Journal of Experimental Social Psychology 45 (2009) 299–312

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

The perils of double consciousness: The role of thought suppression in stereotype threat ☆

Christine Logel [a,*], Emma C. Iserman [a], Paul G. Davies [b], Diane M. Quinn [c], Steven J. Spencer [a]

[a] Department of Psychology, University of Waterloo, 200 University Avenue, Waterloo, ON, Canada N2L 3G1
[b] Department of Psychology, University of British Columbia, Okanagan, 3333 University Way, Kelowna, BC, Canada V1V 1V7
[c] Department of Psychology, University of Connecticut, 406 Babbidge Road, Storrs, CT 06269, USA

ARTICLE INFO

Article history:
Received 12 July 2007
Revised 4 June 2008
Available online 20 August 2008

Keywords:
Stereotype threat
Thought suppression
Postsuppressional rebound
Women
Mathematics

ABSTRACT

The goal of the present research is to demonstrate, and then alleviate, the role of thought suppression in depressing women's math performance under stereotype threat. We hypothesize that when taking a math test, women (but not men) attempt to suppress thoughts of the math-related gender stereotype. Suppression leads to underperformance when it uses up cognitive resources. In Study 1, women underperform on a math test and show postsuppressional rebound of the stereotype when cognitive resources are reduced. In Study 2, women suppress the stereotype after a math test begins, but show rebound when the test is complete. In Study 3, making the stereotype irrelevant to the test improves performance and reduces postsuppressional rebound. In Studies 4 and 5, we test a strategy women can use to make suppression easier, and show that it restores math performance. Theoretical and practical implications are discussed.

Introduction

> When a woman mathematician enters a room, attends a meeting, goes to a conference, or applies for a job, the first thing that is noticed is that she is a woman... As a result, many women talk about feeling 'guilty until proven innocent.' (Henrion, 1997, p. 70).
>
> Some argue that the Association of Women in Mathematics should take more of a traditional focus, for example sponsoring more talks by women about mathematics, not about being a woman in mathematics... As a result, the topic of what women experience as women in mathematics is shunned in most mathematical contexts (Henrion, 1997, p. 79).

As the first quote above suggests, women in math contexts are often acutely aware that they may be judged negatively because they are women. Yet as the second quote describes, women may also be strongly motivated to avoid this judgment. Drawing on the language of W.E.B. Dubois, women in mathematics are in a state of *double consciousness* (Steele, Spencer, & Aronson, 2002). On one hand, as targets of negative stereotypes, women are hyper-vigilant to the possibility of being stereotyped. Yet on the other hand, they are motivated to deny and avoid the implications of such stereotyping. We argue that this predicament lies at the heart of how stereotype threat undermines women's performance in math—the effort required to avoid thinking about a negative stereotype uses up mental resources needed to perform well in the stereotyped domain. In the present studies, we seek to demonstrate, and then alleviate, the role of this thought suppression in the math underperformance of women who are experiencing stereotype threat.

When individuals take math tests, they may feel pressure to perform well for any number of reasons. They may want to meet a personal goal, impress their family, or achieve a required grade to stay in their major. But women face an additional pressure that men do not face: women are aware that many people believe that they are not as good at math as are men, and thus women know that if they do poorly they could confirm this negative stereotype about their gender (e.g., Spencer, Steele, & Quinn, 1999). This *"stereotype threat"* is present, for example, when women take a math test as part of the SAT or GRE exams or for a mathematics course. These types of tests are commonly understood to measure math ability, and thus are situations in which negative stereotypes about women's ability could be applied. This threat of confirming a negative stereotype about one's group has been shown to undermine

☆ This research was supported by an NSERC Doctoral Fellowship and a SSHRC Doctoral Fellowship to the first author, and a SSHRC Grant to the fifth author. We thank Dan Wegner and Vincent Yzerbyt for their feedback on some of the ideas presented in the manuscript, and Danu Stinson and Gregory Walton for their comments on an earlier draft. We also thank Cory Martin, Esther Wiersma, Sarah Chatland, Richard Johnson, Juliana Parker, Rebecca Gerhardstein, and Abigail Pochert for their help with the data collection.

* Corresponding author.
E-mail address: clogel@uwaterloo.ca (C. Logel).

0022-1031/$ - see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.jesp.2008.07.016

the performance of stigmatized groups in a number of domains, including, for example, Black and Latino students on intellectual tests (Gonzales, Blanton, & Williams, 2002; Steele & Aronson, 1995), students with lower socioeconomic status on intellectual tests (Croizet & Claire, 1998), and even White athletes on measures of natural athletic ability (Stone, Lynch, Sjomeling, & Darley, 1999). Most relevant to the present research, stereotype threat also adversely affects women's math performance (Spencer et al., 1999).

Underperformance due to stereotype threat has been partially explained by several mechanisms, including anxiety (Ben-Zeev, Fein, & Inzlicht, 2005; O'Brien & Crandall, 2003; Osborne, 2001; Spencer et al., 1999; but see also Leyen, Desert, Croizet, & Darcis, 2000; Stangor, Carr, & Kiang, 1998; Stone et al., 1999), dejection (Keller & Dauenheimer, 2003), lower performance expectancies (Cadinu, Maas, Frigerio, Impaglizaao, & Latinotti, 2003; Sekaquptewa & Thompson, 2002), and stereotype activation (Davies, Spencer, Quinn, & Gerhardstein, 2002; Steele & Aronson, 1995). Thought suppression has been proposed previously, as an explanation for findings that people experiencing stereotype threat have reduced mental resources available for completing a test (e.g., Schmader, Johns, & Forbes, 2008). However, the present studies are the first that we know of to actually examine the role of thought suppression in stereotype threat effects. By identifying the key role that thought suppression plays in stereotype threat, we can develop a way of reducing stereotype threat effects by reducing the effort required to suppress thoughts of the stereotype.

Activation, suppression, and mental load

When individuals experience stereotype threat, they activate thoughts about the stereotype and concerns about performance, and this activation predicts underperformance on a subsequent test. Steele and Aronson (1995) demonstrated this activation effect with black students writing verbal ability tests. Davies et al. (2002) illustrated this effect with women taking math tests. Participants watched television commercials that either featured women behaving stereotypically or did not feature women, and then completed a lexical decision task assessing the accessibility of the negative stereotype about women's math ability. Women who had watched the stereotypic commercials activated the stereotype, and this activation predicted underperformance on a subsequent math test.

Why would activation of a negative stereotype before taking a test contribute to poorer performance? We suggest that once a test begins, women under stereotype threat feel motivated, consciously or unconsciously, to suppress thoughts associated with the stereotype so that they can concentrate on the difficult test questions. The more strongly a stereotype is activated, the more effort they will expend to suppress it. Wegner and colleagues (Wegner, 1994; Wegner, Carter, Schneider, & White, 1987; Wegner, Erber, & Zanakos, 1993) suggest that in order to suppress a thought, people must consciously avoid thinking about the thought while at the same time unconsciously monitoring the environment for the presence of the thought. This complex process of suppressing thoughts of the stereotype would leave few mental resources for solving difficult test problems, resulting in fewer correct test answers and lower test scores.

The effortful process of thought suppression could explain findings that stereotype threat depletes working memory. Schmader and Johns (2003) found that women experiencing stereotype threat displayed deficits in short-term memory capacity, and these memory deficits mediated underperformance on a math test. Croizet and colleagues found that when stereotype threat was high, participants had greater heart rate variability, an indicator of increased mental load, and this was also associated with poorer performance on an intelligence test (Croizet et al., 2004). It seems likely that it was the effort of suppressing thoughts and worries about the negative stereotype that depleted mental resources in these studies.

The present research

Based on this research, we suggest that when women take math tests under stereotype threat, they attempt to suppress thoughts of the negative stereotype, which increases their mental load, and, consequently, impairs their math performance. This hypothesis extends prior stereotype threat research to investigate more closely why and how stereotype threat leads to underperformance by members of stereotyped groups. In addition, it allows us to develop and test a thought suppression strategy to reduce the negative effects of stereotype threat.

Previous studies have already demonstrated that women experiencing stereotype threat activate thoughts about the stereotype, and that the extent to which they activate these thoughts predicts the degree to which they underperform on a subsequent math test (e.g., Davies et al., 2002). Studies have also shown that stereotype threat causes working memory deficits (Croizet et al., 2004; Schmader & Johns, 2003). Thus, it is our goal in the present studies to make the case that thought suppression plays a key role in these and other stereotype threat results. We do this by showing evidence for thought suppression under stereotype threat in four different ways, described in the subsequent paragraphs. Having demonstrated women's need to suppress thoughts of the stereotype, we then test a psychological tool to help women to overcome the underperformance that arises from the difficult task of suppressing thoughts.

First, we test our theory by examining postsuppressional rebound. Research has shown that once individuals are no longer required to suppress particular thoughts, they show a rebound effect, in which the formerly suppressed thoughts become hyperaccessible (Wegner et al., 1987, Study 2). Thus, if women suppress thoughts about math-related gender stereotypes while they are working on a math test under stereotype threat, then they should show a postsuppressional rebound of these stereotypes once the test has been completed and they are no longer required to suppress the thought. Hence, in each study, we predict that women who write a math test under stereotype threat will underperform on the test and show a postsuppressional increase in the activation of the gender stereotype after the test. However, women who are not under stereotype threat, and men, for whom the stereotype does not apply, will not underperform on the test and will show no such rebound.

Second, we directly measure women's thoughts about the stereotype as they take a test. In Study 2, we measure the activation of math-related gender stereotypes as women are beginning a difficult math test under stereotype threat. We expect that these women will show evidence of suppression of thoughts related to the stereotype, and that this will predict the degree to which they underperform on the math test. This pattern will not be evident for men, who should have no such thoughts to suppress.

Third, we test whether reducing the need to suppress thoughts of the stereotype restores women's abilities to succeed on math tests. Thus, in Study 3 we manipulate the relevance of the negative gender stereotype to women's performance. We expect that when women believe that they cannot be judged by the stereotype about their gender, they will not need to suppress the stereotype, and thus will neither underperform nor suppress the stereotype.

Finally, we test whether alleviating the mental effort of thought suppression allows women to perform better. Research by Wegner and his colleagues (1987) shows that using a thought-substitution task makes thought suppression easier. Therefore, in Studies 4 and 5, we give women a thought-substitution strategy, and test whether it eliminates stereotype threat effects on performance.

Study 1

When women find out they are about to take a math test, this knowledge arouses stereotype threat, because they assume that the test is a measure of math ability, and thus a situation in which they could be judged by the negative stereotype that alleges that women are inferior to men in math. If the test is sufficiently easy for them to solve problems accurately even with some of their cognitive resources devoted to the task of thought suppression, they should realize that the negative stereotype about women will no longer apply to them because they are succeeding on the test. This should reduce stereotype threat, and thus they should no longer need to suppress the gender stereotype once they begin working on the test. However, if they do not have sufficient resources to suppress thoughts of the stereotype and solve test problems correctly—if, for example, they are under additional cognitive load from an unrelated task—then they will continue to experience stereotype threat, and continue to need to suppress thoughts of the stereotype. Women in this situation are likely to underperform on the test, and show a postsuppressional rebound when the test is complete and they no longer need to suppress the stereotype.

Based on this logic, for the first test of our thought suppression hypothesis, we conducted a study in which women and men took an easy math test. We expected women and men to do equally well on the test, replicating previous findings (Quinn & Spencer, 2001; Spencer et al., 1999; Study 1: O'Brien & Crandall, 2003). However, half of the participants were put under cognitive load while taking the test. Among these participants, we expected that women would perform worse than men, because after devoting resources to the cognitive load task and to suppressing thoughts of the stereotype, women would not have mental energy left to correctly solve test problems. As evidence that women under cognitive load devoted resources to thought suppression, we expected that they would show postsuppressional rebound of thoughts related to the stereotype on a lexical decision task.

Participants

As part of a mass-testing booklet at the beginning of the term, potential participants completed two items measuring their identification with math: "I am good at math" and "It is important to me that I am good at math." Participants who strongly agreed with both statements, selecting at least 8 on an 11-point scale, were invited to participate in the study. We pre-selected highly identified students because they are motivated to do well even on math tests in the lab, much as students taking a course exam or the SATs are motivated to do well (Spencer et al., 1999; Steele, 1997). For these highly identified participants, the test-taking situation in the lab would come closer to the importance of a real-world test. Forty-two Hope College students (21 male, 21 female) and 8 University of Waterloo students (4 male, 4 female) met these criteria, and participated in exchange for partial course credit in their introductory psychology course.

Procedure and design

When participants arrived at the lab, they were given the cover story that the researchers were interested in the relationship between mathematical and verbal abilities. For this reason, they would be asked to complete a math test and a vocabulary task (actually the lexical decision task). Participants first took the 25-item easy math test on a computer. The questions were taken from the quantitative section of the GRE general exam, including questions requiring knowledge of algebra, trigonometry, and geometry (Spencer et al., 1999, Study 1). Participants read test instructions that advised them not to guess, because each wrong answer would deduct one quarter of a point from their test score. Half of the participants were randomly assigned to be placed under cognitive load during the math test. These participants were asked, while answering the math questions, to watch a series of letters flashing on the upper left corner of the computer screen. Every time a 'T' was followed by a 'U', they were required to press the spacebar.

Following their completion of the math test, participants completed a lexical decision task, portrayed as the verbal ability test, but actually designed to be a measure of postsuppressional rebound. The task required participants to read letter-strings, presented one at a time using Superlab, and press a key to indicate whether each string was a word or a non-word. Letter strings remained on the screen until participants made a lexical decision. The task included 12 words that pretesting showed were part of the math-related gender stereotype (illogical, intuitive, weak, indecisive, irrational, emotional, complaining, uncertain, worried, confused, failure and distracted).[1] Each of the key words was also matched with a neutral word (defining, animated, melon, someone, advisable, door, coal, context, relative, punctuated, saturated, heating), based on length and language frequency, using norms established by Kucera and Francis (1967) when possible. There were also 12 non-word strings, for a total number of 36 trials. Participants were instructed to complete the task both as quickly and accurately as possible.

After completing the lexical decision task, participants were given the opportunity to ask questions, were fully debriefed, and thanked for their participation. The study was a 2 (gender) × 2 (cognitive load vs. no cognitive load) × 2 (word type: stereotypic vs. neutral) mixed-model design.

Results and discussion

Math test performance

We expected that women taking the easy math test under cognitive load would underperform compared to the other groups of participants.

We calculated performance on the math test using procedures common to standardized tests such as the GRE to correct for guessing. We summed the number of items each participant answered correctly, subtracting one quarter of a point for each incorrect answer as a penalty for guessing, and then divided the resulting score by the number of questions on the test.

An 2 (gender) × 2 (cognitive load vs. no cognitive load) analysis of variance (ANOVA) on participants' test scores revealed a main effect of gender, $F(1,46) = 5.54$, $p < .05$, and a main effect of condition, $F(1,46) = 3.98$, $p = .05$. The predicted interaction did not reach significance, $F(1,46) = 2.02$, $p = .16$. Although the interaction was not significant, a Scheffe's post hoc interaction contrast testing the difference between women in the cognitive load condition and the other three cells was significant, $F(1,46) = 11.42$, $p < .05$, and Tukey post hoc contrasts showed women in the cognitive load condition ($M = 48.90$, $SD = 16.65$) scored worse than men in the cognitive load condition ($M = 62.20$, $SD = 9.28$), $z = -3.7$, $p < .05$, men in the no cognitive load condition ($M = 64.22$, $SD = 12.16$), $z = -4.3$, $p < .05$, and women in the no cognitive load condition ($M = 60.94$, $SD = 9.90$), $z = -3.4$, $p < .05$, as illustrated in Fig. 1.[2]

[1] Words were selected based on a pilot test. Undergraduates (similar to the sample used in the study) were asked to generate a list of words reflecting concerns women might have about how they might be characterized when taking a math test.

[2] The residual F testing the significance of the remaining variance not accounted for by the contrast between women in the cognitive load condition and participants in the other three conditions was also non-significant, $F(3,46) = .33$, $p = .19$, indicating that the remaining variance is not more than would be expected by chance.

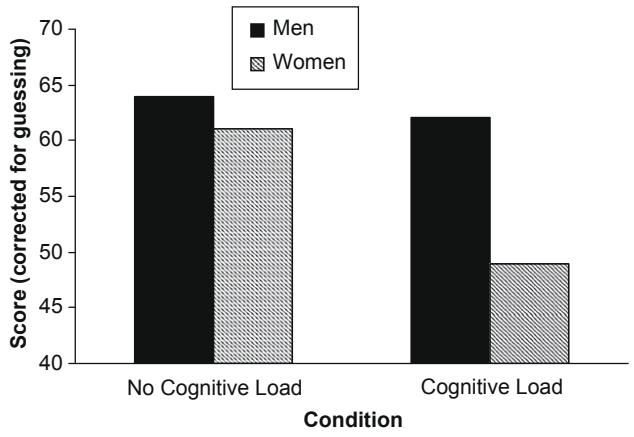

Fig. 1. Test performance by gender and condition in Study 1.

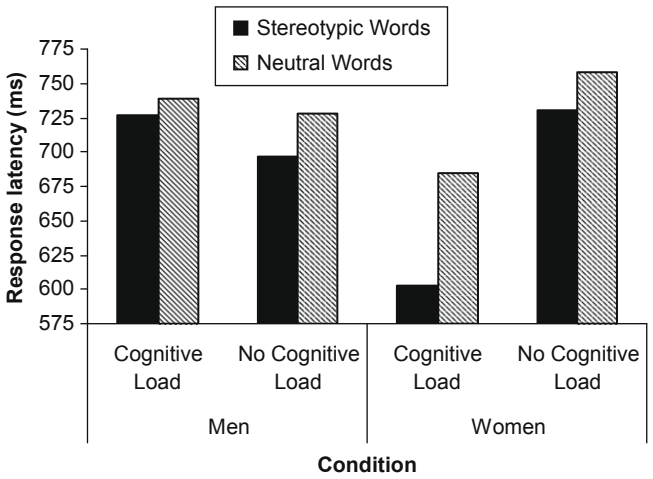

Fig. 2. LDT response latencies by gender and condition in Study 1. *Note:* Standard deviations corresponding to bars on the graph progressing from left to right: 84.02; 78.07; 113.65; 71.65; 82.70; 82.06; 159.02; 117.49.

Test performance and cognitive load task performance

If women underperformed on the math test under cognitive load because the tasks of suppressing the stereotype in addition to monitoring the letters on the cognitive load task used up mental resources required to accurately solve math problems, we would expect performance on the cognitive load task to be negatively related to performance on the math test. In the current study we did not collect performance on the cognitive load task, so we conducted an additional study. Thirteen male and 13 female University of Waterloo students completed only the cognitive load condition of the previously described procedure. Regression analysis revealed a significant interaction between gender and the number of correct identifications on the cognitive load task when predicting test score, $t(25) = 2.2$, $p < .05$. The more correct identifications women made on the cognitive load task, the worse they scored on the math test. Men did not show this relation. This suggests that because they are using up resources suppressing the stereotype about their group, women under stereotype threat must choose to focus their remaining attention either on the math test or on the cognitive load task.

Evidence for postsuppressional rebound

We predicted that after taking the math test under cognitive load, women would respond faster to words relevant to the gender stereotype on the lexical decision task than participants in the other conditions, suggesting that they were experiencing postsuppressional rebound.

To arrive at a final score for stereotypic words and neutral words on the lexical decision task, we excluded data from trials in which participants made errors. We eliminated outliers using the procedures developed by Van Selst and Jolicoeur (1994). In this procedure, outliers are removed from each cell in the design based on their distance in standard deviation units from the cell mean and the number of participants in each cell. Cut-offs for elimination of data were established using Monte Carlo procedures. Using these procedures, 5% of the total trials were removed. We then calculated an average score for each participant on stereotypic and neutral words.

We analyzed participants' responses to stereotypic words compared to neutral words in a 2 (gender) × 2 (condition: cognitive load vs. no cognitive load) × 2 (word type: stereotypic vs. neutral) mixed-model ANOVA, wherein gender and cognitive load condition were between-subjects factors and word type was a within-subjects variable. Results are illustrated in Fig. 2. Results revealed a main effect of word type, $F(1,46) = 22.17$, $p < .01$, a two-way interaction between word type and gender, $F(1,46) = 4.37$, $p < .05$, and a two-way interaction between condition and gender, $F(1,46) = 4.81$, $p < .05$. These effects were qualified by a three-way interaction, $F(1,46) = 5.46$, $p < .05$.

Table 1
Adjusted means and standard deviations on stereotypic words controlling for neutral words: Study 1

Condition	Gender			
	Men		Women	
	M	*SD*	*M*	*SD*
Cognitive load	713.78	64.02	649.27	40.83
No cognitive load	694.96	70.10	695.36	42.26

Note: All times in milliseconds (ms).

To understand the pattern of the interaction, we first conducted a 2 (gender) × 2 (cognitive load vs. no cognitive load) ANOVA predicting neutral words. There were no significant main effects, $Fs < 1.6$, $ps > .22$, but there was an unexpected marginally significant interaction $F(1,46) = 2.81$, $p = .10$. Simple effects revealed that women who completed the math test under cognitive load responded faster to neutral words than women who were not under cognitive load $F(1,46) = 4.26$, $p < .05$. There were no other differences across conditions, all $Fs < 2.32$, all $ps > .13$.

To control for this interaction in general response speed when predicting stereotypic words, we used speed of responding to neutral words as a covariate in a 2 (gender) × 2 (cognitive load vs. no cognitive load) ANCOVA. Adjusted means and standard deviations are displayed in Table 1. This analysis revealed a main effect of gender, $F(1,45) = 3.90$, $p = .05$, which was qualified by a marginally significant two-way interaction, $F(1,45) = 3.48$, $p = .07$. There was no significant main effect of condition, $F = .53$, $p = .47$.

Simple effects tests revealed that women who completed the math test under cognitive load responded faster to stereotypic words than men who were under cognitive load, $F(1,45) = 7.37$, $p = .01$, and marginally faster than women who were not under cognitive load, $F(1,45) = 3.36$, $p = .07$, controlling for neutral words. As expected, men's speed in responding to stereotypic words did not differ by condition, and women who were not under cognitive load did not differ from men in the same condition, all $Fs < 1$.

These results suggest that compared to other groups, women who were under cognitive load were experiencing postsuppressional rebound of concepts associated with the math-related gender stereotype. However, because two of the effects obtained were only marginally significant, we designed subsequent studies to replicate the finding that women who complete a math test while experiencing stereotype threat show postsuppressional rebound.

Test performance and rebound

We have suggested that the more resources women are required to devote to thought suppression, the fewer they have to devote to the math test. Thus, we might expect that the degree to which women suppress thoughts of the stereotype during the test would predict the degree to which they underperform. We thus examined the relation between postsuppressional rebound (as a measure of the degree to which women had been suppressing the stereotype during the test) and math test performance among women in the high cognitive load (stereotype threat) condition. Somewhat surprisingly, test performance was unrelated to postsuppressional rebound, $r = -.07$, $p = .84$. In Study 2, we measure thought suppression directly, which will allow us to determine whether the degree of thought suppression itself predicts the degree of underperformance on the math test.

The results of this study support our hypothesis that thought suppression plays a role in underperformance under stereotype threat. Women who took an easy test performed equally to men and did not show evidence of postsuppressional rebound. This suggests that they did not continue to experience stereotype threat while taking the test and thus did not need to suppress the stereotype about their gender. However, women who took the same test under cognitive load underperformed on the test compared to men and showed evidence that they had been suppressing thoughts of the stereotype, by their postsuppressional rebound after the test was complete. This suggests that they did not have the cognitive resources to accurately complete test questions while suppressing thoughts of the stereotype.

We conducted Study 2 in order to conceptually replicate the underperformance and postsuppressional rebound results of Study 1, while also measuring thought suppression directly as women begin a test.

Study 2

The purpose of Study 2 was to provide further evidence for the role of thought suppression in stereotype threat, this time by measuring thought suppression directly. This design would also allow us to determine whether the degree of thought suppression would predict the degree of underperformance on the test for women under stereotype threat.

In the present study women and men completed a relatively difficult math test and completed the same lexical decision task from Study 1. We chose a more challenging math test for Study 2, so no cognitive load manipulation was required for women to experience stereotype threat during the test. One condition was designed to conceptually replicate the results of Study 1. Half of the participants were randomly assigned to take the math test and then complete the lexical decision task. We expected women to underperform compared to men, and to show postsuppressional rebound on the lexical decision task after the test. The other condition was designed to measure thought suppression directly. Participants began the math test, but after they had read the test instructions and a sample problem and were about to begin working on the test problems, they were interrupted and asked to take the lexical decision task, before completing the math test. We expected women to show thought suppression on the lexical decision task, and that this thought suppression would predict underperformance on the test.

Participants

Participants were 71 (35 male, 36 female) students at the University of Waterloo, who participated for course credit. As in Study 1, participants all strongly agreed with the statements "I am good at math" and "It is important to me that I am good at math."

Procedure and design

Participants were brought into the lab one at a time, given the same cover story used in Study 1, and told that they would begin by completing a math test. The test was comprised of 20 items used by Spencer et al. (1999, Study 3) and was chosen to be more challenging than the test used in Study 1. Participants were given a page of test instructions providing the same information as in Study 1 (a warning that points would be deducted for wrong answers) and a difficult sample question. Half of the participants then completed the math test, followed by the same lexical decision task as in Study 1. The other half of the participants were interrupted after they had read the math test instructions and were about to go on to work on the test questions. The experimenter told them that an error had been made, and they were supposed to do a verbal task (the lexical decision task) before completing the math test. After completing the lexical decision task and the math test, all participants were then fully debriefed and thanked for their participation. The study was a 2 (gender) × 2 (lexical decision task: as test is beginning vs. after test) × 2 (word type: stereotypic vs. neutral) mixed-model design.

Results and discussion

Math test performance

Because all women in this study were taking the math test under stereotype threat, we expected them to underperform compared to men. We did not predict that the order of the tasks would affect test scores.

We calculated test performance using the same methods as in Study 1. An initial 2 (male vs. female) by 2 (lexical decision task as test is beginning vs. after test) ANOVA on participants' test scores revealed the predicted main effect of gender, $F(1,67) = 16.37$, $p < .001$. Women ($M = 15.13$, $SD = 12.23$) on average scored lower on the test than men ($M = 29.11$, $SD = 16.21$). There was no effect of order of tasks, and the interaction did not reach significance, all Fs < 1.

Evidence for suppression and postsuppressional rebound

We predicted that women who took the lexical decision task after completing the math test would show postsuppressional rebound of stereotypic words, replicating Study 1. More importantly, we predicted that women who took the lexical decision task as the math test was beginning would show suppression on the stereotypic words, and the suppression would predict the degree to which they underperformed on the math test.

We eliminated outliers from the lexical decision data using the procedures described in Study 1. Using these procedures, 5% of the total trials were removed. The trials that were removed were spread evenly across conditions. Ten participants who made errors on more than 20% of the words were removed from the analysis. We then calculated an average score for each remaining participant on stereotypic and neutral words.

We analyzed participants' responses to stereotypic words compared to neutral words in a 2 (gender) × 2 (lexical decision task: as test is beginning vs. after test) × 2 (word type: stereotypic vs. neutral) mixed-model ANOVA, wherein gender and lexical decision task order were between-subjects factors and word type was a within-subjects variable. No main effects or two-way interactions were significant, all Fs < 1.10, all ps > .29, except for a two-way interaction between order of lexical decision task and word type, $F(1,57) = 7.61$, $p < .01$. This effect, however, was qualified by the predicted three-way interaction, $F(1,57) = 20.19$, $p < .001$, as illustrated in Fig. 3.

To examine this three-way interaction we conducted a 2 (gender) × 2 (order of lexical decision task) ANOVA predicting neutral

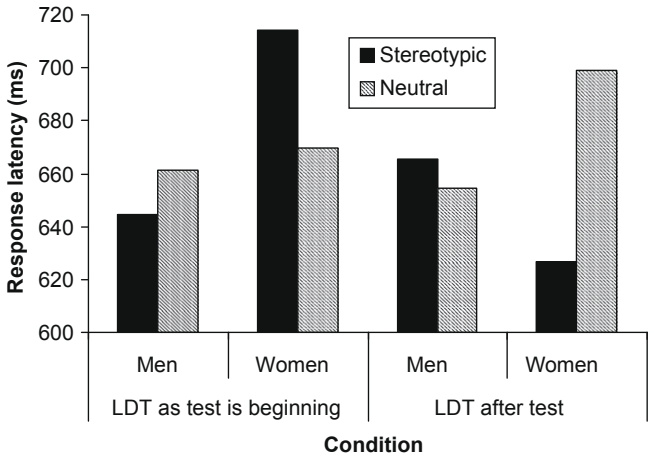

Fig. 3. LDT response latencies by gender and condition in Study 2. *Note:* Standard deviations corresponding to bars on the graph progressing from left to right: 118.25; 103.29; 105.57; 123.31; 104.14; 118.25; 101.67; 95.20.

words, which did not reveal significant main effects or an interaction, all Fs < 1. To control for individual differences in general response speed we used speed of responding to the neutral words as a covariate in a 2 (gender) × 2 (order of lexical decision task) ANCOVA predicting speed of responding to stereotypic words. There was no significant main effect of gender, $F = .30$, $p = .59$, but there was a main effect of condition, $F(1,56) = 7.80$, $p < .01$, which was qualified by the predicted two-way interaction, $F(1,56) = 20.72$, $p < .001$.

Simple effects tests revealed that women who completed the lexical decision task after the math test responded faster to stereotypic words than men in the same condition, $F(1,56) = 20.49$, $p < .001$, and faster than women who completed the lexical decision task as the math test was beginning, $F(1,56) = 29.01$, $p < .001$, controlling for neutral words. This indicates that they were experiencing postsuppressional rebound, replicating the results of Study 1. Women who completed the lexical decision task as the math test was beginning responded slower to stereotypic words than men in the same condition, $F(1,56) = 6.16$, $p < .05$, controlling for neutral words, suggesting that they were suppressing thoughts of the stereotype. As expected, men's speed in responding to stereotypic words did not differ by condition, $F < 2.40$, $p > .12$. Adjusted means and standard deviations are displayed in Table 2.

In summary, women suppressed concepts relevant to the math-related gender stereotype as they were beginning the math test, but showed postsuppressional rebound on these concepts after the test was complete.

Testing mediation

We also predicted that when women completed the lexical decision task as the math test was beginning, the degree to which they suppressed stereotypic words would predict their underperformance on the math test. We dummy coded the categorical variables, mean centered the continuous variables, and regressed gender, condition, stereotypic words, and all higher-order interactions on to math test score, controlling for neutral words. Replicating the ANOVA results, there was a marginal main effect of gender, $\beta = .33$, $t(1,52) = 2.77$, $p = .08$. As previously reported, men's math test scores were higher than women's. There was also a marginal main effect of stereotypic word speed, $\beta = -.59$, $t(1,52) = -1.66$, $p = .08$, and a gender by stereotypic word speed interaction, $\beta = .63$, $t(1,52) = 2.56$, $p = .01$. These effects were qualified by the predicted three-way interaction, $\beta = -.48$, $t(1,52) = -2.14$, $p < .05$. No other main effects or interactions were significant, all ts < 1.02, all ps > .31.

Among women who completed the lexical decision task as the math test was beginning, the more they suppressed the stereotypic words (e.g., the slower they responded) the worse they performed on the math test, $\beta = -.59$, $t(1,52) = -1.66$, $p = .10$, albeit marginally. Predicted math scores revealed that women who were the highest suppressors (+1SD) scored only 11.30% on the math test, whereas women who were the lowest suppressors (−1SD) scored 28.00% on the math test, similar to the scores of men. There was no significant relation between speed of responding to stereotypic words and math test performance for men in either condition, all ts < 1.22, all ps > .22. Finally, replicating the surprising results of Study 1, there was no relation between postsuppressional rebound and test performance among women who took the lexical decision task after completing the math test, $t = -.55$, $p = .58$.

In summary, Study 2 provides evidence for thought suppression under stereotype threat two ways: by replicating the findings in Studies 1 that women show postsuppressional rebound after taking a math test under stereotype threat, and by directly measuring their suppression of the stereotype as they are beginning the test. We also predicted that because attempting to suppress thoughts is cognitively taxing, it would be a mechanism by which stereotype threat leads to diminished performance. Consistent with this prediction, suppression of stereotype related words as the test was beginning predicted women's underperformance on the test.

The purpose of Study 3 was to use a third method to provide evidence for thought suppression under stereotype threat. If thought suppression plays a role in stereotype threat-based underperformance, then alleviating women's need to suppress the stereotype about their gender should alleviate women's underperformance.

Table 2
Adjusted means and standard deviations on stereotypic words controlling for neutral words: Study 2

Condition	Gender			
	Men		Women	
	M	*SD*	*M*	*SD*
LDT as test is beginning	678.03	81.40	731.69	57.59
LDT after test	710.16	52.82	616.49	45.03

Note: All times in milliseconds (ms).

Study 3

In Study 3, women and men completed an easy math test under cognitive load, which, according to our arguments and the results of Study 1, should lead women to experience stereotype threat while they complete the test. However, before the math test, half of the participants were given instructions that previous research had found no gender differences in scores on the test. Studies have shown that when a test is described as gender-neutral, stereotype threat is reduced (Spencer et al., 1999; Quinn & Spencer, 2001).

Replicating earlier research, we expected that women would perform better when the test was described as gender-neutral then when it was not characterized in this way. In addition, we expected that women who were not told that the test was gender-neutral would attempt to suppress thoughts about the stereotype, and thus would show postsuppressional rebound, replicating the results of Studies 1 and 2. However, women who were told that the test did not show gender differences would not think that the gender stereotype could be applied to their test performance, so they would not need to suppress thoughts of the stereotype, and would not show postsuppressional rebound after the test was complete.

Participants

Participants were 54 University of Waterloo students (29 male, 25 female) who participated in exchange for partial course credit in their introductory psychology course. As in Studies 1 and 2, all participants strongly agreed with the statement "I am good at math" and "It is important to me that I am good at math."

Materials

Participants took the same easy math test and lexical decision task used in Study 1.

Procedure and design

Participants were given the same cover story and instructions as in Studies 1 and 2. Half of the participants read additional math test instructions telling them that the test had been found to be gender-neutral (stereotype irrelevant condition). The other half of the participants read instructions that gave no further information about the test (stereotype threat condition). All participants completed the math test along with the same cognitive load task used in Study 1, and then completed the lexical decision task. They were then fully debriefed, and thanked for their participation. The study was a 2 (gender) × 2 (test information condition: no information vs. no gender differences) × 2 (word type: stereotypic vs. neutral) mixed-model design.

Results and discussion

We expected women who took an easy test under cognitive load to underperform compared to men and to show postsuppressional rebound, replicating the results of Study 1. However, when the test was portrayed as gender-neutral, we expected women to perform better on the test and not to show postsuppressional rebound.

Math test performance

We calculated test performance using the same methods as in Studies 1 and 2. A 2 (gender) × 2 (test information condition: no information vs. no gender differences) ANOVA on participants' test scores revealed a main effect of gender, $F(1,50) = 20.55$, $p < .01$. The main effect was qualified by a significant interaction, $F(1,50) = 4.29$, $p = .05$. As can be seen in Fig. 4, replicating past research, women ($M = 58.64\%$, $SD = 10.35$) underperformed compared to men ($M = 80.00\%$, $SD = 11.36$) when they were given no further information about the test $F(1,50) = 22.7$, $p < .05$. When they were instructed that the test did not have gender differences, women ($M = 68.72\%$, $SD = 17.01$) performed better than when given no information $F(1,50) = 3.64$, $p = .06$. Unexpectedly, within the no gender differences condition, a marginally significant difference in performance between men ($M = 75.67\%$, $SD = 8.31$) and women remained $F(1,50) = 2.86$, $p = .09$.[3]

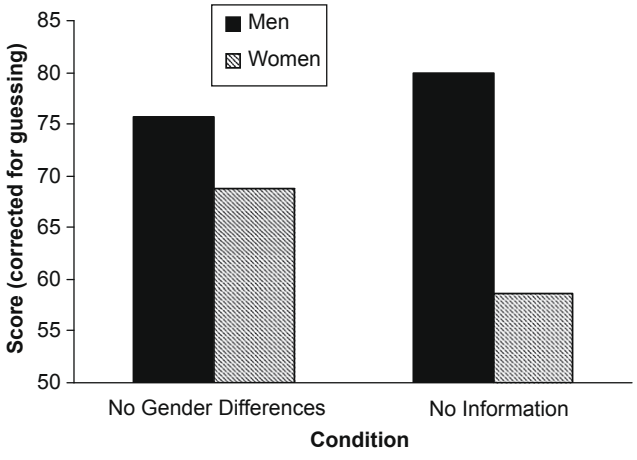

Fig. 4. Test performance by gender and condition in Study 3.

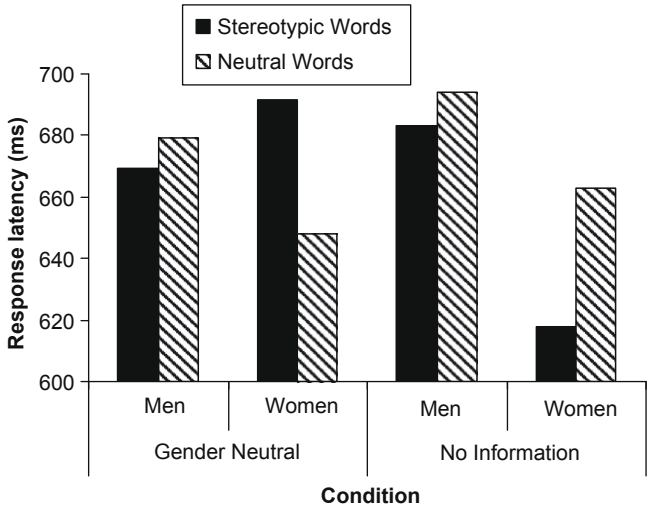

Fig. 5. LDT response latencies by gender and condition in Study 3. *Note:* Standard deviations corresponding to bars on the graph progressing from left to right: 103.12; 92.45; 162.04; 115.97; 145.82; 144.18; 79.13; 121.85.

Evidence for postsuppressional rebound

We predicted that women who took the math test with no additional information would respond faster to stereotype relevant words on the lexical decision task than would other participants.

We eliminated outliers from the lexical decision data using the procedures described in Studies 1 and 2. Using these procedures, 5% of the total trials were removed. The trials that were removed were spread evenly across conditions. Two participants who made errors on more than 20% of the words were removed from the analysis. We then calculated an average score for each participant on stereotypic and neutral words.

We analyzed participants' response times to stereotypic words compared to neutral words in a 2 (gender) × 2 (test information condition: no information vs. no gender differences) × 2 (word type: stereotypic vs. neutral) mixed-model ANOVA wherein gender and test information condition were between-subjects factors and word type was a within-subjects variable. There were no significant main effects or interactions, all Fs < 1, except for a significant two-way interaction between word type and test information condition, $F(1,48) = 4.30$, $p < .05$ and, more importantly, the predicted three-way interaction, $F(1,48) = 4.06$, $p = .05$, as illustrated in Fig. 5.

[3] Women's performance tended to be lower than men's even under low stereotype threat. This result may reflect a tendency for men in the participant pool to have more extensive mathematics backgrounds than women. We were unable to obtain data on the math backgrounds of the participants in the study. However, to gain some insight into the probable math backgrounds of these participants, we examined a sample of participants from the same pool—although from a different term—from which the sample in this study was drawn. In this sample, 46 students were enrolled in the Faculty of Arts (including social sciences, humanities, languages, and fine arts). Thirty-nine of those students were female, while only 7 were male. Forty-two students were enrolled in the Faculties of Math and Engineering; 35 of those students were male, while only 7 were female. Women who met the inclusion criteria for the study reported lower marks in high school calculus than did men who met the inclusion criteria ($M = 84.80$ vs. $M = 88.10$), $t(172) = 2.2$, $p < .05$.

Table 3
Adjusted means and standard deviations on stereotypic words controlling for neutral words: Study 3

Condition	Gender			
	Men		Women	
	M	*SD*	*M*	*SD*
No gender differences	663.70	60.06	711.81	78.34
No information	664.88	70.83	626.25	88.58

Note: All times in milliseconds (ms).

To understand the pattern of the interaction, we conducted a 2 (gender) × 2 (test information condition) ANOVA predicting neutral words. There were no significant main effects, nor was there an interaction, all Fs < 1. In order to be consistent with Studies 1 and 2 and control for general response speed, we used speed of responding to neutral words as a covariate in a 2 (gender) × 2 (test information condition) ANCOVA predicting stereotypic words. This analysis revealed a main effect of test information condition, $F(1,47) = 4.18$, $p < .05$, which was qualified by the predicted two-way interaction, $F(1,47) = 4.34$, $p < .05$. There was no main effect of gender, $F < 1$.

Simple effects tests largely supported our predictions. Women who completed the math test with no additional information responded faster to stereotypic words than women who were told that the test did not have gender differences, $F(1,47) = 7.88$, $p = .01$, controlling for response speed to neutral words. Men's response speed did not differ by condition, $F < 1$. However, women who were given no additional information about the test did not respond significantly faster than men in that condition, $F(1,47) = 1.74$, $p = .19$, although as illustrated in Table 3, the adjusted means trend in the predicted direction.[4]

In summary, in Study 3, women who were told that a math test was gender-neutral performed better on a math test than women who were not given that information, and showed no evidence of postsuppressional rebound after the test.

In the process of testing our proposal that thought suppression is a mechanism of stereotype threat, we have shown that it is possible to reduce the effect of stereotype threat-based thought suppression on test performance, and allow women to perform closer to their true abilities. In Study 3, we did so by telling participants the test was gender-neutral. However, many real-world testing environments are high stereotype threat situations in which characterizing the test as gender-neutral is not feasible. To develop a way to alleviate underperformance due to thought suppression under stereotype threat, we needed to develop a psychological tool that test-takers could use in real test-taking situations. This was our first purpose in conducting Studies 4 and 5.

In addition, in order to increase the generalizability of the findings, in Studies 4 and 5 we modified our selection criteria somewhat. Studies 1–3 used participants who were highly identified in math. The results of the studies showed that these women were vulnerable to stereotype threat. However, it has been suggested that the most talented individuals are also the most vulnerable to stereotype threat (Steele, 1997; Steele & Aronson, 1995), which makes it all the more important to alleviate it. In Studies 4 and 5, we selected only participants who had shown in the past that they were high-achievers in math.

[4] We meta-analysed the lexical decision task effects across Studies 1, 2 and 3. The three-way interaction of gender × condition × word type was highly significant ($z = 4.66$, $p < .001$), as was the two-way interaction of gender × condition predicting response time to stereotypic words, controlling for neutral words ($z = 4.33$, $p < .001$). Women showed postsuppressional rebound compared to men ($z = 4.81$, $p < .001$) and compared to women who were not under threat or were not expected to be suppressing ($z = 4.61$, $p < .001$).

Finally, we thought that based on the results of Studies 1–3, some might argue, however implausibly, that women underperformed on the math tests because they were actually inferior to men in math. According to this argument, women performed equally to men on an easy test in Study 1 because it was easy, but performed more poorly under cognitive load because they, unlike men, did not have the ability to complete easy questions when some of their resources were also devoted to the cognitive load task. In Study 3, when the stereotype was made irrelevant, women's performance improved, but was still marginally lower than men's performance.

We find this account unlikely because it is inconsistent with previous findings that when stereotype threat is reduced, women perform equally to men on even very difficult math tests (e.g., Spencer et al., 1999). However, in Study 4 we gave half of the participants a way to make thought suppression easier. We expected that when they were given this suppression strategy, women would perform equally to men even on a difficult math test, and would not show a postsuppressional rebound.

Study 4

In the studies we have described thus far, we demonstrated the role of thought suppression in stereotype threat-based underperformance by measuring postsuppressional rebound (Studies 1–3), by measuring thought suppression directly (Study 2), and by manipulating the relevance of the stereotype to the testing situation (Study 3). In Studies 4 and 5, we hoped to provide additional evidence for our hypothesis by making thought suppression easier. We expected that this would eliminate the underperformance and postsuppressional rebound effects that typically accompany efforts to suppress the negative math gender stereotype.

This design would also allow us to develop a psychological tool that women can use to cope with stereotype threat. We based our manipulation on thought suppression research showing that participants have an easier time following instructions to suppress thoughts when they are given a suppression strategy—when they are told to replace any thoughts they want to suppress with other thoughts (i.e., to suppress thoughts of a white bear, replace those thoughts with thoughts of a red Volkswagen; Wegner et al., 1987). Without mentioning gender or stereotypes, we instructed participants in the suppression-strategy condition that if they were worried about their performance on the math test, they should think about an important personal identity. We hoped that women could replace worries about the stereotype with thoughts of this personal identity, to make it easier to suppress the stereotype. In Study 5, we test the same hypothesis using a neutral suppression strategy, a red Volkswagen.

We expected that women who were given the suppression strategy would perform equally to men on the math test. Because research has shown that if suppression is subjectively easy, postsuppressional rebound is reduced (Forster & Liberman, 2001), we did not expect women in this condition to show postsuppressional rebound.

Method

Participants

As explained above, we selected participants who had previously demonstrated excellence in math. As part of a mass-testing booklet, potential participants had to report achieving a grade of at least 85% in their final year of high school calculus and major in math, science, or engineering. In order for participants to be able to handle the difficult test questions, we only selected those who reported that they were currently taking introductory calculus or had taken it the pre-

vious term. Finally, participants strongly agreed with the statements "I am good at math," and "It is important to me that I am good at math." Seventy-five first year students (38 females, 37 males) at the University of Waterloo met these criteria, and participated in exchange for partial course credit. One male participant was removed from the data set for failure to follow instructions.

Procedure and materials

Participants were given the same cover story as in the previous studies. They told they would be completing three tasks: a writing task, a math task, and a verbal decision task. The writing task was the suppression-strategy manipulation.

Suppression-strategy condition. The 36 participants in the suppression-strategy condition were given the following instructions to write about something they considered important about their personal identity:

> In the space below, please briefly describe something you consider important about yourself. This should be something you consider an important part of who you are. What is something that you consider important for your personal identity? Why is it important to you? Please note, we ask that this item of importance not be related to academics or academic achievement.

After completing the writing task, participants were immediately given a 12-item, difficult math test (Davies et al., 2002; Spencer et al., 1999). The test consisted of questions taken from the advanced GRE subject exam in mathematics, and required knowledge of advanced calculus, abstract algebra, and real variable theory. Test instructions indicated that they would have 20 min to complete as many questions as they could, but they should not guess, because a one quarter point penalty would be taken from their scores for each incorrect answer.

These participants were given additional written instructions (to which the experimenter drew their attention) that if, at any time during the test, they felt anxious, nervous, or worried, that they were to try and replace those thoughts with thoughts of the personal identity they had written about earlier:

> Sometimes, while writing tests such as this, people get nervous. Please recall briefly the item you identified as important about yourself in the writing task. Should you feel nervous, anxious, or worried you don't know what to do while taking this test, please try and replace those thoughts with thoughts about the personal identity you generated in the writing task.

We expected that if women used the personal identity as a thought substitution, it would improve their math test performance by making it easier for them to suppress the stereotype. However, there were two reasons why the task may have improved their performance. We designed the control conditions to address two of these alternative explanations. Because a limited number of participants met the selection criteria for this study, and in the interests of maximizing the power of the design, we divided our control condition into two subsets.

Campus building control condition. It seemed plausible that writing about a personal identity might improve women's math test performance simply because the writing-task distracted women from anxiety they may have experienced before taking the test. To account for this possibility, half of the control participants completed a neutral writing task before taking the math test:

> In the space below, please briefly describe your attitude towards and feelings about the Student Life Centre. What do you like about it? What don't you like about it? What would you like to see added or changed?"

These participants were not given instructions to use what they had written as a thought substitution.

Personal identity control condition. It also seemed plausible that writing about a personal identity might improve women's math test performance because it provided a self-affirmation (Steele, 1988) right before women began the test. To account for this possibility, the other half of the control participants wrote about an important identity, following the same writing-task instructions as those in the suppression-strategy condition. However, these control participants were not given the instructions on the test to use the personal identity as a thought substitution.

Because none of the control participants were instructed to use what they had written as a thought-substitution task, we did not expect there to be differences in test performance between the two control tasks, so we planned to collapse the data across these two conditions.

After they were given 20 min to work on the math test, participants completed the same lexical decision task from the previous studies. Participants were then probed for suspicion, fully debriefed about the purposes of the study, given the opportunity to ask questions, and thanked for their participation.

Results and discussion

We predicted that the two control conditions (writing about a personal identity but not being instructed to use it as a thought replacement; writing about a building on campus) would have the same effects on performance. We expected that women in these conditions would experience stereotype threat, and would have lower math test scores than men. We expected that women who were given the suppression strategy would find it easier to suppress thoughts of the stereotype, and thus would perform as well on the test as men.

Math test results

We adjusted participants' math test scores using the same correction for guessing as in the previous studies. We then compared means in the control groups to determine if we could collapse our two control groups into one. A 2 (gender) $\times$ 2 (control task: campus building vs. personal identity) ANOVA revealed the predicted gender main effect, $F(1,34) = 7.16$, $p = .01$. Women ($M = 8.30\%$, $SD = 14.18$) scored lower on the test than men ($M = 23.00\%$, $SD = 18.73$). There was no effect of control task or an interaction, all Fs < 1, so we collapsed the control conditions into a single control group for the remaining analyses.

After collapsing the control conditions together, a 2 (gender) $\times$ 2 (suppression strategy vs. no suppression strategy) ANOVA on participants' test scores revealed a main effect of gender, $F(1,70) = 3.89$, $p = .05$, and a main effect of suppression-strategy condition, $F(1,70) = 4.55$, $p = .02$. These main effects were qualified by a significant interaction, $F(1,70) = 4.41$, $p < .05$. As illustrated in Fig. 6, women who were given the strategy to aid thought suppression performed significantly better ($M = 24.42\%$, $SD = 12.48$), than women who were not given the suppression strategy ($M = 8.33\%$ $SD = 14.18$), $F(1,70) = 10.21$, $p < .01$, and just as well as men given the same instructions, ($M = 23.96\%$, $SD = 16.13$), $F < 1$. However, women who did not have the suppression-strategy performed worse than men who did not have the suppression strategy ($M = 23.03\%$, $SD = 18.74$), $F(1,70) = 8.54$, $p < .01$.

Therefore, women who were given a strategy to make thought suppression easier scored as well as men on the math test, whereas women who were not given the suppression-strategy underperformed on the test compared to men.

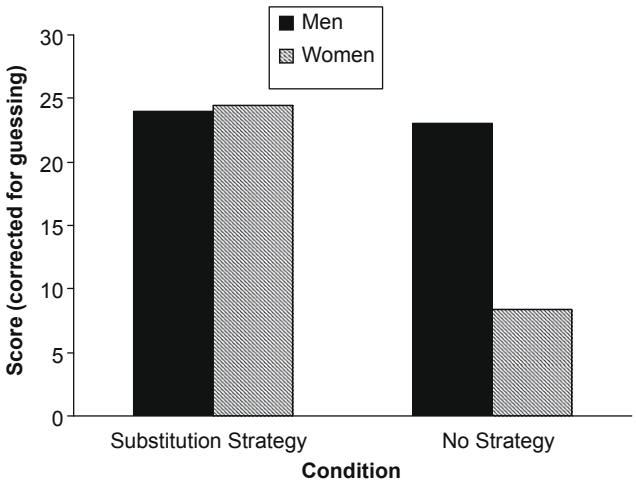

Fig. 6. Test performance by gender and condition in Study 4.

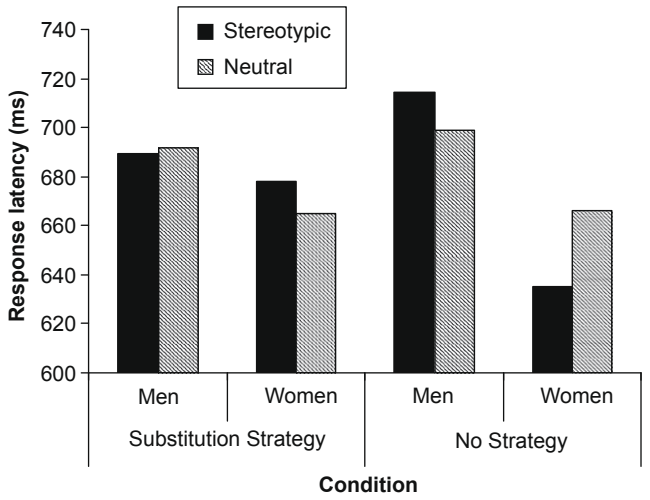

Fig. 7. LDT response latencies by gender and condition in Study 4. *Note:* Standard deviations corresponding to bars on the graph progressing from left to right: 193.903; 188.12; 144.44; 112.12; 133.16; 132.79; 128.46; 104.41.

Evidence for postsuppressional rebound

We predicted that women who were not given the suppression strategy (for whom suppressing negative thoughts would be difficult) would experience postsuppressional rebound, showing activation of stereotypic words on the lexical decision task, whereas women given the suppression strategy would have an easier time suppressing the unwanted thoughts and would therefore not show postsuppressional rebound.

Data was cleaned following the same procedures described in the previous studies. Five percent of the trials were removed using this method. Three participants could not be included due to a failure of the computer program to save their data.

We first compared response latencies within the control conditions to determine whether we could collapse the two conditions into one. A 2 (gender) × 2 (writing task: campus building vs. personal identity) × 2 (word type: stereotypic vs. neutral) mixed-model ANOVA revealed the predicted gender by word type interaction, $F(1,32) = 5.47$, $p = .03$. Women responded faster than men to the stereotypic words ($M = 637.04$ vs. $M = 714.59$), but responded at about the same speed to the neutral words ($M = 668.98$ vs. $M = 698.76$, all times in milliseconds). There was an unanticipated gender by writing-task interaction, $F(1,32) = 4.30$, $p = .05$. This interaction occurred because women's response speed did not vary by condition ($M = 674.66$ vs. $M = 631.36$), but men tended to respond faster to both stereotypic words and neutral words in the campus building writing task than in the personal identity writing task ($M = 647.79$ vs. $M = 765.55$). No other effects were significant, all Fs < 2.0, all ps > .17. Despite the unanticipated interaction between gender and writing task, for simplicity of presentation and because this interaction did not qualify further analyses, we collapsed across the two control conditions for the remaining analyses.

We analyzed participants' response times to stereotypic words compared to neutral words in a 2 (men vs. women) × 2 (suppression strategy vs. control) × 2 (stereotypic words vs. neutral words) mixed-model ANOVA. There were no significant main effects or two-way interactions, Fs < 1.3, ps > .25, but the predicted three-way interaction did emerge, $F(1,67) = 4.59$, $p < .05$, as illustrated in Fig. 7.

As in the previous studies, we next conducted a 2 (gender) × 2 (no information vs. no gender differences) ANOVA predicting neutral words. As expected, there were no significant main effects, nor was there an interaction, all Fs < 1. We next conducted a 2 (gender) × 2 (no information vs. no gender differences) ANCOVA with neutral words as a covariate to control for individual differences in response speed. There were no significant main effects, Fs < 1.31, ps > .25) but there was the predicted two-way interaction, $F(1,66) = 4.66$, $p < .05$.

Simple effects tests revealed that women who were not given a suppression-strategy responded faster to stereotypic words than men in that condition, $F(1,66) = 5.53$, $p < .05$, and faster than women who were given a suppression strategy, $F(1,66) = 4.55$, $p < .05$, controlling for neutral words. Women's response speed did not differ from men's when they were given a suppression strategy, and men's speed in responding to stereotypic words did not differ by condition, all Fs < 1. Adjusted means and standard deviations are displayed in Table 4.

Thus, women who were given a suppression strategy did just as well as men on a difficult math test, whereas women who were not given the suppression-strategy underperformed on the test and showed postsuppressional rebound. These results suggest that the suppression strategy made it easier for women to suppress thoughts of the stereotype about their gender. By showing that women who have a suppression-strategy score equally to men on a difficult math test, these results also refute the suggestion that women are simply inferior to men in mathematics.

There is a potential alternative explanation for these findings. It is possible that the personal identity manipulation was self-affirming to participants, and that the self-affirmation is what restored women's performance. It might be plausible that when women replaced worries about their performance on the test with thoughts of an important personal identity, this made them feel better about themselves and thus less afraid of being judged by a negative stereotype. In this way, the strategy may have worked because it reduced stereotype threat.

On the one hand, this alternative explanation for why the Study 4 manipulation worked does not take away from the finding that women can use a simple psychological tool to improve their math performance under stereotype threat. On the other hand, because

Table 4
Adjusted means and standard deviations on stereotypic words controlling for neutral words: Study 4

Condition	Gender			
	Men		Women	
	M	*SD*	*M*	*SD*
Suppression strategy	678.62	59.04	693.06	63.08
No suppression strategy	697.10	54.57	649.22	65.83

Note: All times in milliseconds (ms).

the goal of the present research is to demonstrate and then alleviate the role of thought suppression in stereotype threat-based underperformance, we conducted a final study to provide evidence that a neutral suppression strategy, that would not be affirming, could restore women's math performance.

Study 5

In Study 5, we gave participants a neutral suppression strategy. If women attempt to suppress thoughts related to the stereotype about their gender, and the manipulation in Study 4 restored their test performance because it made thought suppression easier, then a neutral suppression strategy should also be effective in improving test performance.

We selected the thought suppression method used by Wegner and colleagues (1987, Study 2) as the strategy participants would be given. Participants were asked to think about a red Volkswagen if they were anxious or nervous about their performance on the math test. We predicted that women who were given this suppression strategy would perform better on the test than women who were not given a strategy.

Method

Participants

The same selection criteria as in Study 4 were used to recruit high-achieving students who were identified with math and could handle difficult test questions. However, because we established in Studies 1 and 4, along with previous research (e.g., Spencer et al., 1999) that women who are not experiencing stereotype threat perform equally to men, we concluded that we could determine the effect of our manipulation simply by comparing women who received it to women who did not, so we did not recruit men for the final study. Participants were 90 first year students (all women) at the University of Waterloo who participated in exchange for partial course credit.

Materials

The cover story, math test and lexical decision task were the same as those used in Study 4.

Procedure and design

Participants were brought into the lab one at a time and given the same cover story as in the previous studies. The cover page of the math test had the standard instructions that they had 20 minutes to complete the test and there would be a penalty for guessing. For participants randomly assigned to the suppression-strategy condition, the instructions also read:

Sometimes, while writing tests such as this, people get nervous. Please think now of a red Volkswagen. Should you feel nervous, anxious, or worried you do not know what to do while taking this test, please try and replace those thoughts with thoughts about a red Volkswagen.

The experimenter ensured that participants had read the instructions, then left the room for 20 min while participants worked on the test. After 20 min had elapsed, participants completed the lexical decision task, were probed for suspicion, fully debriefed and thanked for their participation. The study was a 2 (suppression strategy vs. no suppression strategy) × 2 (word type: stereotypic vs. neutral) mixed-model design.

Results and discussion

Math test performance

Test scores were calculated in the same way as in the previous studies, by subtracting one quarter of a point for each wrong answer and dividing by the total number of questions. As predicted, women who were given the suppression strategy had higher test scores ($M = 19.01\%$, $SD = 18.70$) than those who were not given the suppression strategy ($M = 9.15\%$, $SD = 16.06$), $t(88) = 2.69$, $p < .01$.

Evidence for postsuppressional rebound

The data were cleaned using the procedures described in the previous studies. Four percent of the total trials were removed using this method. Five participants could not be included because the computer failed to save their data correctly (three in the suppression-strategy condition, two in the no strategy condition).

We predicted that women who were not given the suppression strategy would show postsuppressional rebound of stereotype relevant words on the lexical decision task. A 2 (suppression strategy vs. no suppression strategy) × 2 (word type: stereotypic words vs. neutral words) mixed-model ANOVA revealed a main effect of word type, $F(1,83) = 10.71$, $p < .01$, and a main effect of condition, $F(1,83) = 9.61$, $p < .01$. These main effects were qualified by a significant interaction, $F(1,83) = 5.50$, $p < .05$, as illustrated in Fig. 8.

An analysis predicting speed of responding to neutral words revealed that, similar to Study 1, women who were not given the suppression-strategy responded faster to the neutral words than women who were given the strategy, $F(1,83) = 6.12$, $p < .05$. As in the previous studies, we used speed of responding to neutral words as a covariate to predict response speed to stereotypic words. This analysis revealed that women who were not given the suppression-strategy responded faster to the stereotypic words than women who were given the strategy $F(1,82) = 5.63$, $p < .05$, controlling for the neutral words. Adjusted means and standard deviations are reported in Table 5.

These results replicate the results of Study 4. Women who were given the suppression-strategy performed better on the math test than women who were not given the suppression strategy and did not show postsuppressional rebound. These results were obtained using a neutral thought replacement, suggesting that a suppression strategy does not have to include an affirmation to be effective at reducing the effect of stereotype threat on women's math performance.

General discussion

W.E.B. Dubois wrote about the concept of double consciousness, and this idea is easily applicable to a woman's experience when

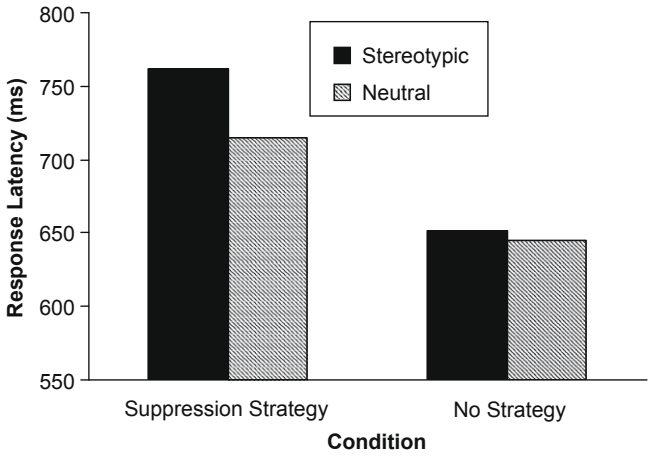

Fig. 8. LDT response latencies by condition in Study 5. *Note:* Standard deviations corresponding to bars on the graph progressing from left to right: 173.28; 145.33; 113.16; 114.66.

Table 5
Adjusted means and standard deviations on stereotypic words controlling for neutral words: Study 5

Condition	*M*	*SD*
Suppression strategy	727.86	89.90
No suppression strategy	685.70	60.14

Note: All times in milliseconds (ms).

taking a math test. On the one hand, a woman taking a math test is aware that she may be judged in terms of a negative stereotype about women. On the other hand, she wants to avoid that stereotype. We believe this combination of awareness and avoidance leads to attempts to suppress negative thoughts. Unfortunately, as shown by the present research, attempts at suppression actually interfere with performance and lead to women confirming the very stereotypes they are trying to avoid.

The goal of the present research was to demonstrate, and then alleviate, the role of thought suppression in underperformance due to stereotype threat. Results of Study 1 suggest that thought suppression decreases mental resources under stereotype threat. Women who took an easy math test scored as highly as men did, but women who took the test under cognitive load scored lower than men and showed postsuppressional rebound on words related to the negative gender stereotype. This suggests that women underperformed on the test because the extra effort of suppressing thoughts of stereotype, in addition to completing the cognitive load task, used up mental resources needed to work on test problems.

Results of Study 2 demonstrated directly the relation between stereotype threat, thought suppression and underperformance. Conceptually replicating Study 1, women who took a lexical decision task after a math test underperformed on the test compared to men and showed postsuppressional rebound of thoughts related to the stereotype. More importantly, women who took the lexical decision task as they were beginning the math test showed suppression of thoughts related to the stereotype, and the degree of suppression predicted the degree of to which they underperformed on the test.

Study 3 showed that reducing stereotype threat reduces both the need to suppress thoughts of the stereotype, and the underperformance that results from the effort of suppression. Women underperformed on a math test and showed postsuppressional rebound, unless they were told that the math test did not show gender differences. These women performed better on the test and did not show postsuppressional rebound.

Studies 4 and 5 showed that the negative effects of the need to suppress thoughts of the stereotype can be overcome with a simple psychological tool. In Study 4, women who were instructed to replace worries about their test performance with thoughts of an important personal identity performed equally to men on a difficult math test and did not show postsuppressional rebound. Women who were given control instructions scored lower than men on the test and showed postsuppressional rebound of thoughts of the negative stereotype. Study 5 replicated these results and showed that a neutral replacement thought, of a red Volkswagen, also improved women's performance and eliminated postsuppressional rebound. These results suggest that by making it easier for women to suppress thoughts of the stereotype, women have cognitive resources left to solve difficult test problems.

Implications for understanding stereotype threat

The present studies demonstrate that thought suppression is an underexamined and potentially important mediator of stereotype threat. By considering these studies together with previous stereotype threat research, a story emerges about the sequence of events that begin when women find themselves experiencing stereotype threat. First, they activate thoughts of the negative stereotype about their gender (Davies et al., 2002). Next, if they are required to perform in the stereotyped domain, for example by taking a math test, they suppress thoughts of the stereotype when the test begins (Study 2) and underperform on the test (Studies 1–5). When the test is complete and there is no longer a need to suppress the stereotype, they show postsuppressional rebound (Studies 1–5). However, if the test is easy (Study 1) or gender-neutral (Study 3) and therefore the stereotype does not apply, then women do not underperform or show postsuppressional rebound, suggesting that they did not suppress the stereotype. If they are given a strategy to make suppression easier (Studies 4 and 5), they do not underperform or show postsuppressional rebound. These findings support our hypothesis that women underperform under stereotype threat because the effort required to suppress thoughts of the negative stereotype uses up cognitive resources needed for solving test problems.

If suppressing the stereotype leads to lower math test scores, would embracing the stereotype lead to higher scores? Research on stereotype endorsement suggests just the opposite. Schmader, Johns, and Barguissau (2004) found that women who endorsed the stereotype were *more* susceptible to the negative effects of stereotype threat on their math test performance. Embracing the stereotype may release women from the double consciousness of recognizing the stereotype and trying to avoid it, but it does not protect them from the stereotype's negative effects.

These results add to a growing literature that highlights just how powerful and pernicious these negative effects can be. Stereotype threat affects women taking a math test even without special high-threat instructions (Studies 2–5, see also Spencer et al., 1999 Study 3). Simply taking a test of mathematical ability invokes the pressures of the negative stereotype. Stereotype threat affects people who have the most to offer to the field, who are highly identified with the domain (Studies 1–5, see also Steele, 1997) and who are highly skilled at math (Studies 4 and 5, see also Spencer et al., 1999) and have presumably succeeded in threatening environments in order to make it into a mathematics, science or engineering major.

Some researchers have suggested an ideomotor explanation for stereotype threat findings (e.g., Wheeler & Petty, 2001). Pointing to research showing that behavior can be a consequence of priming effects, they suggest that when a stereotype becomes activated, stereotype-consistent behavior may follow automatically from that activation. For example, a woman faced with a math test will activate the stereotype in her mind that women are not good at math, and her behavior (answering the questions incorrectly) follows directly from that activation. Although the ideomotor account probably explains differences in performance in some situations, it does not provide a compelling account of the present results. Ideomotor theory would suggest that activation of the stereotype should predict impaired performance, whereas in Study 2, we found just the opposite, that suppression of the stereotype predicted underperformance on the test. This suggests that although the ideomotor effect can account for some behaviors, it is not an appropriate explanation in this case.

Implications for thought suppression research

This research builds on the thought suppression literature by expanding it into a new and important domain. Our use of a thought-substitution task to improve test performance is, to our knowledge, a unique application of this previous thought suppression research.

One question that might interest thought suppression researchers is whether women are aware of their motive to suppress

thoughts of the stereotype or whether the urge to suppress these thoughts occurs at a less conscious level. This question is beyond the scope of the present studies, but these data do show at least that the solution can be conscious. Women in Study 3 were consciously aware of test instructions indicating that the stereotype did not apply to the current test-taking situation, and women in Studies 4 and 5 were consciously aware of the thought-substitution strategy, which was designed to be used consciously.

One surprising result that may be of interest to thought suppression researchers is the apparent lack of relation between math test performance under stereotype threat and postsuppressional rebound. In Study 2, we found that women's success in suppressing the stereotype as the test was beginning negatively predicted their success on the test. It is somewhat surprising, then, that in Studies 1 and 2, math test performance was unrelated to postsuppressional rebound. Examinations of the data for Studies 3, 4 and 5 and also failed to find a relation, all *ps* > .45. The lack of relation suggests that there may be a moderating variable that is not being accounted for, such as the strategy women choose to cope with the problem of not having enough mental resources to suppress the stereotype and accurately solve math problems. Some women may choose to devote resources to successful thought suppression. This leaves fewer resources to devote to the test, which leads to lowered performance, but less postsuppressional rebound. Other women may choose to devote more mental resources to the test. This leaves fewer resources to devote to thought suppression and more intrusion of unwanted thoughts, which leads to lowered performance, but more postsuppressional rebound. These two strategies lead to opposite relations between test performance and postsuppressional rebound, so if women equally tend to do one or the other, the overall relation between test performance and postsuppressional rebound will be approximately zero. More research is required to fully understand this issue.

Implications for alleviating stereotype threat in the real world

Ideally, the solution to stereotype threat-based underperformance would be to eradicate the false, negative stereotypes that create extra pressure for women in male-dominated domains. Until that can be achieved, a more realistic strategy is to give women a tool they can use to perform well in spite of the negative stereotypes. The current studies show that a simple thought-substitution task can successfully restore women's performance on a difficult math test in the lab. Will this tool be useful in the real world?

We endeavored to create a testing atmosphere that approximated the importance of real-world math tests that allowed the necessary experimental control. Because there were no particular incentives towards performing well on the math tests, we needed participants who would be intrinsically motivated to do well. For this reason, we selected participants who self-identified with mathematics (that is, who believed math to be an important part of their identities). In most real-world testing situations, such as course exams, SATs or GREs, even women who are not identified with math will be motivated to achieve the highest possible score because of the real-world consequences to their performance. For this reason, we think these results should generalize to testing situations outside of the lab.

Although all the participants in our studies identified with math, we selected a range of mathematical abilities. In Studies 1–3, participants had moderately strong math backgrounds, and in Studies 4 and 5 participants had very strong math backgrounds. The similar pattern of results under stereotype threat (Studies 1–5) and when threat is alleviated (Studies 3–5) suggest that these results will generalize to women with a variety of abilities and experience in math.

Still, more research may be needed before applying the thought-substitution suppression strategy (Studies 4 and 5) to low-achieving women. On the one hand, a suppression strategy may be especially effective for low-achievers because in addition to group-level threats, they may have other performance concerns competing for cognitive resources (e.g., a history of poor scores in the domain, or others' low expectations of them). On the other hand, thinking of a red Volkswagen (as in Study 5) may not be a powerful enough tool to allow them to overcome all of those pressures. For them, a thought replacement that is an important personal identity (as in Study 4) may be more effective—writing about an important value has been shown to improve the academic success of low-achieving minority students (Cohen, Garcia, Apfel, & Master, 2006).

Although results of Studies 4 and 5 suggest that a thought-substitution task can alleviate performance decrements in the short term, additional research should examine its value as a long-term strategy for coping with stereotype threat. It is possible that over time, test-takers could begin to associate the replacement thought with the worries it is meant to replace. Thought-substitution may be one of several strategies that women can use to restore cognitive resources in persistently threatening environments. For example, one study found fewer intrusive thoughts among participants who were instructed to monitor and record these thoughts, rather than attempting to suppress them (Salkovskis & Campbell, 1994).

If women can use thought-substitution tasks and other strategies to overcome the barriers they face in male-dominated fields, it may create a snowball effect. Their role modeling will facilitate other women's success, leading over time to the diminishing of negative stereotypes, and an end to stereotype threat.

References

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*, 174–181.

Cadinu, M., Maas, A., Frigerio, S., Impaglizaao, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33*, 267–285.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social–psychological intervention. *Science, 313*, 1307–1310.

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

Croizet, J. C., Désprés, G., Gauzins, M., Huguet, P., Leyens, J., & Méot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30*, 721–731.

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615–1628.

Forster, J., & Liberman, N. (2001). The role of attribution of motivation in producing postsuppressional rebound. *Journal of Personality and Social Psychology, 81*, 377–390.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin, 28*, 659–670.

Henrion, C. (1997). *Women in mathematics: The addition of difference*. Bloomington: Indiana University Press.

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29*, 371–381.

Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

Leyen, J. P., Desert, M., Croizet, J. C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin, 26*, 1189–1199.

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.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race ans sex differences in achievement? *Contemporary Educational Psychology, 26*, 291–310.

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.

Salkovskis, P. M., & Campbell, P. (1994). Thought suppression induces intrusion in naturally occurring negative intrusive thoughts. *Behavioural Research and Therapy, 32*, 1–8.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 440–452.

Schmader, T., Johns, M., & Barguissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles, 50*, 835–850.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336–356.

Sekaquptewa, D., & Thompson, M. (2002). The differential effects of solo status on women's math performance. *Journal of Experimental Social Psychology, 38*, 194–201.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–8.

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

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 21). San Diego, CA: Academic Press.

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

Steele, C. M., & Aronson, J. (1995). Contending with a stereotype: African-American intellectual test performance and stereotype threat. *Journal of Personality and Social Psychology, 69*, 797–811.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.). *Advances in experimental social psychology* (Vol. 34). San Diego, CA: Academic Press.

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

Van Selst, M. V., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology A, 47*, 631–650.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101*, 34–52.

Wegner, D. M., Carter, S. R., Schneider, D. J., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology, 53*, 5–13.

Wegner, D. M., Erber, R., & Zanakos, S. (1993). Ironic processes in the mental control of mood and mood-related thought. *Journal of Personality and Social Psychology, 65*, 1093–1104.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*, 797–826.