achievement and performance is lower in cultures where this stereotype is stronger (1). Although there are many reasons why women underperform at math, one important factor that affects not only math performance but also career choices and career achievement is stereotype threat: The mental and behavioral states that accompany the activation of this stereotype in women (2). Most notably, when this stereotype is activated in women's minds, they worry about confirming this pejorative stereotype and their math performance drops (3–5). Similar findings exist for other performance stereotypes held about other populations (5–7). The reasons for the performance drop are not entirely clear. Perhaps some kind of general anxiety causes a general decrement in women's math performance (8). Stereotype threat (ST) has been shown to increase arousal (9), increase negative thoughts (3, 10), decrease cognitive resources (5), and increase the perseverance of incorrect problem-solving strategies (11–13); all of these reactions partially explain performance decrements. Given the varied negative responses to threat, it seems a natural extrapolation to guess that it might inhibit women's ability to learn new information. That theory is the target of the present research.

Demonstrating that ST reduces or inhibits learning is important because it broadens the domain of ST and provides a specific unstudied mechanism by which ST impacts performance. Not only might ST hurt women's performance of a well-learned math skill when they are under pressure, it also might inhibit their ability to learn math skills. A failure to learn would then reduce their performance on other occasions when they are not experiencing threat. Distinguishing the effects of ST on generalized performance from effects on learning is not trivial. For example, eliminating ST reduced the achievement gap between Caucasian and African American middle-school students over a 2-y period (14). However, this narrowing of the achievement gap could have been a result of improved performance associated with reductions in ST, or it could have been because of increased learning after ST was reduced. All tasks, whether or not they involve learning, have components that could reflect performance differences (15). Because ST can affect women's math performance devoid of any learning deficits (3–5), to make conclusions about how ST affects learning, it is essential to clearly delineate the impact of ST on learning from its impact on performance.

Distinguishing learning effects from other performance effects is not easy in mathematical tasks. We therefore used a well-studied task, visual search, for which learning is routinely observed, for which the types and causes of learning are understood, and for which the learning component can be isolated from performance components. In visual search a target character is shown first, followed by a display of several characters, one of which might be the target (it is present on half the trials). Women's decisions about character presence are highly accurate and the primary measure of interest is the time to make the response. The task we use is known as "consistent mapping": The target characters remain targets, and the foil characters remain foils, across all trials. Under these circumstances, two kinds of learning can take place: (*i*) The observer learns to attend automatically to the target characters, so that attention moves directly to the target and extended search becomes unnecessary (16). The effect is often called "popout," in analogy to the way that a red target appears to pop out from green foils. (*ii*) The observer learns to unitize a character, so that each comparison of display item to test item involves a single comparison, whereas before learning a single comparison might involve several comparisons of features (17).

Three visual-search studies used an experimental group of women exposed to ST instructions (reminding them of the negative stereotype about women's math and visual processing performance) and a control group of women not given such instructions. The stimuli were initially unknown Chinese characters, 5 always used as targets and 195 always used as foils. One target preceded the display, and the display consisted of either two or four characters (Fig. 1). Half the displays contained targets. The roughly 75 min-long training session consisted of several blocks of 80 trials, with the threat group receiving ST instructions at the outset, and a shortened version of these instructions being reiterated between blocks. There were six blocks in exp. 1 and eight in exp. 2 and 3.

Performance in visual-search tasks without learning (for example when targets and foils exchange roles from trial to trial) is often characterized by serial self-terminating search: Display items are compared with the test item one at a time successively, at a fixed average rate. The search terminates when the target is reached (target-present trials) or when all items have been

Author contributions: R.J.R., R.M.S., and M.T.R. designed research; R.J.R., K.L.B., K.V.L., and M.T.R. performed research; R.J.R. analyzed data; and R.J.R., R.M.S., K.L.B., and K.V.L. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: rjrydell@indiana.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1002815107/-/DCSupplemental.

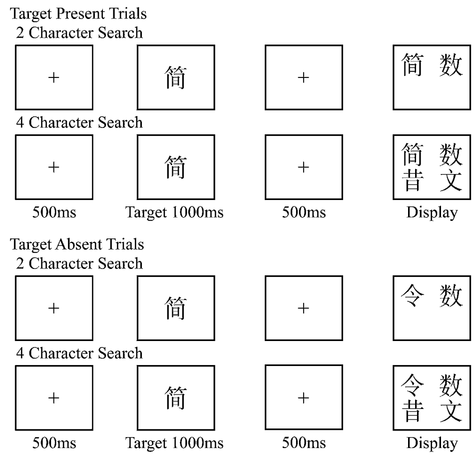

Fig. 1. Ordering, timing, and examples of the four different types of trails in the visual search task.

compared (target-absent trials). For target-absent trials the slope is an estimate of the comparison time. For target-present trials the target is reached on average half-way through the display, so the slope is halved. In most visual-search tasks the results do not match the predictions of this model exactly, but it is common in the field to calculate comparison times on the basis of this model, and use the reduction in these times as a measure of learning (16, 17). We adopt this convention as well. As is often the case in other studies, serial self-terminating search is not a perfect description of our results, either at the outset of training or for the ST group (for example, the slopes do not exhibit a 2:1 ratio for target-absent to target-present trials). Nonetheless, we will track slopes for the two groups to assess the presence of learning. Letting observed response time be denoted T, base time B, comparison time per character C, and display size n, we have $T = B + nC$ for target-absent trials and $T = B + [(n + 1)/2]C$ for target-present trials. The comparison time C is the slope, and serves as an estimate of rate of search. For target-absent and -present trials, the respective estimates of C are $[T(4) - T(2)]/2$ and $T(4) - T(2)$. Prior research has shown that C drops with training, thereby indicating learning (16, 17), and the greater the drop in C the greater is the learning that is inferred. It is also often useful to estimate the base time B, indicating the part of the response time other than that used to carry out comparisons (perception time, motor-response time), which for target absent and present is estimated respectively as $2T(2) - T(4)$ and $T(2) - (3/2)C$. A reduction in base time (or B) across blocks indicates increased performance for the components of responding that are not related to learning.

Exp. 1 (Dataset S1) compared visual search for women randomly assigned to receive control instructions or ST instructions. Fig. 2 *A* to *C* gives the estimates of slope or *C* for successive blocks, averaged across women in each of the two groups, for target-present, -absent, and -combined trials. Responses for women in these conditions start together but quickly diverge. The control group showed typical learning, with a steady decrease in slope. The women made aware of the stereotype that "women are bad at math" showed no change in slope. Perhaps these average results fail to reveal that some women responded to ST and some did not. We therefore examined the slope results by individual. Fig. 3 shows individual estimates of slope for the first and second half of the blocks, and the difference between the two. Clearly, most control women learned: slope dropped between the first and second half of training. The women under threat did not on average exhibit decreased slope from the first to the second half of training. Particularly interesting, the variability of the slope changes did not differ for the women in the two groups. If some women had failed to respond to the threat manipulation and learned, and others had responded so strongly that they increased their slope across blocks, then the threat group would have exhibited larger variance. Thus, it seems that most women were responding similarly to the threat manipulation. Because the threat instructions were reiterated before each block, it is possible that

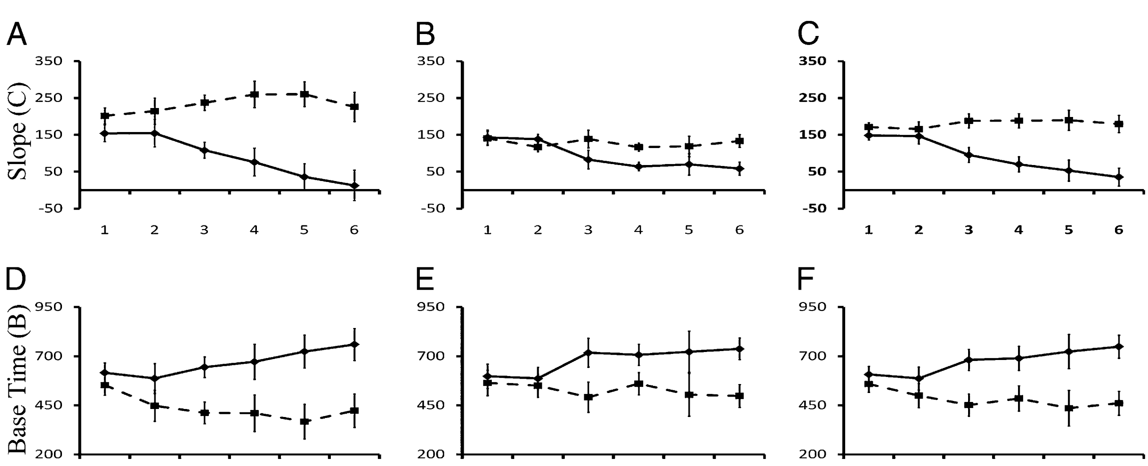

Fig. 2. The slope or *C* (measure of leaning) and base-time or *B* (measure of performance not based on learning) results for exp. 1 as a function of ST and training (block). The results are plotted separately for target-present trials, target-absent trails, and these trials combined. Results for the control condition are shown with solid lines (and diamonds) and results for the ST condition shown with dashed lines (and squares). The slope showed the same pattern of results for the present (*A*), absent (*B*), and combined trials (*C*). The control condition showed a steady reduction in slope with training indicative of learning [present, $F(1,11) = 5.98$, $P = 0.03$; absent, $F(1,11)=14.98$, $P < 0.01$; combined, $F(1,11) = 10.84$, $P < 0.01$ (Unless otherwise noted, the simple effects of training (block) within an experimental condition tested for linear effects)] but there was no reduction of slope or learning in the ST condition [present, $F(1,10) = 0.97$, $P = 0.35$; absent, $F(1,10) = 0.22$, $P = 0.65$; combined, $F(1,10) = 0.62$, $P = 0.45$; ST $\times$ Block interaction, present $F(5,105) = 2.92$, $P = 0.02$; absent, $F(5,105) = 2.31$, $P = 0.049$; combined, $F(5,105) = 4.20$, $P < 0.01$]. The base-time results differ somewhat across the present (*D*), absent (*E*), and combined trials (*F*). On all trials, the ST condition showed lower base times than the control condition, indicating better performance under ST [present, $F(1,21) = 12.56$, $P < 0.01$; absent, $F(1,21) = 4.83$, $P = 0.04$; combined, $F(1,21) = 9.68$, $P < 0.01$]. However, when the present and absent trials were combined, base time had a weak tendency to increase in the control condition [$F(1,11) = 2.61$, $P = 0.13$] and decrease in the ST condition [$F(1,10) = 2.52$, $P = 0.14$; ST $\times$ Block interaction, $F(5,105) = 2.07$, $P = 0.075$].

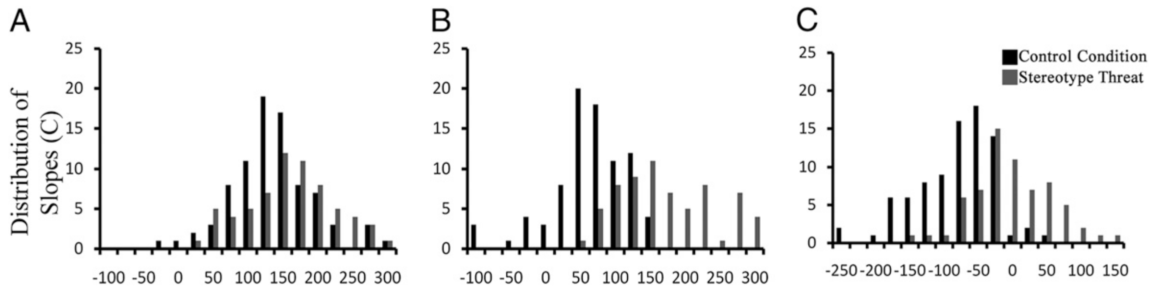

Fig. 3. The distribution of participants slope scores across all three experiments are presented as a function of ST. The slopes were calculated separately for the target-present and target-absent trials and slopes from both types of trials were used to construct distributions. For the first half of trials (*A*), the average of slopes for the first three blocks of exp. 1 and 2 and the first four blocks of exp. 3 were computed, and their count is displayed in 25-ms increments. The same distributions were created for the second half of trials [the average of Blocks 4–6 in exp. 1 and 2 and Blocks 5–8 in exp. 3 (*B*)]. The distribution of the difference in slopes was computed by subtracting the slopes for the first half of trials from the slopes for the second half of trials, and the count is displayed in 25-ms increments (*C*).

most or all women responded by increasing their caution and effort, emphasized accuracy over speed, and stayed with a serial search strategy that may have inhibited learning (18, 19).

The base times *B* are shown in Fig. 2 *D* to *F*. The results exhibit the reverse of the slope results: The base times tend to rise for the control group and fall for the threat group. If the control women learned to switch from an effortful and attention-intensive serial-terminating search to a more automatic and easier search in which they waited for attention to be "drawn" to the target (if present), it would be quite possible for base times to increase. On the other hand, the women under threat may have continued using effortful serial search, speeding their base time as the session continued, but not speeding or even slowing the rate of comparisons (12). A less persuasive but similar account could be proposed if learning in the control group consisted of unitization of characters.

These results do not fit very well with an account in which threat causes a general disruption of mental function. The results suggest instead that women under threat are induced to exert more effort searching (12, 20, 21), in particular continuing to use a more effortful serial search strategy throughout the session; the control women apparently learn to switch to an easier automatic strategy as the targets come to attract attention or become unitized.

Is it the case that women under ST fail to learn automatic attention to targets or unitization of characters? An alternative hypothesis would hold that learning does occur but is overwritten by an effortful strategy, and hence is not expressed. Women in the ST group may have learned an attention-attraction response to targets, but may have inhibited use of that response because of a strong tendency to continue serial search (perhaps because of a belief that serial search would be the best way to ensure high performance). In a similar vein, women under ST may have unitized targets, but nonetheless continued comparing displayed characters to the test item feature by feature, so that the expression of the unitization was inhibited. The next two studies examined this issue.

Exp. 2 (Dataset S2) used three groups of women. The control group replicated the control group of exp. 1, but continued for eight blocks rather than six. The control-plus-ST group replicated the control group of exp. 1 for the first six blocks, at which time the ST manipulation was introduced. The ST-plus-release group replicated the threat group of exp. 1 for the first six blocks and then introduced a self-affirmation manipulation (22) that in previous research has been shown to reduce the effects of ST (23).

The slope results in Fig. 4 *A* to *C* showed typical learning for the control group and for the first six blocks of the control-plus-ST group. The introduction of threat after block six had a strong deleterious effect in the control-plus-ST group. Because learning had already taken place, such a result shows that threat can produce a change in strategy that prevents expression of much of the learning that has taken place. The simplest interpretation is that women switched their strategy from the easier automatic processes and returned to effortful serial search (7). Women under threat did not show decreased slopes as a result of training in the first six blocks and self-affirmation provided no benefit thereafter. If they had learned but were not expressing the learning when under threat, then the release presumably caused by self-affirmation should have allowed expression of the learning. Because this did not happen, the somewhat indirect inference would be that learning had not occurred (and that new learning does not have time to develop in the final two blocks of training). The base time estimates are shown in Fig. 4 *D* to *F*, and are somewhat noisy, but replicate the inverse pattern (compared with slopes) that was observed in exp. 1. These findings are most parsimoniously explained in the same way.

Because the self-affirmation manipulation did not produce a shift in performance, the inference that learning had not taken place in earlier blocks is indirect and weaker than would be desirable. Exp. 3 therefore introduced a different test designed to distinguish learning without expression from lack of learning. Similar to exp. 1, exp. 3 (Dataset S3) used two groups of women who received control or ST instructions, and then all women completed eight blocks of training. Following training for both groups, a new task was introduced for which visual search for Chinese characters was irrelevant: The women were shown two color patches (of the same color) and asked to choose the more saturated of the two. Superimposed on the patches were two irrelevant Chinese characters. In one condition, neither of these characters had been trained as targets. In the critical condition, one of the two had been trained previously as a target. If such a target had developed the ability to attract attention automatically, then one might expect some interference with the color-saturation judgment. Because the superimposed characters are task-irrelevant, any such interruption is presumably an automatic reaction to prior learning, and would not be inhibited by reactions to threat.

The slope and base-time results are given in Fig. 5. The slope and base-time results replicate those of the corresponding conditions in the first two studies.

The critical results for the irrelevant task are shown in Fig. 6. For the control women (for whom learning had occurred), the time to judge color saturation slowed when one of the colors had a superimposed target character (i.e., showed interference from the target character); in addition, the interference was less when the target was on the more saturated color (i.e., match trials) as opposed to the less saturated color (i.e., mismatch trials). Both effects are consistent with an account in which the superimposed

PSYCHOLOGICAL AND COGNITIVE SCIENCES

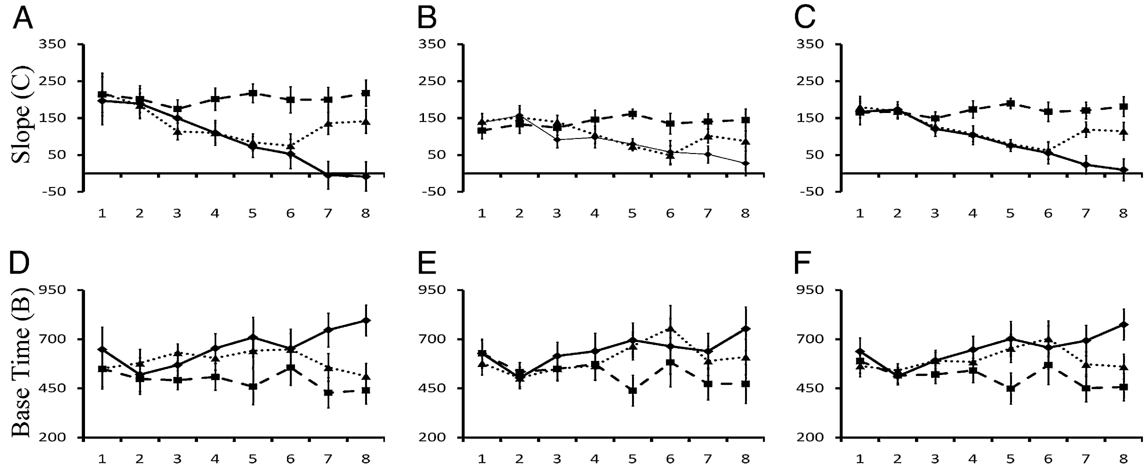

Fig. 4. The slope or *C* and base-time or *B* results for exp. 2 as a function of ST and training (block). The results are plotted separately for target-present trials, target-absent trails, and these trials combined. Results for the control condition are shown with solid lines (and diamonds), results for the control-plus-ST condition are shown with dotted lines (and triangles), and results for the ST-plus-release condition shown with dashed lines (and squares). The slope showed the same pattern of results for the present (*A*), absent (*B*), and combined trials (*C*). Women in the control condition showed a steady decrease in slope with training [present, $F(1,7) = 17.68$, $P < 0.01$; absent, $F(1,7) = 7.71$, $P < 0.03$; combined, $F(1,7) = 29.89$, $P = 0.001$], women in the control-plus-ST condition showed a decrease in slope until the introduction of ST caused the slope to increase [present, Quadratic $F(1,11) = 40.66$, $P < 0.001$; absent, Cubic $F(1,11) = 7.70$, $P = 0.02$; combined, Quadratic $F(1,11) = 66.78$, $P < 0.001$], and women in the ST-plus-release condition showed no change in slope, even after engaging in self-affirmation [present, $F(1,9) = 0.03$, $P = 0.86$; absent, $F(1,9) = 0.59$, $P = 0.46$; combined, $F(1,9) = 0.29$, $P = 0.60$; ST × Block interaction, present $F(14,189) = 1.87$, $P = 0.03$; absent, $F(14,189) = 2.27$, $P < 0.01$; combined, $F(14,189) = 3.55$, $P < 0.001$]. The results for the base time or *B* measure in the target-present trials only showed faster performance in the ST-plus-release condition than in the control and control-plus-ST conditions [$F(2,27) = 3.13$, $P = 0.06$] (*D*). In the target-absent trials, there are no significant effects (*E*). When the present and absent trials were combined (*F*), the base time increased with training in the control condition [$F(1,7) = 6.16$, $P = 0.04$], increased with training until ST was introduced in the control-plus-ST condition [Quadratic $F(1,11) = 3.83$, $P = 0.076$], but decreased with training in the ST-plus-release condition [$F(1,9) = 3.79$, $P = 0.08$; ST × Block interaction, $F(14,189) = 1.74$, $P = 0.05$].

target character automatically engages and attracts attention. For women under threat, such effects were absent.

The control results (exhibiting interference caused by irrelevant superimposed targets) suggest that at least some of the learning that had occurred was the learning of attention to targets. Because it is hard to see why women in this group would not express such learning in an irrelevant color task, the absence of such an effect for the ST group suggests that learning had never occurred. Combining the results of exp. 2 and 3, we conclude that learning was prevented for the women under threat and, when learning does occur before threat, its expression can be reduced significantly by a subsequent induction of ST. That is,

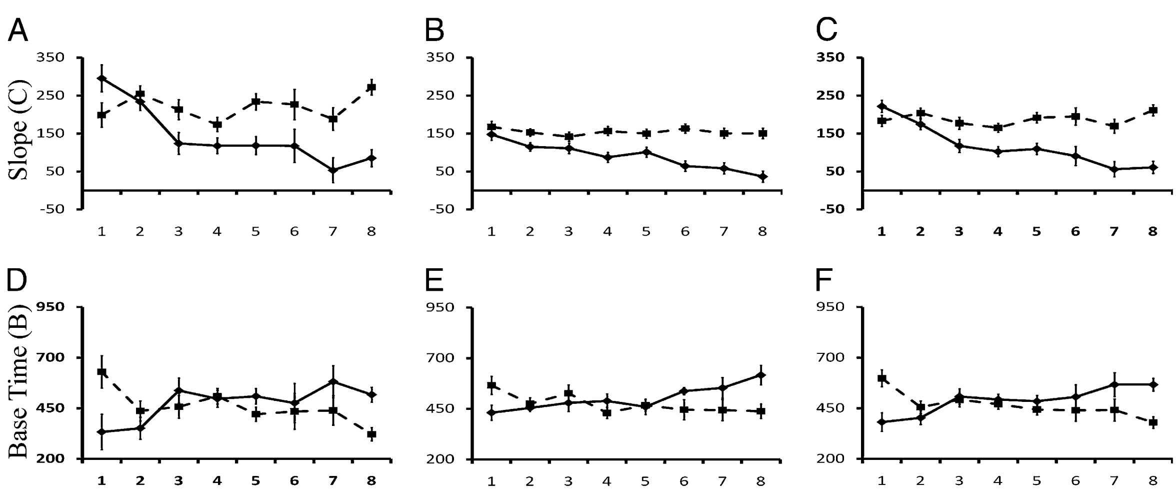

Fig. 5. The slope or *C* and base-time or *B* results for exp. 3 as a function of ST and training (block). The results are plotted separately for target-present trials, target-absent trails, and the combination of present and absent trials. Results for the control condition are shown with solid lines (and diamonds) and results for the ST condition shown with dashed lines (and squares). The slope showed the same pattern of results for the present (*A*), absent (*B*), and combined trials (*C*). The slope decreased with training in the control condition [present $F(1,9) = 29.52$, $P < 0.001$; absent, $F(1,9) = 60.31$, $P < 0.001$; combined, $F(1,9) = 46.08$, $P < 0.001$], but was unaffected by training in the ST condition when the target was present [present $F(1,11) = 0.50$, $P = 0.49$; absent, $F(1,11) = 0.02$, $P = 0.90$; combined, $F(1,11) = 0.36$, $P = 0.56$; ST × Block interaction, present $F(7,140) = 3.35$, $P < 0.01$; absent, $F(7,140) = 3.87$, $P = 0.001$; combined, $F(7,140) = 5.28$, $P < 0.001$]. The base times also showed the same pattern of results for the present (*D*), absent (*E*), and combined trials (*F*). The base time increased with training in the control condition [present, $F(1,9) = 4.93$, $P = 0.05$; absent, $F(1,9) = 7.62$, $P = 0.02$; combined, $F(1,9) = 7.04$, $P = 0.03$], but decreased with training in the ST condition when the target was present [present, $F(1,11) = 8.23$, $P < 0.02$; absent, $F(1,11) = 6.60$, $P = 0.03$; combined, $F(1,11) = 14.17$, $P < 0.01$; ST × Block interaction, present, $F(7,140) = 2.36$, $P < 0.03$; absent, $F(7,140) = 3.38$, $P < 0.01$; combined, $F(7,140) = 3.94$, $P = 0.001$].

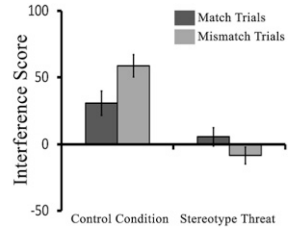

Fig. 6. The interference results for the color task from exp. 3 are plotted as a function of ST and whether the target character matched the correct answer (i.e., was superimposed on the more saturated color swatch) or did not match (mismatch) the correct answer (i.e., was superimposed on the less saturated color swatch). Interference scores indexed the amount of time taken to correctly respond when the target character matched vs. mismatched the correct answer (controlling for response latencies when untrained characters were presented in both color swatches). There was greater interference in the control condition than the ST condition [$F(1,20) = 36.55$, $P < 0.001$]; however, although there was less interference in the control condition on match than mismatch trials [$F(1,20) = 5.82$, $P = 0.03$], interference did not differ between trial type in the ST condition [$F(1,20) = 1.75$, $P = 0.20$; ST $\times$ Trial Type, $F(1,20) = 7.15$, $P = 0.02$].

both effects take place: ST prevents learning and ST can also reduce expression of prior learning.

Discussion

Stereotype threat theory attempts to understand how negative performance stereotypes affect stigmatized group members' performance, and how activation of the stereotype produces and amplifies the resultant performance decrements (2, 5). The present research refines our understanding of the effects of ST by using a task that allows us to distinguish learning from general performance, and thereby allowed us to demonstrate that ST can prevent learning.

The learning that was prevented was attentional/perceptual learning in visual search. In the visual-search learning literature, this learning had been assumed (implicitly) to occur automatically as a consequence of task training. The present results show that such learning can be prevented and therefore is not an automatic consequence of training, and researchers interested in visual-search processes might well want to revisit this issue. Perhaps participants have learned during development to associate effortful processing with superior performance. Then, when reminded of the stereotype in our task and motivated to prove the stereotype wrong, they continually engage in effortful serial search. An attention response to the target would alter and interfere with the serial search, and hence this attention response might be inhibited. This inhibition hypothesis is not unprecedented: Previous research has shown that an untrained target can receive a benefit by being paired with previously trained foils (24). These ideas are interesting but somewhat orthogonal to the main purposes of the present research.

At least in the present task setting, we see that overt emphasis on the existence of the stereotype both prevents learning (exp. 1 and 3) and, to a significant degree, prevents expression of learning that has already occurred (exp. 2). We note that visual search is quite rapid, and its internal processes and their results are not generally available to introspection and awareness. Thus, it would be reasonable for participants under threat to use or switch to an effortful and attention-demanding mode of search in the belief that this would enhance performance (20, 21). The present results do not fit very well with a view that threat simply causes a general deterioration in performance, given that the base-time estimates are, if anything, lower for the women placed under ST. The results seem to fit with the view that the women under threat try harder to carry out the task (12), thereby persisting in effortful serial search throughout training, and failing to find and learn an alternative strategy that makes search easier and less effortful.

We have noted that two kinds of learning have previously been shown to occur in consistent mapping visual search: learned popout (attention attracted to targets automatically) and learned unitization (comparisons of two characters can be made in one step rather than feature by feature in multiple steps). Our research (especially the color task in exp. 3) shows that the control women likely had learned to attend automatically to targets (thereby interfering with color judgments when a task-irrelevant target was superimposed on a color patch). It is possible that unitization had also been learned, although we have no direct evidence concerning this possibility. Whatever had been learned by women in the control group, it seems not to have been learned by women under ST.

The present results are of course found in a specific task: visual search. It remains to be seen whether they generalize to other task settings. The type of automated learning prevented by ST in this research could affect women's math performance in at least two ways: By reducing the extent to which women automate lower-level math tasks, making more complicated math tasks more difficult (25, 26) and by reducing women's ability to differentiate the relevant from the irrelevant information provided in math problems (13). It is of course essential to explore tasks other than perceptual/attention learning to establish the generality of the conclusions. Nonetheless, the present results serve as an existence proof that learning can be prevented by ST, and serve as a warning that the results could be more general.

Conclusion

Stereotype threat is an important problem in many domains, especially in mathematical settings, where women need and would like to perform well. Understanding the role of negative stereotypes in causing decrements in mathematical performance is an important step in finding solutions to the problem. The present research demonstrates a critical component of the deficit: ST can prevent learning, at least somewhat independent of general performance. This result points to the importance of creating environments that reduce the impact of ST during mathematical skill acquisition by women. If creating such environments is not done, the learning deficits that result could well be cumulative, causing problems that continually worsen as development proceeds. Society presently recognizes the importance of finding ways to facilitate the learning of skills to help women enter into careers in science and mathematics, where they are currently underrepresented. Finding ways to reduce the impact of negative stereotypes is one step that should help.

Methods

Participants. Across three experiments, 75 women participated in partial fulfillment of a class requirement. All women provided informed consent before participating in this research. Women in exp. 1 ($n = 23$) and 3 ($n = 22$) were randomly assigned to a control condition ($n = 12$ and $n = 10$, respectively) or a ST condition ($n = 11$ and $n = 12$, respectively). Women in exp. 2 ($n = 30$) were randomly assigned to a control condition ($n = 8$), a control-plus-ST condition ($n = 12$), or a ST-plus-release condition ($n = 10$).

Procedure. After a tutorial about the visual-search task, women were presented with either control instructions or ST instructions. In the control condition, the search task was described as assessing individual differences in problem solving; gender was not mentioned.* In the ST condition, women were told that women underperform relative to men in math and that one reason for this difference is that women have difficulty distinguishing relevant from irrelevant information on math problems, a skill purportedly

*The control instructions read: "In this laboratory, we have been researching how people solve a number of different types of problems. We hope to understand what strategies people use and how these strategies impact problem solving. Thus, we have created a task to attempt to look at how different people approach solving problems in the laboratory. It is important that you take this task seriously so that we can learn more about learning styles and problem solving."

PSYCHOLOGICAL AND COGNITIVE SCIENCES

measured by the search task.[†] These instructions were presented, in shortened form, before each block of trials.[‡]

In the visual-search task, participants searched for a target Chinese character in a display of nontarget characters. Participants pressed one response key if the target character was presented in the display, and another key if it was absent from the display. [Accuracy was high (exp. 1 = 93.3%, exp. 2 = 92.8%, exp. 3 = 94.5%) and not affected by the manipulation of ST or training in any of the experiments. Latencies from trials with incorrect responses were not included in reaction-time measures.] Each block consisted of 80 trials, 20 from each of the trial types presented in Fig. 1. Participants were randomly assigned to receive one of three different sets of five target characters. The location of the characters on the final display (upper right, upper left, lower right, lower left) was counterbalanced across blocks and participants were randomly assigned to one of three orders of blocks. The two counterbalancing factors did not affect the results.

In exp. 2, additional manipulations were presented after Block 6. In the control condition, women took a 30-s break; in the control-plus-ST condition, women read the ST instructions; and in the ST-plus-release condition women engaged in self-affirmation. The self-affirmation procedure involved: (*i*) ranking the personal importance of 11 characteristics (artistic skills, athletics, business, creativity, musical ability, physical attractiveness, relations with friends/family, romantic values, sense of humor, spontaneity, social skills) and (*ii*) writing for 5 min about why the characteristic ranked first held a place of importance (27).

After the visual-search task, women in exp. 3 completed a 120 trial-color task. On each trial, two swatches of the same color (blue, green, red, or yellow) were presented, with one swatch appearing to the participant's left and another to her right. One swatch of color was more saturated than the other and women indicated which side of the screen contained the saturated swatch by pressing the "e" key if it was to her left and the "i" key if it was to her right. To make sure participants understood the task, they completed four practice trials (one for each color) that included feedback. Furthermore, the correct answer appeared equally often on each side of the screen for each trial type. All color swatches had a Chinese character superimposed on them. For the 60 control trials, neither of the superimposed Chinese characters was a target character from the visual search task; for the 60 critical trials, one of the superimposed characters was a target character and the other was not. The 30 match trials had the target character superimposed on the more-saturated color swatch and 30 mismatch trials had the target character superimposed on the less-saturated color swatch.

[†]The bulk of the ST instructions were taken verbatim from the Appendix of ref. 3, p. 276. We added: "One reason why women do more poorly on some math tasks, like word problems on the SAT or ACT, is that they have a more difficult time correctly choosing the information needed to solve the problem from the irrelevant and distracting information that is also provided. The current research is looking at this skill and why women perform more poorly at this skill than men." This addition was to explicitly connect math to the search task.

[‡]The shortened form of the control instructions read: "As was mentioned at the outset, we hope to understand what strategies people use and how these strategies impact problem solving. Thus, we have created a task that attempts to look at individual differences in problem solving style in the laboratory. It is important that you take this task seriously so that we can learn more about learning styles and problem solving." The shortened form of the ST instructions read: "As was mentioned at the outset, one reason why women are poorer on some math tasks, like word problems on the SAT or ACT, is that they have a more difficult time correctly choosing the information needed to solve the problem from the irrelevant and distracting information that is also provided. The current research is looking at this skill and why women perform more poorly at this skill than men. In our lab, we are interested in how underperformance in information selection impacts math performance."

1. Nosek BA, et al. (2009) National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proc Natl Acad Sci USA* 106:10593–10597.
2. Steele C, Spencer S, Aronson J (2002) Contending with group image: The psychology of stereotype and social identity threat. *Adv Exp Soc Psychol* 34:379–440.
3. Beilock SL, Rydell RJ, McConnell AR (2007) Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *J Exp Psychol Gen* 136:256–276.
4. Spencer S, Steele C, Quinn D (1999) Stereotype threat and women's math performance. *J Exp Soc Psychol* 35:4–28.
5. Schmader T, Johns M, Forbes C (2008) An integrated process model of stereotype threat effects on performance. *Psychol Rev* 115:336–356.
6. Steele CM, Aronson J (1995) Stereotype threat and the intellectual test performance of African Americans. *J Pers Soc Psychol* 69:797–811.
7. Beilock SL, Jellison WA, Rydell RJ, McConnell AR, Carr TH (2006) On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Pers Soc Psychol Bull* 32:1059–1071.
8. Bosson J, Haymovitz E, Pinel E (2004) When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *J Exp Soc Psychol* 40:247–255.
9. Murphy MC, Steele CM, Gross JJ (2007) Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychol Sci* 18:879–885.
10. Cadinu M, Maass A, Rosabianca A, Kiesner J (2005) Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychol Sci* 16:572–578.
11. Carr P, Steele C (2009) Stereotype threat and inflexible perseverance in problem solving. *J Exp Soc Psychol* 45:853–859.
12. Jamieson JP, Harkins SG (2007) Mere effort and stereotype threat performance effects. *J Pers Soc Psychol* 93:544–564.
13. Quinn DM, Spencer SJ (2001) The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *J Soc Issues* 57:55–71.
14. Cohen GL, Garcia J, Purdie-Vaughns V, Apfel N, Brzustoski P (2009) Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science* 324:400–403.
15. Thorndike E (1932) *The Fundamentals of Learning* (Teachers College Press, New York).
16. Shiffrin R, Schneider W (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and general theory. *Psychol Rev* 84:127–190.
17. Shiffrin R, Lightfoot N (1997) Perceptual learning of alphanumeric-like characters. *Psychol Learn Motiv* 36:45–81.
18. Grimm LR, Markman AB, Maddox WT, Baldwin GC (2009) Stereotype threat reinterpreted as a regulatory mismatch. *J Pers Soc Psychol* 96:288–304.
19. Seibt B, Förster J (2004) Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *J Pers Soc Psychol* 87:38–56.
20. Solomons L, Stein G (1896) Normal motor automatism. *Psychol Rev* 3:492–512.
21. Downey J, Anderson J (1915) Automatic writing. *Am J Psychol* 26:161–195.
22. Steele C (1988) The psychology of self-affirmation: Sustaining the integrity of the self. *Adv Exp Soc Psychol* 21:261–302.
23. Martens A, Johns M, Greenberg J, Schimel J (2006) Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *J Exp Soc Psychol* 42:236–243.
24. Shiffrin R, Dumais S (1981) The development of automatism. *Cognitive Skills and Their Acquisition*, ed Anderson J (Erlbaum, Hillsdale), pp 111–140.
25. Kellman P, et al. (2008) Perceptual learning and the technology of expertise: Studies in fraction learning and algebra. *Learn Tech Cog* 16:356–405.
26. Vallacher R, Wegner D (1987) What do people think they're doing? Action identification and human behavior. *Psychol Rev* 94:3–15.
27. Cohen G, Aronson J, Steele C (2000) When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Pers Soc Psychol Bull* 26:1151–1164.