THE ROLE OF MOTIVATION IN BLATANT STEREOTYPE THREAT, SUBTLE STEREOTYPE THREAT, AND STEREOTYPE PRIMING

A dissertation presented

by

Jeremy Paul Jamieson

to
The Department of Psychology

In partial fulfillment of the requirements of the degree of
Doctor of Philosophy

in the field of

Psychology

Northeastern University
Boston, Massachusetts
June, 2009
THE ROLE OF MOTIVATION IN BLATANT STEREOTYPE THREAT, SUBTLE STEREOTYPE THREAT, AND STEREOTYPE PRIMING

by

Jeremy Paul Jamieson

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology in the Graduate School of Arts and Sciences of Northeastern University, June, 2009
ABSTRACT

Stereotype threat refers to the concern that is experienced when a stigmatized individual feels at risk of confirming, as self-characteristic, a negative stereotype about his/her group (Steele & Aronson, 1995). Although a multitude of research demonstrates that threat debilitates performance in stereotyped domains, little is known about how stereotype threat produces this debilitating effect on performance. The current research focuses on the potential role of motivation in producing these effects. More specifically, the experiments presented here test Jamieson and Harkins’s (2007) mere effort account, which argues that stereotype threat motivates participants to want to perform well, which potentiates prepotent responses. If the prepotent response is correct, performance is facilitated. If incorrect, and participants do not know, or lack the knowledge or time required for correction, performance is debilitated.

Each of the experiments in this research examines gender-math stereotypes, and indexes performance using problems taken from the quantitative section of the GRE general test. The GRE quantitative test includes two types of problems: solve problems, which require the solution of an equation; and comparison problems, which require the use of logic and estimation. Research shows that test-takers’ prepotent tendency on math problems is to apply a solving approach (i.e., use known formulas and equations to compute an answer) (e.g., Gallagher et al., 2000). Consistent with mere effort predictions, Experiments 1 and 2 demonstrate that, regardless of problem difficulty, threatened participants perform better than controls on solve problems (prepotent response correct), but worse than controls on comparison problems (prepotent response incorrect).
Experiment 3 shows that a simple instruction as to the correct solution approach eliminates the performance deficit on comparison problems.

Stone and McWhinnie (2008) argue that although motivation may account for performance under blatant (i.e., explicit) threat, working memory interference accounts for performance under subtle threat (e.g., solo status). Experiment 4 examined whether the motivation-based, mere effort explanation generalizes across different types of stereotype threat. Females completed GRE-Q problems under subtle threat, and the results replicated those observed in Experiment 1. Threatened females performed more poorly on comparison problems and better on solve problems, suggesting that motivation can also account for the effect of subtle threat on performance.

Stereotypes can also influence performance through priming processes. That is, activating stereotype constructs can lead individuals to behave stereotypically (e.g., Bargh, Chen, & Burrows, 1996). However, little research has focused on differentiating stereotype threat and stereotype priming effects. Experiment 5 sought to determine if threat and priming operate via different or similar mechanisms. As was the case in Experiments 1-4, stereotype threat debilitated performance on problems for which the prepotent solution approach was incorrect, but facilitated performance when the prepotent solution approach was correct. In contrast, female gender primes impaired performance on both types of quantitative GRE problems. These findings suggest that performance debilitation under stereotype threat may be the result of trying too hard, whereas stereotype priming effects may result from trying too little. In sum, this research illustrates the role that motivation plays in producing the effects of stereotype threat on performance and supports Jamieson and Harkins’s (2007) mere effort account.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>5</td>
</tr>
<tr>
<td>List of Figures</td>
<td>6</td>
</tr>
<tr>
<td>List of Tables</td>
<td>7</td>
</tr>
<tr>
<td>Introduction</td>
<td>8</td>
</tr>
<tr>
<td>Chapter 1: Mere Effort and Stereotype Threat on the GRE</td>
<td>11</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>17</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>25</td>
</tr>
<tr>
<td>Chapter 2: A Motivation Based Intervention Strategy</td>
<td>34</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>35</td>
</tr>
<tr>
<td>Chapter 3: The Role of Motivation in Subtle Stereotype Threat</td>
<td>49</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>53</td>
</tr>
<tr>
<td>Chapter 4: Distinguishing Stereotype Threat from Stereotype Priming</td>
<td>66</td>
</tr>
<tr>
<td>Experiment 5</td>
<td>68</td>
</tr>
<tr>
<td>Chapter 5: General Discussion</td>
<td>81</td>
</tr>
<tr>
<td>References</td>
<td>94</td>
</tr>
<tr>
<td>Footnotes</td>
<td>101</td>
</tr>
<tr>
<td>Appendix A</td>
<td>103</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: Experiment 1, overall % correct 22
Figure 2: Experiment 3, mediation 44
Figure 3: Experiment 4, overall % correct 57
Figure 4: Experiment 4, mediation 61
Figure 5: Experiment 5, mediation 74
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Experiment 1, GRE performance</td>
<td>24</td>
</tr>
<tr>
<td>Table 2</td>
<td>Experiment 2, GRE performance</td>
<td>31</td>
</tr>
<tr>
<td>Table 3</td>
<td>Experiment 3, GRE performance</td>
<td>41</td>
</tr>
<tr>
<td>Table 4</td>
<td>Experiment 3, use of solving approach</td>
<td>43</td>
</tr>
<tr>
<td>Table 5</td>
<td>Experiment 3, total % correct</td>
<td>47</td>
</tr>
<tr>
<td>Table 6</td>
<td>Experiment 4, GRE performance</td>
<td>59</td>
</tr>
<tr>
<td>Table 7</td>
<td>Experiment 4, use of solving approach</td>
<td>60</td>
</tr>
<tr>
<td>Table 8</td>
<td>Experiment 5, GRE performance</td>
<td>72</td>
</tr>
<tr>
<td>Table 9</td>
<td>Experiment 5, use of solving approach</td>
<td>76</td>
</tr>
</tbody>
</table>
Introduction

Stereotypes are "beliefs about the characteristics, attributes, and behaviors of members of certain groups" (Hilton & von Hipple, 1996, p. 240). These stereotypical beliefs underlie our attitudes towards members of stereotyped groups and also influence how we behave towards these individuals. Thus, the existence of negative stereotypes can damage intergroup relations through prejudice (attitudes) and discrimination (behavior). Because of the many negative consequences of discrimination, we have laws discouraging those behaviors in the present day United States. However, the passing of laws and the public decrees of politicians does not instantly reverse years of enculturation. Thus, stereotypes persist and can cause problems, while society remains unaware of the potential negative impact of derogatory beliefs, comments, and jokes on intergroup relations.

Although holding stereotypical beliefs about members of particular groups can have a negative effects on the holders of those stereotypes (e.g. Vanman, Paul, Ito, & Miller, 1997), stereotypes are more likely to negatively affect the targets of stereotypes. Obviously, discriminatory behavior on the part of the holders of the negative stereotypes will impact their targets. Moreover, research shows that knowledge of the stereotype’s existence can impact its targets’ behavior, even if the target does not believe in the content. Stereotype threat refers to the concern that is experienced when one feels at risk of confirming, as self-characteristic, a negative stereotype about one’s group (Steele & Aronson, 1995). As long as stereotypes about groups of people persist in society,
stigmatized individuals will be at a disadvantage in situations diagnostic of ability in stereotyped domains.

Since its conception, the relationship between stereotype threat and performance has been an intensely-studied topic in the social psychological literature. A wide range of stereotypes have been tested, including women’s lack of ability in math and science domains (Ben-Zeev, Fein, & Inzlicht, 2005; Brown & Josephs, 1999; Brown & Pinel, 2003; Davies, Spencer, Quinn, & Gerhardstein, 2002; Jamieson & Harkins, 2007; Johns, Schmader, & Martens, 2005; Keller & Dauenheimer, 2003; O’Brien & Crandall, 2003; Pronin, Steele, & Ross, 2003; Schmader & Johns, 2003; Sekaquaptewa & Thompson, 2002; Spencer, Steele, & Quinn, 1999); African-American’s underperformance on standardized tests (Aronson, Fried, & Good, 2001; Blascovich, Spencer, Quinn, & Steele, 2001; Steele & Aronson, 1995); and White males’ athletic inferiority (Stone, 2002; Stone, Lynch, Sjomeling, & Darley, 1999). In each case, concern about confirming the relevant stereotype has been shown to negatively impact the performance of stigmatized individuals.

Research has since shifted from describing threat effects to examining how stereotype threat produces its debilitating effects on performance. Although uncovering the mechanism(s) of stereotype threat performance effects is an important step in alleviating the impact of negative group stereotypes, it has proven to be a difficult task. A number of explanations have been proposed to account for threat effects in the literature (e.g., anxiety: Bosson, Haymovitz, & Pinel, 2004; Spencer et al., 1999; Steele & Aronson, 1995; expectancy: Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; arousal: Ben-Zeev et al., 2005; Blascovich et al., 2001; O’Brien & Crandall, 2003; working
memory interference: Schmader, Johns, & Forbes, 2008; motivation: Jamieson & Harkins, 2007); cognitive load: Croizet, Després, Gauzins, Huguet, Leyens, & Méot, 2004; withdrawal of effort: Stone, 2002; Stone et al., 1999), but no consensus has been reached as to the mechanism(s) though which stereotype threat has its effect on performance.

Until recently, most of the research on mediating mechanisms of stereotype threat has focused on cognitive explanations (see Schmader et al., 2008 for a review). Researchers have assumed that threatened participants are motivated to perform well (e.g. Steele & Aronson, 1995), but have not argued that motivation plays a direct role in producing threat effects. Recently, Jamieson and Harkins (2007) suggested that increases in motivation may directly lead to the performance debilitation observed under threat. The current series of experiments builds on this work and more closely explores the relationship between motivation and performance.
Chapter 1: Mere Effort and Stereotype Threat on the GRE

Researchers have argued that stereotype threat effects “are likely to be mediated in multiple ways—cognitively, affectively, and motivationally” (Steele, Spencer, & Aronson, 1992, p. 397). In the current work, we focus on the contribution of motivation to stereotype threat performance effects, specifically on Jamieson and Harkins’s (2007) mere effort account.

The mere effort account was suggested by Harkins’s (2006) analysis of the effect of evaluation on performance. This account argues that evaluation motivates participants to want to perform well which potentiates whatever response is prepotent, or most likely to be produced in a given situation (e.g., reading the color-word rather than the color of the word in the Stroop Color-Word Test). If the prepotent response is correct, the potential for evaluation facilitates performance. If the prepotent response is incorrect, and participants do not know, or lack the knowledge or time required for correction, performance is debilitated. However, if participants are able to recognize that their prepotent tendencies are incorrect and are given the opportunity to correct, performance will be facilitated. Harkins and his colleagues have found support for the mere effort account of the effect of evaluation on performance on the Remote Associates Task (Harkins, 2006), on anagrams (McFall, Jamieson, & Harkins, 2009, Exp. 1), the Stroop task (McFall et al., 2009, Exps. 2 & 3), and the antisaccade task (McFall et al., 2009, Exp. 4).
Jamieson and Harkins (2007) argued that stereotype threat, like the potential for evaluation, motivates participants, bringing into play the same process that Harkins and his colleagues (Harkins, 2006; McFall et al., 2009) implicate in the evaluation-performance relationship. Of course, other stereotype threat research has also argued that the participants subject to threat are motivated to perform well. For instance, Steele and Aronson (1995) suggest that the performance of threat participants may suffer as a result of the fact that they alternate their attention between trying to answer the items and trying to assess the self-significance of their frustration. That is, according to this account, they are indeed motivated to solve the problems. However, it is not this motivation in and of itself that impacts their performance, but rather the time spent assessing their frustration. In contrast, the mere effort account argues that motivation plays a direct role in producing the effects of threat on performance.

Mere effort shares the notion that stereotype threat energizes prepotent or dominant responses with the arousal/drive explanations proposed by O’Brien and Crandall (2003) and Ben-Zeev et al. (2005). For example, O’Brien and Crandall (2003) argue that the arousal produced by stereotype threat “is non-specific and serves to energize behavior in a nondirective way. For this reason, arousal enhances the emission of dominant responses.” (p. 783). However, as is also the case for the arousal/drive explanations for social facilitation (e.g., Cottrell, 1972), these accounts do not incorporate the correction process proposed by Harkins and his colleagues (Harkins, 2006; McFall, et al., 2009).

Jamieson and Harkins (2007) tested the mere effort account using an inhibition task, the antisaccade task (Hallett, 1978). This task requires participants to respond to a
target presented randomly on one side of the display or the other. Before the target appears, a cue is presented on the opposite side of the display. Participants are explicitly instructed to not look at this cue, but rather to look to the opposite side of the display where the target will appear. However, a reflexive-like prepotent tendency to look at the cue must be inhibited or corrected to optimize performance.

The mere effort account predicts that because the motivation potentiates the prepotent response, participants under stereotype threat will look the wrong direction, toward the cue, more often than controls. However, because this response is obviously incorrect and participants have the time required for correction, those subject to threat would also be motivated to launch correct saccades (saccades launched toward the target following successful inhibition of the reflexive saccade) and corrective saccades (saccades launched toward the target following reflexive saccades) faster than controls. Once the participants’ eyes arrive at the target area, they must then determine the target’s orientation and press the appropriate response key. The motivation to perform well would make stereotype threat participants try to respond as quickly as possible. Thus, when the participants see the target, participants subject to threat should respond more quickly than controls. Jamieson and Harkins (2007; Experiment 3) found support for each of these predictions, which taken together, produced faster overall reaction times for participants subject to stereotype threat than for no threat participants with no difference in accuracy.1

Jamieson and Harkins’s (2007) research supports the mere effort account of performance on the antisaccade task, an inhibition task framed as a measure of visuospatial capacity. However, most stereotype threat research that has focused on the female math-ability stereotype has used problems taken from standardized tests, such as
the quantitative section of the GRE general test (e.g., Ben-Zeev et al., 2005; Inzlicht & Ben-Zeev, 2003; Schmader & Johns, 2003; Spencer et al., 1999). On this task, like the RAT and anagrams, but unlike inhibition tasks like the Stroop and the antisaccade task, the correct answer is not apparent. The present research tests the contribution of mere effort to performance on this task.

The GRE quantitative test is primarily made up of two types of problems, solve problems and comparison problems, which differ in the solution approach that tends to be most efficient (see Appendix A for examples). Solve problems are standard word problems which tend to be most efficiently solved by using an equation or algorithm. Comparison problems require the test-taker to compare the quantity in one column to the quantity in the other. In this case, the most efficient approach tends to be simplifying the terms in each column or use logic, estimation, and/or intuition to find the correct answer.²

To test the mere effort account on GRE items, we must first identify the prepotent response. For example, in the case of the RAT, our own research (Harkins, 2006) identified the prepotent response and then demonstrated the effect of evaluation on this response. For anagrams (McFall et al., 2009, Exp. 1), previous research had identified the prepotent response, the tendency to try consonants in the first position of the word (e.g., Witte & Freund, 2001). The Stroop (McFall et al., 2009, Exps. 2 & 3) and the antisaccade task (McFall et al., Exp. 4; Jamieson & Harkins, 2007) are inhibition tasks, and, as such, the prepotent response (reading the word and looking at the peripheral cue, respectively) is well-established.
In the case of quantitative GRE problems, previous research suggests that participants’ prepotent tendency is to take a conventional or solving approach (i.e., compute the answer using a rule or an equation). For example, in their research on the performance of high school students on math problems taken from the SAT, Gallagher, De Lisi, Holst, McGillicuddy-De Lisi, Morely, and Cahalan (2000, Exp. 2) found that participants used a conventional (i.e., solving) approach 55.5% of the time, whereas they used the unconventional (i.e., comparison) approach 10% of the time, with the remainder of the trials comprising guesses (16.5%), omissions (9%), and unknowns (9%). Other research that has examined participants’ performance on the same problems used by Gallagher et al. (2000) has found the same pattern of results: Participants use the solving approach significantly more often than any other solution approach. For example, Gallagher and De Lisi (1994) found that their participants used the solving approach on 63% of the problems as opposed to the comparison method (32%) or guessing (5%), and Quinn and Spencer (2001; Experiment 2) found that their participants relied on the solving approach on 57% of the problems, as opposed to the comparison approach (28%), guessing (8%), or an unknown approach (7%). Furthermore, this pattern held true whether the participants were working on solve or comparison-type problems. For instance, Gallagher et al. (2000, Experiment 2) found that on solve-type problems, participants used the solving approach 66% of the time and the comparison approach 9% of the time. On the comparison problems the solving approach was used 45% of the time versus 11% for the comparison approach.

Taken together, these findings strongly suggest that when participants are asked to solve math problems, the prepotent tendency is to attempt to compute the answer using a
known formula or an algorithm (i.e., the solving approach), which is consistent with the training in mathematics that they receive. For example, Katz, Bennett and Berger (2000) point out that: “Traditional strategies [the solving approach] are the formal methods traditionally emphasized in U.S. mathematics education” (p. 41). Likewise, Stigler and Hiebert (1999) note that in the United States, “teachers present definitions of terms and demonstrate procedures for solving specific problems. Students are then asked to memorize the definitions and practice the procedures” (p. 27).

Given this prepotent response, mere effort predictions on the GRE quantitative test are straightforward. On solve problems, the prepotent tendency is often the most efficient approach. As long as the test-taker knows the correct equation to apply, solving the equation will lead to a correct answer. The stereotype threat manipulation should motivate the females to do well, which should facilitate performance on these problems. However, on comparison problems the prepotent tendency to use the solving approach will be much less efficient and often may not work at all. It is also unlikely that the participant will recognize that s/he is not using the best approach. Thus, on these problems, the potentiation of the prepotent response, solving equations, should debilitate the performance of females working under stereotype threat.

If stereotype threat facilitates performance on solve problems, but debilitates it on comparison problems, how can mere effort account for the finding that stereotype threat debilitates overall performance (e.g., Brown & Pinel, 2003; O’Brien & Crandall, 2003)? In fact, our analysis suggests that the effect of the debilitation of performance on the comparison problems should be more profound than the effect of facilitation on the solve problems. We argue that stereotype threat potentiates the prepotent response, solving
equations. Because this response is prepotent, everyone will tend to take the correct approach to solve-type problems, but females subject to stereotype threat will be more motivated to perform the computations than no threat females. In addition, the females in the stereotype threat and control groups should not differ in their basic knowledge of the relevant formulas or operations. Thus, the effect of the heightened motivation of females subject to stereotype threat is limited to the fact that they will try harder than controls to solve as many of these problems as they can. And, of course, the time limit on the test provides an upper bound for this effort.

In contrast, on the comparison problems, the potentiation of the solving approach will make it highly unlikely that the threatened females will ever even adopt the correct approach to the problem. As a result, they will be stymied from the outset. So, even though they have the same knowledge as females in the control group, this knowledge will be much less likely to be brought to bear on the problems. And the more motivated they are, the worse off they will be as they continue to attempt to use the solving approach on comparison problems (cf., producing close associates on RAT problems (Harkins, 2006). Consequently, stereotype threat females should perform better on solve problems, more poorly on comparison problems, and more poorly overall than females in a control group. This hypothesis was tested in Experiment 1.

Experiment 1

The GRE test in Experiment 1 included comparison and solve problems. Because mere effort predicts that stereotype threat will potentiate the prepotent response (Jamieson & Harkins, 2007), threatened females are predicted to perform more poorly on comparison problems and better on solve problems than females not subject to stereotype
threat. We also predicted that overall performance on the test would be worse for females subject to stereotype threat than females in the control group, because the debilitation of performance on the comparison problems will be greater than the facilitation of performance on the solve problems. The performance of males should not differ as a function of stereotype threat.

Method

Participants

64 Northeastern University undergraduates (32 male, 32 female) participated in this experiment in exchange for course credit.

Materials

The math test consisted of 30 multiple-choice problems taken from the quantitative section of the GRE. The test included 15 comparison problems and 15 solve problems, and was presented as a paper and pencil test with scratch-paper provided for showing work.

Because each problem actually appeared on a GRE test, we obtained performance norms, as indexed by the proportion of test-takers answering each problem correctly out of all those who attempted that problem. We selected problems by first randomly picking 12 problems of each type from problems that varied in their solution rates from 35% to 65%. We then picked the final three problems for each type so that mean overall accuracy averaged 50% for each problem set. (comparison range = 38% to 60%; solve range = 42% to 63%).

Problems were randomized throughout the test with the constraint that no one type of problem could appear in more than three consecutive problems. This procedure
produced a test on which comparison and solve problems were dispersed evenly throughout the test. Of the first 16 problems, 8 were comparison and 8 were solve problems. Participants worked on the test for 20 minutes and were instructed to complete as many problems as accurately as possible. All participants were given 2 practice problems (1 comparison, 1 solve) prior to beginning and were instructed not to use calculators.

Procedure

Threat manipulations were adapted from Jamieson and Harkins (2007). In the threat condition participants were instructed they would be taking a math test which had been shown to produce gender differences. In the no threat conditions, participants were also told they were taking a math test, but that the test had been shown not to produce gender differences. This explicit stereotype threat manipulation has been shown to produce performance effects in previous research (e.g., Brown & Pinel, 2003; Jamieson & Harkins, 2007; Keller, 2002; Keller & Dauenheimer, 2003; O’Brien & Crandall, 2003; Spencer et al., 1999). No specific mention was made as to whether men outperformed women or vice versa, only that gender differences did or did not exist on the task. Participants were expected to infer that women would perform more poorly than men based on the societal stereotype that men are superior to women in mathematical ability.

Each participant responded to a questionnaire upon completion of the math test. Two questions allowed us to evaluate the effectiveness of the stereotype threat manipulation: “To what extent are there gender differences in performance on this task?” (1 = “no gender differences” and 11 = “gender differences”); and “Who do you believe performs better on this task?” (1 = “males perform better,” 6 = “males and females
perform the same,” and 11 = “females perform better”). Participants were also asked to rate how difficult the test was, how interesting the task was, how anxious they felt about their performance, and how much effort they put into the task, all on 11-point scales.

Results

Manipulation Check for Stereotype Threat

The manipulation checks were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) x 2 (gender: male vs. female) ANOVAs. Condition and gender were both analyzed as between subject factors. Participants in the stereotype threat condition reported that gender differences existed to a greater extent ($M = 6.63, SD = 2.94$) than participants in the no stereotype threat condition ($M = 4.03, SD = 2.40$), $F(1, 60) = 15.58$, $p < .001$, $d = 1.02$. Participants in the stereotype threat condition also reported that men performed these tasks better than women to a greater extent ($M = 3.91, SD = 1.38$) than participants in the control condition ($M = 5.44, SD = 1.48$), $F(1, 60) = 17.91$, $p < .001$, $d = 1.09$. These results indicate that the stereotype threat manipulation used in the current experiment was successful. Participants in the threat condition were aware of the negative group stereotype and women were expected to perform more poorly than men.

Math Test Performance

To test for the traditional stereotype threat effect on GRE test performance (e.g., Schmader & Johns, 2003, Experiment 3), we analyzed the total percentage of problems solved (total number solved correctly/total attempted (problems for which answers were given)) in a 2 (condition: stereotype threat vs. no stereotype threat) x 2 (gender: male vs. female) ANOVA. Replicating previous research, we found a Stereotype Threat x Gender interaction, $F(1, 60) = 4.76, p = .03$, $d = .56$. Pairwise contrasts (Kirk, 1995) revealed
that female participants in the stereotype threat condition performed more poorly ($M = 42.44\%, SD = 15.06\%$) than females in the no stereotype threat control group ($M = 54.16\%, SD = 13.85\%), F(1, 60) = 6.18, p < .03, d = .64$, whereas males did not differ as a function of stereotype threat, $p > .50$ (see Figure 1). Thus, we were successful in replicating previous findings in this domain.

We also tested for differences in the total number of problems attempted in the same 2 x 2 design, and found none, $ps > .20$. Overall, participants attempted 17.25 problems.

To test for the effects suggested by the mere effort hypothesis we examined the performance of participants on solve and comparison-type problems on two measures: the number of problems attempted and the percentage of these problems that were correctly answered. These data were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) x 2 (gender: male vs. female) x 2 (problem type: comparison vs. solve) ANOVAs with condition and gender as between subjects factors and problem type as a within subjects factor (see Table 1 for means and standard deviations). Pairwise contrasts (Kirk, 1995) were used to decompose significant interactions.
Figure Caption. Experiment 1: Total percentage correct as a function of gender and stereotype threat condition. ST = stereotype threat, NST = no stereotype threat

Problems Attempted. Analysis of the number of problems attempted revealed a significant three-way interaction between gender, stereotype threat, and problem type, $F(1, 60) = 12.94, p = .001, d = .93$. Females subject to stereotype threat attempted more solve-type problems ($M = 9.44, SD = 1.83$) than females who were not subject to threat ($M = 7.81, SD = 2.43$), $F(1, 60) = 13.25, p < .01, d = .94$. In contrast, males in the stereotype threat condition attempted fewer solve problems ($M = 8.13, SD = 2.13$) than males in the no stereotype threat condition ($M = 9.06, SD = 2.74$), $F(1, 60) = 4.24, p < .05, d = .53$. 
Neither females nor males differed as a function of stereotype threat in the number of comparison problems attempted, \( ps > .20 \). All means and standard deviations are presented in Table 1.

*Percentage Solved.* Analysis of the percentage of problems solved also yielded a significant three-way interaction between stereotype threat, gender, and problem type, \( F(1, 60) = 6.28, p < .02, d = .65 \). On the solve problems, there were no differences in the percentage of problems solved by females or males, \( ps > .20 \).

However, females in the stereotype threat condition solved a smaller percentage of the comparison problems they attempted (\( M = 27.07\%, SD = 14.00\% \)) than females in the no stereotype threat control condition (\( M = 54.33 \%, SD = 14.06\% \)), \( F(1, 60) = 26.45, p < .001, d = 1.33 \). Males did not differ as a function of stereotype threat, \( p > .20 \). All means and standard deviations are presented in Table 1.

Table 1: *Experiment 1: GRE performance as a function of gender, problem type, and stereotype threat condition.*
Ancillary Measures. Participants did not differ in their ratings of how interesting the task was, how difficult the task was, how much effort they put into the task, or how anxious they felt about their performance, \( ps > .20 \)

**Discussion**

In Experiment 1 we found that overall GRE math performance for females subject to stereotype threat was worse than for females who were not subject to threat. However, we also found that females subject to stereotype threat attempted more solve problems than females under no threat but solved the same percentage of these problems correctly.
Thus, threatened females outperformed no threat females (i.e., correctly solved more solve problems). In contrast, on comparison problems, females under threat attempted the same number of problems as control females, but solved a smaller percentage of these problems correctly. Therefore, on comparison problems, females under threat performed more poorly than controls.

Consistent with the mere effort account, when the prepotent, conventional approach was correct, stereotype threat participants performed better than controls, whereas when this response was incorrect, their performance was debilitated. Using methods suggested by Meng, Rosenthal, and Rubin (1992) for comparing correlated correlation coefficients, we compared the magnitude of the effect of stereotype threat on comparison problems to the magnitude of the effect of threat on solve problems, ignoring the difference in sign. As predicted, the magnitude of the debilitation effect on the number of comparison problems solved was significantly greater ($r = -.59$) than the magnitude of the facilitation effect on the number of solve problems solved ($r = .21$), $Z = 3.08$, $p < .001$, accounting for the overall finding of poorer performance on the part of females subject to stereotype threat than females not subject to threat.\(^3\)

**Experiment 2**

Previous stereotype threat research has examined the effect of problem difficulty on females’ math performance. For example, O’Brien and Crandall (2003) found that stereotype threat facilitated performance on a set of simple math problems (e.g., $148 \times 253 = ?$), but debilitated performance on a set of difficult math problems (quantitative SAT). However, the mere effort account would suggest that it is not problem difficulty, per se, that accounts for the performance effects observed by O’Brien and Crandall
(2003), but rather the fact that the prepotent response is generally correct for simple problems but is more often incorrect for difficult problems.

Solving O’Brien and Crandall’s (2003) simple multiplication problems is an extremely well-learned response; stereotype threat should potentiate the prepotent response, facilitating performance. The difficult SAT math problems used by O’Brien and Crandall (2003) are just like the problems studied by Gallagher et al. (2000). Experiment 1 showed that when these problem types are intermixed, potentiation of the prepotent response, the solving approach, ensures that, overall, females in the threat condition perform more poorly than females in the control condition. This is consistent with what O’Brien and Crandall (2003) found on their difficult SAT problems. However, in fact, in Experiment 1, the females subject to threat performed better than females in the no threat condition on which the prepotent response was correct (i.e., solving approach applied to solve problems) even though the problems were difficult.

In Experiment 2, we tested the hypothesis that threat participants would perform more poorly than control participants on problems on which the prepotent response is incorrect (comparison problems) even though the problems are “easy.” This possibility was tested by selecting solve and comparison problems that had solution rates that either averaged 75% (easy) or 50% (difficult). If stereotype threat potentiates the prepotent tendency, the pattern of performance should be the same on both easy and difficult items. That is, the experience of stereotype threat should debilitate performance on comparison problems and facilitate performance on solve problems at each level of problem difficulty. The overall level of performance should just be higher for easy problems than for difficult ones.
In addition, in Experiment 2, participants worked on problems of only one type: comparison problems or solve problems. Manipulating problem type as a between, rather than a within subjects variable, allows us to rule out the possibility that the pattern of findings of Experiment 1 is a result of the fact that participants were exposed to both types of problems during task performance.

Method

Participants

120 Northeastern University undergraduate students participated in this experiment in exchange for course credit. For the remaining experiments, only females were used because the stereotype threat manipulation did not significantly affect the performance of males in Experiment 1.

Materials

Participants were given one of four 15 problem, multiple-choice math tests. All problems were taken from the quantitative section of the GRE test. Each math test consisted of only one type of problem, either all comparison or all solve problems. In addition, a difficult and an easy version of each comparison and solve test were administered. Thus, the four tests differed on two dimensions, problem type and difficulty. The difficult tests consisted simply of the 15 problems of a particular type from the test used in Experiment 1. The easy comparison and solve tests were taken from the same pool of questions as the difficult problems, but selected for higher scores. The easy comparison test had a mean accuracy of 75% with a range of 67% to 86% and the easy solve test had a mean accuracy of 75% with a range of 62% to 83% accuracy. Each participant was given 2 practice trials of whichever problem type they would be working
on prior to beginning the test. Participants worked on the test for 10 minutes and were instructed to complete as many problems as they could as accurately as possible. The ratio of number of problems presented to time to complete the test remained consistent from Experiment 1 to Experiment 2.

Procedure and Data Preparation

All experimental manipulations, questionnaires, and data preparation techniques were the same as those described in Experiment 1.

Results

The manipulation checks, GRE test performance, and ancillary measures were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) x 2 (problem type: comparison vs. solve) x 2 (difficulty level: difficult vs. easy) ANOVAs. Condition, problem type, and difficulty level were all between subjects factors.

Manipulation Checks

Stereotype Threat. Stereotype threat participants reported that gender differences existed on the test to a greater extent ($M = 6.07, SD = 2.63$) than no stereotype threat participants ($M = 4.25, SD = 2.97$), $F(1, 112) = 12.64, p = .001, d = .67$. Also, stereotype threat participants reported that males perform better than females on the test to a greater extent ($M = 4.21, SD = 1.51$) than no stereotype threat participants ($M = 5.68, SD = 1.37$), $F(1, 112) = 32.21, p < .001, d = 1.07$. As in Experiment 1, the stereotype threat manipulation was successful.

Test Difficulty. Participants rated the difficult math test as more difficult ($M = 7.52, SD = 1.70$) than the easy math test ($M = 4.88, SD = 2.53$), $F(1, 112) = 46.55, p < .001, d = 1.29$. Thus, the math tests which were designed to be more difficult were
perceived as such. Participants also rated the solve problem sets as more difficult ($M = 6.60$, $SD = 2.51$) than the comparison problem sets ($M = 5.80$, $SD = 2.50$), $F (1, 112) = 4.30$, $p = .04$, $d = .39$.

**Math Performance**

As in Experiment 1, we analyzed two measures of performance on the GRE problems: problems attempted and percentage solved. And, once again, pairwise contrasts (Kirk, 1995) were used to decompose significant interactions (see Table 2 for means and standard deviations).

**Problems Attempted.** Analysis of problems attempted yielded a significant two-way interaction between stereotype threat and problem type, $F (1, 112) = 18.77$, $p < .001$, $d = .82$. As in Experiment 1, females in the stereotype threat condition attempted more solve problems ($M = 10.37$, $SD = 2.24$) than females in the control condition ($M = 9.13$, $SD = 2.37$), $F (1, 112) = 4.94$, $p < .03$, $d = .42$. On the other hand, females in the stereotype threat condition attempted fewer comparison problems ($M = 11.17$, $SD = 2.63$) than females not subject to stereotype threat ($M = 13.33$, $SD = 1.58$), $F (1, 112) = 70.42$, $p < .001$, $d = 1.59$. The main effect for problem type, $F (1, 112) = 40.59$, $p < .001$, $d = 1.20$, must be interpreted in the context of this interaction.

Participants also attempted more easy problems ($M = 11.67$, $SD = 2.43$) than difficult ones ($M = 10.33$, $SD = 2.80$), $F (1, 112) = 11.55$, $p < .001$, $d = .64$, but problem difficulty did not interact with stereotype threat, $p > .20$.

**Percentage Solved.** Analysis of the percentage of problems solved also yielded a significant two-way interaction between stereotype threat and problem type, $F (1, 112) = 12.94$, $p < .001$, $d = .68$. As in Experiment 1, females under threat did not solve a smaller
percentage of solve problems than control participants. In fact, there was a tendency for them to solve more of these problems than no threat participants ($M_{\text{stereotype threat}} = 70.23\%, \ SD = 22.00\%; \ M_{\text{no stereotype threat}} = 63.86\%, \ SD = 24.84\%$), $F(1, 112) = 2.51, \ p = .12, \ d = .30$. Also as in Experiment 1, females under stereotype threat solved a smaller percentage of the comparison problems that they attempted ($M = 45.88\%, \ SD = 19.31\%$) than females not under threat ($M = 60.70\%, \ SD = 17.96\%$), $F(1, 112) = 12.65, \ p < .001, \ d = .67$. All means and standard deviations are presented in Table 2. The main effect for problem type, $F(1, 112) = 21.81, \ p < .001, \ d = .88$, must be interpreted in the context of the Stereotype Threat x Problem Type interaction.

As was the case for problems attempted, there was a highly reliable problem difficulty effect, $F(1, 112) = 86.64, \ p < .001, \ d = 1.76$. Participants solved a greater percentage of easy problems ($M = 73.88\%, \ SD = 17.10\%$) than difficult problems ($M = 46.46\%, \ SD = 19.26\%$), but problem difficulty did not interact with stereotype threat, $p > .20$.

Table 2. Experiment 2: GRE performance as a function of test difficulty, problem type, and stereotype threat condition. ST = stereotype threat, NST = no stereotype threat.
Ancillary Measures

Participants reported that they exerted more effort on the easy problems ($M = 9.10$, $SD = 1.54$) than on the difficult problems ($M = 8.25$, $SD = 1.87$), $F (1, 112) = 7.21$, $p < .01$, $d = .51$. This effect did not interact with problem type or stereotype threat condition, $ps > .20$. Participants did not differ in their ratings of how interesting the task was or how anxious they felt about their performance, $ps > .20$.

Discussion

In Experiment 1 we found that female participants subject to stereotype threat performed more poorly overall than participants in the other conditions, which did not differ among themselves. However, when we looked at performance as a function of
problem type, consistent with mere effort predictions, females subject to stereotype threat performed more poorly on comparison problems, but better on the solve problems than control participants.

In Experiment 2, problem type and problem difficulty were manipulated orthogonally to determine whether the pattern of performance observed in Experiment 1 would remain consistent across difficulty level. Research by O’Brien and Crandall (2003) shows that stereotype threat facilitates performance on simple math problems but debilitates performance on difficult problems. This interpretation would suggest that if the math test used in Experiment 1 were made up of easier problems, then stereotype threat participants should perform better than controls. However, mere effort argues that it is problem type, not difficulty, which accounts for the pattern of performance observed in Experiment 1. That is, the pattern of performance should not differ as a function of difficulty level, but instead should depend on whether the prepotent response is correct or not.

Consistent with the motivation based, mere effort account, females under threat attempted fewer comparison problems, and solved a smaller percentage of these problems correctly than control participants. So, as in Experiment 1, females under threat performed more poorly than females in the no threat condition on comparison problems. However, in Experiment 2, this pattern characterized performance not only on difficult problems, as in Experiment 1, but also on easy ones. Consistent with the mere effort account, whether the problems were easy or difficult, females subject to stereotype threat performed better than females not subject to threat on solve problems, but performed more poorly than females not subject to threat on comparison problems. Thus,
Experiment 2 demonstrates that when one controls for the prepotency of the correct response, problem difficulty does not have an effect on the pattern of performance produced by stereotype threat. Of course, participants solve more easy problems than difficult ones, but whether the problems are easy or difficult, stereotype threat potentiates the prepotent response, producing better performance on solve problems, but worse performance on comparison problems.

Because the mere effort hypothesis argues that stereotype threat participants are more motivated to perform well, they should be motivated to correct a prepotent response when that response is incorrect, they recognize that it is incorrect, and they have the opportunity to correct it (Jamieson & Harkins 2007, Exp. 3). Experiments 1 and 2 demonstrate that threatened females perform more poorly on comparison problems than females who are not under threat. In this case, the participants may not know that their approach is inappropriate, and even if they do recognize this, they do not know the appropriate approach to take.

Thus, the next step in this research plan is informing threatened participants that their tendency to employ the traditional solving approach on all math problems is incorrect, and instead, they should use logic/intuition to solve the comparison type problems. This information should allow threatened participants to recognize when the prepotent solving approach is incorrect and allow them to perform at least on par with non-threatened stigmatized individuals.
Chapter 2: A Motivation Based Intervention Strategy

The previous research suggests that stereotype threat motivates participants to want to do well and this motivation potentiates whatever response is prepotent on the given task. However, motivation to perform well should also lead to an effort to correct incorrect responses if the participant recognizes that his/her response is incorrect, knows the correct response, and has the opportunity to make it. On inhibition tasks, the fact that the prepotent response is incorrect is quite obvious to the participants, as is the way in which it can be corrected. Thus, given the opportunity, the more motivated participants (participants who are subject to evaluation or stereotype threat) are able to make the correction, and respond more quickly than less motivated participants (not subject to evaluation or stereotype threat).

In contrast, in Harkins’s (2006) research on the Remote Associates Test, regardless of how motivated the participants might have been, they are unlikely to have been aware that their prepotent tendency (generating close associates for triad members) was incorrect. In a pilot study, Harkins (2006) tested the efficacy of a simple instruction on how to approach the task. All participants were told that their performance would be subject to evaluation. One third of these participants were told that if they wanted to succeed, they should refrain from generating close associates. Instead they were to simply register the triad members, and then wait for the answer to “pop up.” Another third were told that if they wanted to succeed, they should generate as many close associates as possible. The final third, a control condition, were told nothing. Harkins (2006) found
that participants, who were told not to generate close associates, but to wait for the answer to emerge, outperformed the other two groups, which did not differ from each other.

**Experiment 3**

In Experiment 3, we took a similar approach to improving the performance of females subject to stereotype threat by informing participants as to the correct solution approach to take to each type of problem. We hypothesized that stereotype threat participants’ debilitated performance on comparison problems results from the potentiation of the prepotent response (to use a solving approach), which is incorrect for comparison problems. The instructions not only make it more likely that participants will recognize when their solution approach to a problem is incorrect, but also suggests the correct approach. If stereotype threat participants are able to reduce their reliance on the solving approach on comparison problems we would expect that their performance would improve compared to stereotype threat participants not given instructions. On the solve problems, however, stereotype threat participants should still outperform controls because the prepotent response, which is potentiated by threat, is correct and the instructions should act to reinforce this tendency.

Previous research has also examined intervention strategies for reducing the effect of threat on females’ math performance. For example, Johns et al. (2005) taught females about stereotype threat, informing them that any anxiety that they were feeling could be the result of “negative stereotypes that are widely known in society and have nothing to do with your actual ability to do well on the test” (p. 176). They found that these females performed better than females not given this information and argued that teaching threat
enabled females to attribute any anxiety felt during the math test to gender stereotypes. In contrast, the intervention strategy tested in the current research provides participants with information that allows them to recognize the behavior that produces the performance debilitation (reliance on the solving approach), providing them with the opportunity to channel their motivation more effectively.

Method

Participants

64 Northeastern University female undergraduates participated in this experiment in exchange for course credit. Only females were used in this research because the stereotype threat manipulation did not significantly affect the performance of males in Experiment 1.

Materials

All materials were identical to those described in Experiment 1.

Procedure

Manipulations and questionnaires were identical to those described in Experiment 1, with the exception that participants in the instruction condition received instructions prior to beginning the practice problems. Instructions were adapted from a GRE preparation book (Lurie, Pecsenye, Robinson, & Ragsdale, 2004). Participants receiving instructions were told, “This test consists of two types of problems, comparison and solve problems. You may be familiar with the solve problem format. For these problems, you are given a problem, offered five solutions, and asked to pick one. These problems generally ask you to apply an equation that you have previously learned or to use mathematical rules to work through an equation to obtain a solution. For example, you
may be given the dimensions of a geometrical figure and asked to obtain the area. To solve a problem such as this you just need to apply the area formula.

The remaining problems consist of comparison problems, which require you to compare the quantity in Column A to the quantity in Column B. It is important to remember that you do not have to calculate the exact values in each column to solve the problem. After all, your goal is simply to compare the two columns. It’s often helpful to treat the two columns as if they were two sides of an equation. Anything you do to both sides of an equation, you can also do to the expressions in both columns. Try planning your approach and simplifying each column rather than calculating.” Participants not assigned to the instruction condition were not given these instructions prior to working on the practice problems, and their procedure was identical to that described in Experiment 1.

In Experiment 3, participants were again provided with scratch-paper. We argue that the prepotent response on the GRE problems is to use a solving approach, which is potentiated by stereotype threat, debilitating performance on comparison problems. Thus, the threat participants should show evidence of using the solving approach more on comparison problems than control participants. The effect of the instruction should be to reduce the threat participants’ reliance on this approach. To test these predictions, a rater blind to condition computed the percentage of problems on which the participants’ scratch-paper showed evidence that they used the solving approach, whether or not the attempt produced the correct answer. 4

Results

Manipulation Checks
Manipulation checks were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) x 2 (instruction: instructions given vs. no instructions given) ANOVAs. Both condition and instruction were analyzed as between subjects factors. Stereotype threat participants rated the test as revealing gender differences to a greater extent ($M = 6.47, SD = 2.85$) than no stereotype threat participants ($M = 3.56, SD = 2.56$), $F (1, 60) = 18.35, p < .001, d = 1.11$. Participants in the stereotype threat condition also indicated that males performed better than females on this test to a greater extent ($M = 4.00, SD = 1.83$) than controls ($M = 5.30, SD = 1.14$), $F (1, 60) = 11.55, p = .001, d = .88$.

Math Performance

To test for the traditional stereotype threat effect, we analyzed the total percentage solved correctly (total number correct/total attempted) in a one-way (stereotype threat/no instructions vs. no stereotype threat/no instruction) ANOVA. As in Experiment 1, stereotype threat participants solved a smaller percentage of problems ($M = 39.23\%, SD = 9.25\%$) than no stereotype threat participants ($M = 49.53\%, SD = 3.09\%$), $F (1, 60) = 7.12, p < .02, d = .97$. We also tested for differences in the total number of problems attempted in the same one-way design, and found none, $F < 1$. Overall, participants attempted 16.97 problems.

Problems attempted and percentage solved were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) x 2 (instruction: instructions given vs. no instructions given) x 2 (problem type: comparison vs. solve) ANOVAs. Condition and instruction were analyzed as between subjects factors whereas problem type was analyzed as a within subjects factor. Pairwise contrasts (Kirk, 1995) were used to
decompose significant interactions. All means and standard deviations are presented in Table 3.

Problems Attempted. Analysis of the problems attempted revealed a Stereotype Threat x Problem Type interaction, $F(1, 60) = 37.66, p < .0001, d = 1.58$. Participants in the stereotype threat condition attempted more solve problems ($M = 9.41, SD = 1.97$) than participants in the no threat condition ($M = 8.25, SD = 2.03$), $F(1,60) = 29.74, p < .0001, d = 1.41$, whereas participants in the no stereotype threat condition attempted more comparison problems ($M = 9.00, SD = 2.24$) than participants in the stereotype threat condition ($M = 8.44, SD = 2.26$), $F(1,60) = 8.03, p < .02, d = 1.03$. No other effects were reliable.

Percentage Solved. Analysis of the percentage of problems solved correctly yielded a reliable Stereotype Threat x Instruction x Problem Type interaction, $F(1, 60) = 6.18, p < .02, d = .64$. As in Experiment 1, when not given instructions, participants in the stereotype threat condition did not differ from no stereotype threat participants in the percentage of solve problems solved ($M_{stereotype} = 53.09\%, SD = 17.37\%; M_{no stereotype} = 47.10\%, SD = 20.75\%$), $p > .20$. Replicating Experiment 1, under no instructions participants in the stereotype threat condition solved a smaller percentage of the comparison problems ($M = 23.49\%, SD = 14.48\%$) than participants in the no stereotype threat condition ($M = 51.39\%, SD = 11.43\%$), $F(1, 60) = 31.40, p < .001, d = 1.45$.

In Experiment 3, we were interested in whether describing the appropriate approach to take to each type of problem would be sufficient to improve threat participants’ performance on comparison problems, and this is exactly what we found. In
fact, when instructions were provided, stereotype threat participants solved the same percentage of comparison problems ($M = 44.77\%, SD = 5.64\%$) as participants not subject to threat ($M = 46.71\%, SD = 12.73\%$), $p > .20$. There was also a tendency for stereotype threat/instruction participants to correctly answer a higher percentage of solve problems ($M = 59.01\%, SD = 17.37\%$) than no threat/instruction participants ($M = 51.80\%, SD = 19.18\%$), $F (1, 60) = 2.10, p = .15, d = .37$.

The overall analysis also produced a significant instruction main effect, $F (1, 60) = 6.19, p < .05, d = .64$, a Stereotype Threat x Instruction interaction, $F (1, 60) = 6.17, p < .02, d = .64$, a problem type main effect, $F (1,60) = 20.11, p < .001, d = 1.16$, and a Stereotype Threat x Problem Type interaction, $F (1, 60) = 18.69, p < .001, d = 1.12$, each of which must be interpreted in the context of the three-way interaction.

Table 3. Experiment 3: GRE performance as a function of instruction, problem type, and stereotype threat condition.
Solution Approach Analysis. The percentage of trials on which participants showed that they used the prepotent, solving approach was analyzed in a 2 (condition: threat vs. no threat) x 2 (instruction: instructions given vs. no instructions given) x 2 (problem type: comparison vs. solve) ANOVA. Condition and instruction were analyzed as between subjects factors whereas problem type was analyzed as a within subjects factor. Pairwise contrasts were used to decompose interactions.

Overall, participants used the solving approach on 65% of the problems, consistent with the argument that this approach is prepotent. The analysis produced a significant Condition x Instruction x Problem Type interaction, $F (1, 60) = 3.74, p = .05, d = .50$. The main effects for problem type, $F (1, 60) = 46.59, p < .001, d = 1.76$,
When not given instruction, threat participants used the solving approach significantly more often on comparison problems ($M = 70.24\%, SD = 19.39\%$) than no threat participants ($M = 48.12\%, SD = 9.98\%, F(1, 60) = 31.46, p < .001, d = 1.45$). This analysis also showed that threat participants significantly reduced their use of the solving approach on comparison problems when given instruction ($M = 44.67\%, SD = 20.78\%), F(1, 60) = 31.48, p < .01, d = 1.45$. In fact, there was no difference between threat and no threat participants in the use of the solving approach on comparison problems when instructions were given ($M = 47.20\%, SD = 15.63\%), F < 1$.

Table 4. Experiment 3: Percentage of problems on which the solving approach was used as a function of instruction, problem type, and stereotype threat condition.
Mediation. We argue that use of the solving approach on comparison problems produces the performance debilitation observed under threat. To test this argument, we conducted a mediation analysis following the procedures suggested by Kenny, Kashy, and Bolger (1998) on the number of comparison problems solved incorrectly. As shown in Figure 2, the use of the prepotent, solving approach, mediated the effect of stereotype threat on comparison problem performance, Sobel $Z = 2.07, p = .039$.\(^5\)

*Figure 2*
Figure Caption. Number of times the prepotent solving approach was used on comparison trials as a mediator of number of comparison problems answered incorrectly in Experiment 3. Coefficients in parentheses indicate zero-order correlations. Coefficients not in parentheses represent parameter estimates for a recursive path model including both predictors. Asterisks (*) indicate parameter estimates or correlations that differ from zero at $p < .05$, double asterisks (**) indicate parameter estimates or correlations that differ from zero at $p < .01$.

Ancillary Measures
Participants subject to stereotype threat reported that the test was more difficult ($M = 7.91$, $SD = 1.75$) than controls ($M = 6.97$, $SD = 1.79$), $F(1, 60) = 4.46, p = .04, d = .55$. There were no significant differences in ratings of how interesting the task was, how much effort was put into the task, performance expectations, or anxiety, $ps > .20$.  

Discussion

When not provided with instruction, the overall GRE math performance for females under threat was worse than that of females who were not subject to threat. As in Experiment 1, when not given instruction, threat participants outperformed no threat participants on solve problems by attempting more problems with no cost in accuracy, whereas they performed more poorly than no threat participants on comparison problems. Consistent with the mere effort account, the solution approach analysis showed that threat participants used the prepotent solving approach significantly more often than no threat participants on these problems, which debilitated performance (see Figure 2). As in Experiment 1, we used the methods suggested by Meng et al. (1992) to compare the magnitude of the effect of stereotype threat on comparison problems to the magnitude of the effect of threat on solve problems. Replicating the previous finding, we found that the magnitude of the debilitation effect on the number of comparison problems solved was significantly greater ($r = -.70$) than the magnitude of the facilitation effect on the number of solve problems solved ($r = .35$), $Z = 3.41, p < .001$, accounting for the overall poorer performance on the part of females subject to stereotype threat than females not subject to threat.

We attempted to improve the performance of threat participants by providing instructions designed to reduce their reliance on the solving approach on comparison
problems. This manipulation had a significant impact on performance, as the difference between threat and no threat participants in the percentage of comparison problems solved was eliminated. The solution approach data suggest that threat participants’ improvement on comparison problems resulted from the fact that females under threat took the instructions into account and reduced the extent to which they relied on the solving approach when they tried to solve comparison problems.

In the current research, although providing instructions did improve the performance of females subject to threat on comparison problems, this improvement was only sufficient to bring their overall performance to parity with the performance of females not under threat. As shown in Table 5, when instructions were given, overall, females under threat performed as well as females not subject to threat, $F < 1$. If females subject to threat are motivated to do well and have been given instructions as to how to improve their performance, one might argue that their performance should exceed that of females not under threat, who are less motivated. However, females under threat must still cope with the potentiation of the prepotent response, which works against the instructions, and this could limit the amount of improvement that is possible.

*Table 5. Experiment 3: Total percentage correct as a function of instruction and stereotype threat condition. ST = stereotype threat, NST = no stereotype threat.*
Johns et al. (2005) argue that their teaching intervention externalizes arousal, which improves performance. In that research the experimenter made it clear that the negative stereotypes about math performance have nothing to do with the participants’ ability. Under these circumstances, participants may not have felt the need to attempt to disconfirm a stereotype that the experimenter has already discounted. In contrast, the instruction manipulation used in Experiment 3 does nothing to reduce participants’ motivation to disconfirm the stereotype, but rather shows them how to apply their motivation more effectively. Consistent with this argument, we find that even with instructions, threat participants outperform control participants on solve problems, suggesting that they are more motivated than these participants. Thus, the current

<table>
<thead>
<tr>
<th>Condition</th>
<th>Percentage Correct</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Instruction</td>
<td>Instruction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>NST</td>
<td>49.53</td>
<td>12.37</td>
<td>49.03</td>
</tr>
<tr>
<td>ST</td>
<td>38.23</td>
<td>9.25</td>
<td>52.07</td>
</tr>
</tbody>
</table>
research suggests another means of reducing the negative effects of stereotypes on the test performance of stigmatized group members. This research suggests that performance can be improved not only by breaking the link between stereotypes and performance (e.g., Johns et al., 2005), but also by assisting stigmatized individuals in channeling their efforts to best take advantage of their heightened motivational state.

Although the data presented in Experiments 1-3 indicate that mere effort can provide an account for the effect of stereotype threat on performance on the quantitative GRE, each of these experiments used an explicit manipulation of threat. Recent research suggests that the performance effects produced by more subtle manipulations may be due to deficits in cognitive capacities, rather than increases in motivation (Stone & McWhinnie, 2008). Thus, Chapter 3 explores the role of motivation in subtle stereotype threat situations.
Chapter 3: The Role of Motivation in Subtle Stereotype Threat

Although the research presented in Chapter 1 is consistent with a mere effort account of the effect of stereotype threat on GRE-Q performance, current conceptualizations of stereotype threat suggest that threat may be multiply mediated (e.g., Steele et al., 2002). Similarly, Shapiro and Neuberg (2007) argue that stereotype threat is an umbrella term that actually encompasses multiple types of threat, and that different mechanisms, or combinations of mechanisms, may mediate different types of threat.

Consistent with a multiple threats outlook, recent work by Stone and McWhinnie (2008) suggest that the mechanisms that account for the effect of blatant stereotype threat on performance may differ from those that produce subtle stereotype threat effects. Blatant stereotype threat manipulations, like the one used in Experiments 1-3 of the current research, explicitly inform participants that the task at hand measures attributes related to a negative stereotype and/or informs targets that their group tends to perform more poorly on the task in comparison to some other group (e.g., Jamieson & Harkins, 2007; Spencer et al., 1999; Steele & Aronson, 1995; Stone et al., 1999; Stone & McWhinnie, 2008).

On the other hand, subtle stereotype threat manipulations do not make any explicit mention of group stereotypes. Instead, subtle threat is produced by placing stigmatized group members in a minority status performance situation that assess ability in a stereotyped domain (e.g., Danso & Esses, 2001; Inzlicht & Ben-Zeev, 2000; Marx &
Goff, 2005; Sekaquaptewa & Thompson, 2002; Schmader & Johns, 2003; Stone & McWhinnie, 2008).

Stone and McWhinnie (2008) suggest that motivational mechanisms may account for the effects of blatant stereotype threat on performance, whereas subtle manipulations of threat may work through a different mechanism: working memory deficits. When the stereotyped task is explicitly framed as measuring an attribute that relates to a negative ingroup stereotype, there is no ambiguity about the performance situation and Stone and McWhinnie argue that targets attempt to "minimize mistakes and avoid failure that would confirm the negative stereotype" (p. 446). The effects of this motivation to avoid failure can be seen in research that has shown that participants subject to threat take advantage of explanations that allow them to deflect responsibility for their performance (e.g., Keller & Dauenheimer, 2002). The research presented in Chapter 1 also used a blatant manipulation of threat, but the stigmatized participants were not provided with the opportunity to avoid judgment in the context of the stereotype. It is under these conditions that we would argue that threat participants are motivated to perform well, producing the effects predicted by mere effort.

On the other hand, subtle manipulations of threat cue the participant that the experiment might examine stereotypes, but does not make any explicit mention of stereotypes or group differences. For instance, a common subtle threat manipulation requires female participants to perform a math test in the presence of male confederates (e.g. Schmader & Johns, 2003). Knowledge of the negative stereotype regarding women’s math ability and females’ solo status suggests that the female participant’s performance may be compared to that of the males. However, the potential for being
evaluated in the context of the stereotype must be inferred from characteristics of the situation. Thus, without explicit stereotype threat instructions, stigmatized individuals experience ambiguity regarding the presence of threat.

Stone and McWhinnie (2008) argue that the ambiguity produced by subtle manipulations of threat lead participants to expend cognitive and emotional resources on reducing uncertainty about the presence of bias, resulting in a working memory deficit and debilitated performance. Thus, the performance effects reported in research that has used subtle manipulations of threat (e.g., Croizet et al., 2004; Inzlicht & Ben-Zeev, 2003; Schmader & Johns, 2003) may be produced not by motivation, but instead by a cognitive mechanism (i.e. working memory deficits).

If there is a difference in the effects produced by these different types of stereotype threat manipulations, then a manipulation of subtle threat would be expected to produce a different pattern of performance on the GRE-Q test than what was observed in Experiments 1-3. Instead of performing better on solve type problems and worse on comparison type problems compared to controls, as was the case under blatant threat, females subject to subtle threat may perform more poorly on both types of problems because their ability to solve either type of problem could be diminished by deficits in working memory capacity (e.g., Ashcraft & Kirk, 2001; Logie, Gilhooly, & Winn, 1994). Thus, the current research examines the potential for working memory deficits, as opposed to motivational processes, to account for the performance of stigmatized individuals under subtle stereotype threat.

*Working Memory and Stereotype Threat*
The work of Engle and his colleagues (e.g., Engle, 2001; Engle, Tuholski, Laughlin, & Conway, 1999) forms the conceptual basis for the working memory explanation of stereotype threat performance effects (see Schmader & Johns, 2003; Schmader et al., 2008). Engle and his colleagues view working memory as a subset of highly activated long-term memory units, an array of processes that produce and maintain activation of those units, and an executive attention component. The executive attention component of working memory is responsible for maintaining task goals, processing incoming information, and blocking external and internal interference. Engle and his colleagues argue that the domain-free executive attention ability “is important for predicting performance on higher order cognitive tasks (e.g., Engle, Tuholski, et al., 1999; Kane et al., 2004)” (Unsworth, Shrock, & Engle, 2004, p. 1303). Stereotype threat researchers have argued that impairments in working memory capacity accounts for the negative impact of threat on performance (e.g. Schmader & Johns, 2003). Specifically, threatened participants are hypothesized to expend cognitive resources assessing the threat, thus diminishing working memory capacity available for task performance (see Schmader et al., 2008 for a review).

Consistent with Engle’s conceptualization of the working memory system, research in the cognitive literature has demonstrated that experimentally manipulating working memory capacity impairs mathematical performance. For instance, Logie et al. (1994) required participants to complete an articulatory suppression task and a random letter generation task while also performing math computations. Using this dual-task paradigm, both random letter generation, which impairs central executive processing (e.g. Baddeley, 1996), and articulatory suppression, which reduces phonological storage
capacity (e.g., Baddeley, 2003), increased the number of arithmetic errors participants made (Logie et al., 1994). Furthermore, additional research by Ashcraft and Kirk (2001) has also shown that impairing working memory capacity by requiring participants to keep a set of words active during performance debilitates mental arithmetic.

Because simple arithmetic performance was debilitated by experimentally taxing working memory capacity (Ashcraft & Kirk, 2001; Logie et al., 1994), deficits in working memory resulting from the experience of stereotype threat should also impair performance on tests involving computations. Thus, on the GRE test used in Chapter 1, working memory deficits would be expected to impair performance on both comparison and solve type problems. On the other hand, as demonstrated in Chapter 1, motivation produces a different pattern of performance as a function of whether the prepotent solving approach is correct or not.

**Experiment 4**

Stone and McWhinnie (2008) suggest that subtle threat impacts performance via working memory deficits. Since performance of a simple addition task was debilitated by working memory impairments in Logie et al.’s (1994) research, deficits in working memory capacity resulting from the experience of subtle stereotype threat should impair performance on any type of math problem involving computations. Therefore, by examining the pattern of performance produced under subtle threat, Experiment 4 will shed light on the mechanism(s) of subtle threat performance effects.

If participants subject to subtle threat resemble the blatant threat participants from Experiments 1 through 3, then motivation may also account for the effects of subtle threat on performance. However, if subtle threat participants perform more poorly than control
on both types of problems, then one may conclude that working memory deficits are driving the effect.

*Method*

*Participants*

Seventy-two Northeastern University undergraduates participated in this experiment (36 males and 36 females) in exchange for course credit.

*Procedure*

The subtle threat manipulation was adapted from Schmader and Johns (2003) (see also, Inlzicht & Ben-Zeev, 2000). Participants in both the subtle threat and subtle no threat conditions were exposed to two other “participants” (confederates). Male and female participants in the subtle threat condition were administered a math test in an experimental session conducted by a male experimenter and with two male confederates acting as participants. Male and female participants in the subtle no threat condition completed the math test in an experimental session conducted by a female experimenter and with two female confederates acting as participants.

As participants entered the lab, either a male (subtle threat condition) or a female (no threat condition) confederate was already there filling out the consent form and working on practice problems. Upon arrival, all participants received a brief overview of the experimental session and were given a consent form to sign and practice problems to complete. Once the experimenter gave the participant the practice problems, he (subtle threat condition) or she (no threat condition) escorted the first confederate out of the room to presumably go complete the math test in an adjacent cubicle. At this time, while participants were completing the practice problems, a second confederate entered the
room and sat down next to the participant. The experimenter then re-entered the room, gave the second confederate a consent form and practice problems to fill out. Thus, participants were aware that either two males (subtle threat condition) or two females (no threat condition) were also completing the same experiment at the same time.

When the participants finished the practice problems, there were escorted to an adjacent cubicle by the experimenter. In the subtle stereotype threat condition, the experimenter explained that the goal of the study was to administer a math test to collect normative data on males and females. In the no subtle threat condition, participants were told that the purpose of the study was to administer a problem-solving exercise to collect normative data on college students. The experimenter then explained to the participants that they would have 20 minutes to work on the test and they were to try and complete as many problems as they could as accurately as they could. Participants could not use calculators and were given scratch paper. The math test was identical to that reported in Experiment 1.

After the math test, participants filled out test experience questionnaires. Two questions assessed perceptions of stereotype threat: “To what extent are there gender differences in performance on this task?” (1 = “no gender differences,” 11 = “gender differences”); and “Who do you believe performs better on this task?” (1 = “males perform better,” 6 = “males and females perform the same,” 11 = “females perform better”).

Results

Unless otherwise noted, data were analyzed in 2 (condition: subtle stereotype threat vs. subtle no stereotype threat) x 2 (participant gender: male vs. female) x 2
(problem type: comparison vs. solve) ANOVAs with condition and gender as between subjects factors and problem type as a within subjects factor. Pairwise contrasts (Kirk, 1995) were used to decompose interactions.

**Perception of Stereotype Threat**

The perceptions of threat measures were analyzed in 2 (condition: threat vs. no threat) x 2 (participant gender: male vs. female) between subjects ANOVAs. Participants in the subtle threat condition reported that gender differences existed to a greater extent ($M = 4.47, SD = 2.77$) than no threat participants ($M = 2.89, SD = 2.33$), $F(1, 68) = 6.77, p = .011, d = .63$. Threatened participants also believed that males outperform females to a greater extent ($M = 4.83, SD = 1.08$) than no threat participants ($M = 5.53, SD = 1.42$), $F(1, 68) = 5.36, p = .024, d = .56$. Thus, the subtle threat manipulation produced the perception of stereotype threat.

**Math Performance**

Performance was first analyzed across problem type to test for the overall effect of stereotype threat (see Schmader & Johns, 2003, Exp. 3). As shown in Figure 3, consistent with research in this area, this analysis produced a Condition x Participant Gender interaction, $F(1, 68) = 4.16, p = .045, d = .50$. Females subject to subtle threat performed more poorly ($M = 41.66\%, SD = 17.06\%$) than no threat females ($M = 53.11\%, SD = 18.50\%$), $F(1, 68) = 4.33, p = .041, d = .51$. The performance of males was unaffected by the subtle threat manipulation, $F < 1$. This pattern of performance is consistent with those observed in previous stereotype threat studies (e.g., Schamder & Johns, 2003, Exp. 3).
Total number attempted was also analyzed. This analysis produced a main effect for condition, $F(1, 68) = 6.57, p = .013, d = .62$. Participants in the subtle threat condition attempted more problems ($M = 17.56, SD = 5.26$) than those in the subtle no threat condition ($M = 14.72, SD = 3.92$).

*Figure 3*

![Bar graph showing total percentage correct as a function of gender and subtle threat condition](image)

*Figure 3. Experiment 4: Total percentage correct as a function of gender and subtle threat condition*
To test for the possible effects of motivation, the percentage of problems participants answered correctly (number correct / number attempted) and the number of problems they attempted were analyzed as a function of problem type.

**Problems Attempted.** To test whether the data from the current experiment replicated the pattern observed in Experiments 1-3, a contrast (Kirk, 1995) was conducted comparing the number of solve problems subtle threat participants attempted versus the number of solve problems no threat participants attempted. Consistent with the findings from Experiments 1-3, subtle threat females attempted more solve type problems ($M = 8.44, SD = 2.45$) than controls ($M = 7.11, SD = 2.02$), $F(1, 68) = 12.34$, $p < .001$, $d = .85$. As mentioned above, this contrast was made in the context of a condition main effect, $F(1, 68) = 6.57$, $p = .013$, $d = .62$.

**Percentage Correct.** Analysis of percentage correct produced a Condition x Participant Gender x Problem Type interaction, $F(1, 68) = 5.48$, $p = .022$, $d = .57$. Neither males nor females differed as a function of threat in the percentage of solve type problems answered correctly, $Fs < 1$. However, on comparison problems, subtle threat females performed more poorly than no threat females, $F(1, 68) = 14.86$, $p < .001$, $d = .93$, whereas males did not differ as a function of threat on these problems, $F < 1$. All means and standard deviations are provided in Table 6. This pattern of performance replicates that produced in Experiment 1.

Table 6. Experiment 4: GRE performance as a function of instruction, problem type, and stereotype threat condition. ST = subtle threat, NST = no subtle threat
Solution Approach. The pattern of performance produced under subtle threat closely resembles that which was produced by blatant threat in Experiments 1 through 3. If motivation, indeed, accounts for the effect of subtle threat on test performance, females’ prepotent response to use a solving approach should be potentiated by the experience of subtle threat. To test this argument, the percentage of problems on which the females’ scratch-paper showed evidence of using the solving approach was analyzed (see Experiment 3 for a full description).

This analysis produced at Condition x Participant Gender x Problem Type interaction, $F(1, 68) = 4.35, p = .041, d = .45$. As shown in Table 7, females subject to subtle threat used the solving approach significantly more often than females not subject
to threat on comparison problems, $F(1, 68) = 3.83, p = .05, d = .42$. In contrast, subtle threat did not impact females’ use of the solving approach on solve type problems, $F < 1$.

Thus, the experience of threat potentiated females’ use of the solving approach on comparison problems. On the other hand, as can be seen in Table 7, threat had no impact on how often males used the solving approach on either comparison or solve problems.

The main effects for gender, $F(1, 86) = 4.35, p = .041, d = .45$, and problem type, $F(1, 86) = 42.81, p < .001, d = 1.41$, should be analyzed in the context of the 3-way interaction.

Table 7. Experiment 4: Percentage of problems on which the solving approach was used by females as a function of problem type and stereotype threat condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Solving Approach (%)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Comparison Problems</td>
<td>Solve Problems</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$SD$</td>
<td></td>
</tr>
<tr>
<td>NST Males</td>
<td>53.61&lt;sub&gt;a&lt;/sub&gt;</td>
<td>19.06</td>
<td>68.14&lt;sub&gt;a&lt;/sub&gt;</td>
<td>17.44</td>
<td></td>
</tr>
<tr>
<td>ST Males</td>
<td>58.51&lt;sub&gt;a,b&lt;/sub&gt;</td>
<td>16.72</td>
<td>76.59&lt;sub&gt;a,b&lt;/sub&gt;</td>
<td>11.03</td>
<td></td>
</tr>
<tr>
<td>NST Females</td>
<td>57.31&lt;sub&gt;a&lt;/sub&gt;</td>
<td>14.66</td>
<td>82.08&lt;sub&gt;b&lt;/sub&gt;</td>
<td>14.53</td>
<td></td>
</tr>
<tr>
<td>ST Females</td>
<td>67.03&lt;sub&gt;b&lt;/sub&gt;</td>
<td>15.14</td>
<td>76.64&lt;sub&gt;a,b&lt;/sub&gt;</td>
<td>19.42</td>
<td></td>
</tr>
</tbody>
</table>

Note: Means not sharing a subscript differ at $p < .05$
**Mediation**

To test whether potentiation of the solving approach directly led to debilitated comparison problem performance, we conducted a mediation analysis following the procedures suggested by Kenny, Kashy, and Bolger (1998) on the number of comparison problems solved incorrectly. As shown in Figure 4, the use of the prepotent, solving approach, mediated the debilitating effect of stereotype threat on comparison problem performance, Sobel $Z = 2.01, p = .038$. Thus, under subtle threat, females’ use of the solving approach mediated the effect of debilitation on comparison problems, which replicates the blatant threat findings from Experiment 3.

**Figure 4**
**Figure Caption:** Number of times the prepotent solving approach was used on comparison trials as a mediator of number of comparison problems answered incorrectly. Coefficients in parentheses indicate zero-order correlations. Coefficients not in parentheses represent parameter estimates for a recursive path model including both predictors. Double asterisks (**) indicate parameter estimates or correlations that differ from zero at $p < .01$. Subtle stereotype threat condition is dummy coded (subtle stereotype threat = 1, no subtle stereotype threat = 0)

**Discussion**

In Experiment 4, stereotyped threat was manipulated using a subtle threat manipulation, whereas Experiments 1-3 employed a blatant threat manipulation. However, the effect of subtle threat produced the same pattern of performance as was observed in Chapter 1. Overall, females subject to subtle threat performed more poorly on a GRE-Q test than no threat females and males in either threat condition. When performance was broken down by problem type, females subject to subtle stereotype threat attempted more solve problems than no threat females, but solved the same percentage of these problems correctly. In contrast, on comparison problems, females under threat attempted the same number of problems as control females, but solved a smaller percentage of these problems correctly. This pattern of performance is consistent with a mere effort interpretation and replicates the result observed in Experiment 1 of this research.

Unlike Experiments 1-3, threatened participants attempted more comparison type problems in the current research. However, because stereotype threat females exhibited
impaired performance on the comparison problems (lower percentage) the number of comparison problems they attempted is not informative. The prepotent response to use a solving approach is incorrect for comparison problems but participants do not recognize that this response is incorrect. Thus, threatened participants working on comparison problems may persist in trying to use the incorrect approach on the problems, which could ultimately lead them to attempt fewer of these problems as they struggle. However, it is also the case that, unlike solve problems, comparison problems do not require a specific numeric answer. The participants need only indicate whether the expression in Column A is larger than the one in Column B, B is larger than A, they are the same, or there is not enough information to tell. Thus, threatened participants could quickly choose some answer (probably not correct) based on the incorrect solving approach, allowing them to attempt more of these problems than controls. As a result, the number of comparison problems attempted is not informative in light of the deficit in percentage correct as a function of threat.

Analysis of participants’ solution approach suggests that the debilitating effect of subtle threat on performance was the result of threatened females using the incorrect solving approach more often on comparison problems than controls. This conclusion was supported by a mediation analysis showing that threatened females’ reliance on the solving approach was directly responsible for their debilitated performance on the comparison problems. In sum, when the prepotent, solving approach was correct, subtle threat participants performed better than controls, but when this response was incorrect, their performance was debilitated relative to controls.
Previous stereotype threat research that has argued for the mediating role of working memory deficits in stereotype threat performance effects has used subtle stereotype threat manipulations (e.g., Schmader & Johns, 2003). Consistent with that view, Stone and McWhinnie (2008) argued that motivational effects may be produced by blatant threat manipulations, like that used in Experiments 1-3, whereas subtle manipulations of threat, like that used in Experiment 4, may work through a different mechanism: working memory interference. However, inconsistent with a working memory interference interpretation, in the current research, we found that subtle threat debilitated stigmatized participants’ performance on only those problems for which the prepotent response was incorrect. Instead, this data suggests that females were motivated by the subtle threat manipulation. Moreover, participants’ use of the incorrect solution approach mediated the link between threat and poor performance.

Experiment 4 demonstrates that motivation may play a role in producing subtle threat performance effects. However, stereotype threat situations are not the only way stereotypes can impact behavior. For instance, simply activating stereotypes can lead individuals to behave more stereotypically, an effect dubbed stereotype priming (Bargh et al., 1996). Researchers have suggested that the debilitation produced in stereotype threat and priming paradigms are produced by the same process: stereotype activation leads directly to stereotype-related behaviors (e.g., Ambady, Paik, Steele, Owen-Smith, & Mitchell, 2004; Dijksterhuis & Bargh, 2001; Wheeler et al., 2001; Wheeler & Petty, 2000). Recently, however, Marx and Stapel (2006) have suggested that despite the fact that stereotype threat and priming have the same effect on performance, the processes
that produce these effects are different. Specifically, they argue that threat-based concerns distinguish stereotype threat effects from priming effects.

However, it may also be possible to distinguish threat from priming by examining performance. Because primes impact behavior passively through ideomotor processes (e.g. Bargh et al., 1996; Wheeler, Jarvis & Petty, 2001), it is unlikely that primes motivate individuals to want to perform well. In contrast, the data presented in Experiments 1-4 of this research suggests that motivation plays a role in producing the effect of stereotype threat on performance. Thus, the goal of the following Chapter is to extend the theoretical framework provided by Marx and Stapel (2006) by testing the possibility that stereotype threat and priming differ not only in the threat-based concerns produced by the former, but not by the latter, but also in the pattern of performance that results.
Chapter 4: Differentiating Stereotype Threat from Stereotype Priming

Experiments 1-4 of this research explored the role motivation of motivation in accounting for the stereotype threat performance decrements. Like threat, priming negative stereotypes has also been shown to debilitate performance. For example, Wheeler et al. (2001) primed racial stereotypes by requiring participants to write a story from the first person perspective about an individual named either “Erik” or “Tyrone.” Participants who wrote the story from the perspective of Tyrone (a stereotypically African-American name) performed more poorly on a subsequent test of intellectual ability, which is consistent with stereotypes regarding African-Americans (Wheeler et al., 2001).

In fact, some researchers have suggested that the debilitation produced in these two paradigms is produced by the same process: stereotype activation leads directly to stereotype-related behaviors (e.g., Ambady et al., 2004; Dijksterhuis & Bargh, 2001; Wheeler et al., 2001; Wheeler & Petty, 2000). Recently, however, Marx and Stapel (2006) have suggested that despite the fact that stereotype threat and priming have the same effect on performance, the processes that produce these effects are different. Specifically, they propose that stereotype threat effects stem from the fact that the target of the stereotype knows about the stereotype and is a member of the stigmatized group, whereas priming effects simply require knowledge of the stereotype. As they write: “in stereotype threat situations, targets (but not nontargets) are affected because they know the group stereotype (“women are bad at math”) and because they are members of the group that is targeted by the stereotype (“I am a woman”)” (p. 244). It is this combination of
“knowing and being” that gives rise to threat-based concerns, which distinguishes stereotype threat effects from priming effects.

The goal of the research presented in Chapter 4 is to extend the theoretical framework provided by Marx and Stapel (2006). Specifically, this research tests the possibility that the experience of stereotype threat and stereotype priming differ not only in the threat-based concerns produced by threat, but not by priming, but also in the patterns of performance that result from threat and priming. To test the possibility that threat and priming produce distinct patterns of performance, we rely on Jamieson and Harkins’s (2007) motivation-based, mere effort account of stereotype threat performance effects. The current research compares the patterns of performance produced by a manipulation of priming to those resulting from a manipulation of threat. Participants completed the same math test as in Experiments 1-4: a GRE test comprised of two types of problems: comparison and solve problems, which differ in the solution approach that tends to be most efficient.

The experience of threat is characterized by concerns about confirming negative stereotypes (Marx & Stapel, 2006). Thus, threatened individuals are motivated to disconfirm stereotypes because if performing well in the stereotyped domain were not important to stigmatized individuals, they would have no reason to be concerned about performing poorly. However, there is no reason to expect priming gender stereotypes to impact motivation and produce a similar pattern of performance as the experience of threat. Research on behavioral priming indicates that primes affect behavior by activating information associated with whatever construct is primed (e.g., Bargh et al., 1996). In the case of gender, female gender primes activate stereotypic beliefs that females are poor at
math. This stereotype does not incorporate the notion that females are motivated by their concern about confirming the stereotype, which facilitates performance on solve problems, but debilitates it on comparison problems. Instead, the activated stereotypic information passively guides behavior (e.g., Bargh et al., 1996), which should lead to poor performance on both problem types.

To test these predictions, we replicated the stereotype threat effect produced in Experiment 1 and, in parallel, conducted a priming experiment in which we activated the female gender construct. We conducted the threat and priming experiments separately, rather than crossing the two manipulations because we are interested in looking at the pattern of effects produced by each of these manipulations, not at looking at their interactive effects. Of course, these types of effects could be of interest, but they were not the focus of the current research.

Experiment 5

Stereotype threat was manipulated explicitly, such that participants were told that they would be completing a math test that either had (threat) or had not (no threat) been shown to produce gender differences. This manipulation has been successfully used in previous stereotype threat research (e.g., Brown & Pinel, 2003; Jamieson & Harkins, 2007, in press; Keller, 2002; Keller & Dauenheimer, 2003; O’Brien & Crandall, 2003; Spencer et al., 1999).

Because mere effort predicts that stereotype threat will increase the likelihood that females will rely on their prepotent tendency to use the solving approach, threatened females are predicted to perform more poorly on comparison problems and better on solve problems than females not subject to stereotype threat. We also expected that
overall performance on the test would be worse for threatened than non-threatened females because the debilitation of performance on the comparison problems will be greater than the facilitation of performance on the solve problems.

Method

Participants

111 Northeastern University undergraduates participated in this experiment, with 41 females participating in the stereotype threat experiment, and 35 males and 35 females participating in the priming experiment.

Procedure

The priming manipulation was adapted from Wheeler et al. (2001). Participants were informed that they would be completing math problems, but prior to the math test the experimenter explained that he was helping another lab pilot test an experiment “examining the role of hemispheric dominance on creativity,” and asked if the participants would also complete this experiment. All but one participant, who was excluded from the analysis, agreed. Participants were then told that they would write a creative essay about an assigned topic with either their dominant or non-dominant hand. In fact, all participants wrote with their dominant hands.

Participants in the female prime condition were asked to write about a day in life of a female Northeastern University student named “Ashley” for five minutes. Participants assigned to the gender-neutral control condition wrote about a day in the life of Northeastern University, as an institution. That is, participants wrote about general events that went on at the university on a daily basis. This control ensured that gender-specific information was not activated.
After the experimenter collected the essays, he gave the participants two practice math problems and told them that they would have 20 minutes to work on a 30-item test without the use of calculators. The test included 15 comparison and 15 solve problems. Based on population norms, the mean overall accuracy averaged 50% for each problem type. The test was presented in a paper-and-pencil format with scratch paper provided.

All participants in the stereotype threat experiment were told that they were about to complete a test of mathematical ability that either had (stereotype threat) or had not (no stereotype threat) been shown to produce gender differences (e.g., Brown & Pinel, 2003; O’Brien & Crandall, 2003; Spencer, Steele, & Quinn, 1999). Males were not run in the threat conditions because previous research has found that males’ quantitative GRE performance is not debilitated by manipulations of stereotype threat (e.g., Schmader & Johns, 2003; Spencer et al., 1999; Experiment 1, present research).

After the math test, participants filled out test experience questionnaires. Two questions allowed for the evaluation of participants’ perceptions of stereotype threat: “To what extent are there gender differences in performance on this task?” (1 = “no gender differences,” 11 = “gender differences”); and “Who do you believe performs better on this task?” (1 = “males perform better,” 6 = “males and females perform the same,” 11 = “females perform better”). We expected the stereotype threat manipulation to impact participants’ response to these items, whereas the priming manipulation was not expected to have any effect on reports of threat.

Results

**Stereotype Threat Effects**
Unless otherwise noted, data were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) x 2 (problem type: comparison vs. solve) ANOVAs with condition as a between subjects factor and problem type as a within subjects factor. Pairwise contrasts (Kirk, 1995) were used to decompose interactions.

**Perception of Stereotype Threat**

The stereotype threat measures were analyzed in independent samples t-tests (stereotype threat vs. no stereotype threat). Females in the threat condition reported that gender differences existed to a greater extent ($M = 5.57, SD = 2.77$) than no threat participants ($M = 3.55, SD = 2.70$), $t(39) = 2.36, p = .023, d = .76$. Threatened females also believed that males outperform females to a greater extent ($M = 4.62, SD = 1.46$) than no threat participants ($M = 5.85, SD = 1.53$), $t(39) = -2.63, p = .012, d = .84$.

**Math Performance**

First, performance was analyzed across problem type to test for the overall effect of stereotype threat on performance (see Schmader & Johns, 2003, Exp. 3). Replicating previous research, stereotype threat participants performed more poorly ($M = 46.51\%, SD = 14.31\%$) than no threat participants ($M = 56.57\%, SD = 14.23\%$), $F(1, 39) = 4.99, p = .03, d = .72$.

To test for the effects suggested by the mere effort hypothesis we examined the number of problems attempted and percentage correct (# correct / # attempted)

**Problems Attempted.** Replicating the threat effects from Experiments 1-3, analysis of number attempted produced a Condition x Problem Type interaction, $F(1,39) = 3.97, p = .05, d = .64$. Stereotype threat participants attempted more solve problems ($M = 8.95, SD = 2.10$) than controls ($M = 7.90, SD = 1.17$), $F(1,39) = 4.09, p < .05, d = .64$, whereas
participants in the stereotype threat condition attempted the same number of comparison problems ($M = 8.33, SD = 1.83$) as no stereotype threat participants ($M = 8.75, SD = 1.45$), $p > .20$.

**Percentage Correct.** Analysis of percentage correct also produced a Condition x Problem Type interaction, $F(1,39) = 7.02, p = .01, d = .85$. As shown in Table 8, threat participants did not differ from no threat participants in the percentage of solve problems solved, but solved a smaller percentage of the comparison problems than no threat participants, $F(1,39) = 16.00, p < .001, d = 1.28$. This pattern of findings replicates that observed in Experiments 1-4.

**Table 8.** Experiment 5: GRE performance (% correct) as a function of problem type and stereotype threat condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>% Correct Comparison</th>
<th>% Correct Solve</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$ ($SD$)</td>
<td>$M$ ($SD$)</td>
</tr>
<tr>
<td>Stereotype Threat</td>
<td>31.16 ($18.01$)</td>
<td>60.17 ($15.86$)</td>
</tr>
<tr>
<td>No Stereotype Threat</td>
<td>53.71 ($15.87$)</td>
<td>61.60 ($23.89$)</td>
</tr>
</tbody>
</table>

*Note: Means not sharing a subscript are significantly different, $p < .05$*
**Solution Approach.** We argue that the prepotent response to use a solving approach is potentiated by stereotype threat. To test this argument, following the procedures outlined in Experiment 3 of the current research, an experimenter blind to condition computed the percentage of problems on which the participants’ scratch-paper showed evidence of using the solving approach. This analysis produced a Condition x Problem Type interaction, $F(1,39) = 14.25, p = .001, d = 1.21$. As shown in Table 9, threat participants used the solving approach significantly more often than no threat participants on comparison problems, $F(1,39) = 37.42, p < .01, d = 1.96$. In contrast, stereotype threat did not impact participants’ use of the solving approach on solve type problems.

**Mediation**

To test whether potentiation of the solving approach directly led to debilitated comparison problem performance, we conducted a mediation analysis following the procedures suggested by Kenny et al. (1998) on the number of comparison problems solved incorrectly. As shown in Figure 5, the use of the prepotent, solving approach, mediated the debilitating effect of stereotype threat on comparison problem performance, Sobel $Z = 2.58, p < .01$.

*Figure 5*
Figure Caption. Number of times the prepotent solving approach was used on comparison trials as a mediator of number of comparison problems answered incorrectly. Coefficients in parentheses indicate zero-order correlations. Coefficients not in parentheses represent parameter estimates for a recursive path model including both predictors. Asterisks (*) indicate parameter estimates or correlations that differ from zero at $p < .05$, double asterisks (**) indicate parameter estimates or correlations that differ from zero at $p < .01$.

Stereotype Priming Effects

Unless otherwise noted, priming effects were analyzed in 2 (prime: control vs. female) x 2 (gender: male vs. female) x 2 (problem type: solve vs. comparison) ANOVAs.
with both prime and gender as between subjects factors and problem type as a within subjects factor.

Perception of Stereotype Threat

Analyses of the stereotype threat items were analyzed in 2 (prime: control vs. female) x 2 (gender: male vs. female) ANOVAs. As expected, female prime participants did not differ from controls in their beliefs about whether gender differences existed or in their judgment of whether males or females are better at the task.

Math Performance

Consistent with Marx and Stapel’s (2006) finding, participants in the female prime condition performed more poorly overall ($M = 44.15\%$; $SD = 19.69\%$) than controls ($M = 57.37\%$; $SD = 17.14\%$), $F(1,65) = 8.63$, $p = .005$, $d = .73$.

As was the case for stereotype threat, we analyzed the number of problems attempted and the percentage of problems answered correctly to test for the effects of priming on comparison and solve problems.

Problems Attempted. Participants attempted more comparison problems ($M = 9.14$, $SD = 2.24$) than solve problems ($M = 8.70$, $SD = 1.89$), $F(1,65) = 5.24$, $p < .03$, $d = .57$. No other effects were reliable.

Percentage Correct. Participants in the control condition outperformed female prime participants on both types of problems ($M_{control} = 57.78\%$, $SD_{control} = 19.94\%$; $M_{prime} = 42.71\%$, $SD_{prime} = 23.61\%$), $F(1,65) = 9.11$, $p < .01$, $d = .75$. We also observed a main effect for problem type. Participants solved a greater percentage of solve problems ($M = 56.82\%$, $SD = 24.28\%$) than comparison problems ($M = 45.37\%$, $SD = 21.15\%$), $F(1,65) = 16.01$, $p < .001$, $d = .99$. Thus, prime condition did not interact with
problem type, as one would expect if priming produced the same effect on performance as stereotype threat.

Solution Approach. This analysis produced a Prime x Problem Type interaction, $F(1, 64) = 11.98, p < .001, d = .87$. As shown in Table 9, participants given the female prime used the solving approach on solve problems significantly less than participants in the control group, $F(1, 64) = 18.34, p < .001, d = 1.07$, whereas the prime conditions did not differ in the use of the solve approach on comparison problems, $p > .20$.

**Table 9. Percentage of problems on which participants used the prepotent, solving approach as a function of stereotype threat condition, priming condition, and problem type.**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Solving Approach (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M (SD)</td>
<td></td>
</tr>
<tr>
<td>Comparison Problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>StereoType Threat</td>
<td>69.04 (18.28)</td>
<td></td>
<td>76.00 (10.86)</td>
</tr>
<tr>
<td>No Stereotype Threat</td>
<td>49.05 (14.75)</td>
<td></td>
<td>73.46 (13.37)</td>
</tr>
<tr>
<td>Solve Problems</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female Prime</td>
<td>46.87 (19.37)</td>
<td></td>
<td>57.68 (15.88)</td>
</tr>
<tr>
<td>Institution Control</td>
<td>45.05 (13.55)</td>
<td></td>
<td>70.41 (15.13)</td>
</tr>
</tbody>
</table>

*Note:* Threat and Prime were analyzed in separate analyses. Means not sharing a subscript are significantly different, $p < .05$. 

This analysis suggests that prime participants performed more poorly than control participants on solve problems because they did not use the solving approach on these problems to the same extent as controls. Consistent with this argument, when prime participants showed evidence of using the solving approach, they solved the same percentage of solve problems ($M = 56.22\%, SD = 26.37\%$) as control participants ($M = 63.31\%, SD = 21.04\%), $p > .20$. When prime participants did not show evidence of using the solving approach, their performance dropped significantly on solve problems ($M = 37.13\%, SD = 27.70\%$), $F (1,61) = 14.81, p < .001, d = .99$, whereas the performance of control participants was unchanged ($M = 60.29\%, SD = 34.21\%), p > .30$, interaction $F (1,61) = 5.06, p < .03, d = .58$.

We conducted the same analysis on the solution rates for comparison problems. When participants showed evidence of using the solving approach on comparison problems, their performance was debilitated to the same extent in each condition ($M_{\text{prime}} = 35.11\%, SD_{\text{prime}} = 29.86\%: M_{\text{control}} = 34.84\%, SD_{\text{control}} = 27.52\%), p > .20$. However, when participants did not show evidence of using the solving approach, controls outperformed ($M = 61.77\%, SD = 19.68\%$) prime participants ($M = 47.56\%, SD = 30.57\%), $F (1,63) = 6.10, p < .05, d = .62$, interaction $F (1, 63) = 3.22, p < .08, d = .45$.

Discussion

The stereotype threat findings replicate those from Experiments 1-4. Overall, threatened participants performed more poorly than controls, but threat participants were as accurate and solved more solve-type problems than controls, whereas on comparison problems they performed substantially more poorly. Moreover, a mediation analysis
demonstrates that threatened participants’ use of the prepotent solving approach mediated the effect of threat on comparison problem performance.

The priming findings replicated Marx and Stapel’s (2006) priming finding: Overall, both male and female participants given the female prime performed more poorly than controls. However, unlike the pattern observed under threat in Experiment 1, female prime participants (both males and females) performed worse than institutional prime participants on both problem types.

An analysis of participants’ solution approaches and solution rates suggests a possible explanation for this difference in performance. Female prime participants performed as well as controls on the subset of solve problems on which both groups showed evidence of using the solving approach, but prime participants used this approach less often than controls. On the subset of problems on which neither group showed evidence of using the solving approach, the performance of female prime participants dropped significantly, whereas the performance of controls was unchanged. This pattern of findings suggests that the control participants were trying to solve these problems in their heads, whereas the prime participants were simply guessing. That is, when participants did not write anything down, they could be trying to solve the problems in their heads. When both prime and control participants showed written evidence of using the solving approach, the two groups did not differ in their solution rates. As a result, if prime and control participants were using the solving approach in their heads to the same extent on the problems on which there was no written work, there is no reason to believe that they would differ in their solution rates on these problems. However, the results
show that controls outperformed the prime participants on these problems, suggesting that the female prime participants were withdrawing effort and guessing.

On comparison problems, the results indicate that participants under stereotype threat performed more poorly than no threat participants, because they used the inappropriate solving approach more often. In contrast, female prime and control participants used the solving approach on approximately the same percentage of comparison trials and performed equally poorly on these trials, reflecting the debilitation produced by using the incorrect approach. On the remainder of the problems, female prime and control participants had an equal opportunity to use the appropriate approach (logic and estimation), but female prime participants appear to have done so less than controls as the latter participants outperformed the former participants. Once again, these data suggest that female prime participants withdrew effort. Of course, withdrawing effort and guessing is consistent with the stereotype that females are poor at math and, thus, should not even try to perform well on math tests (e.g., Steele & Ambady, 2006).

Thus, priming gender stereotypes likely does not impair one’s ability to perform mathematical computations, but instead the solution rate analysis suggests that activating gender stereotypes lead individuals to “give up” and guess on a greater percentage of problems than controls.

This research suggests that we should exercise caution when generalizing from findings produced in stereotype activation paradigms to the actual behavior of stereotyped individuals. For example, Smith, Jostmann, Galinsky, and van Dijk (2008) reported finding that low power impaired performance on executive-function tasks. They generalized this finding to real world settings, arguing that empowering employees could
reduce costly organizational errors. However, that research used manipulations aimed at activating stereotypes concerning power (e.g., scrambled-sentences priming task). Thus, we may have learned something about the content of stereotypes associated with the powerful and powerless, but not necessarily about the actual behavior of people who occupy these positions.

We must also place this research in the broader context provided by recent work on priming. Wheeler, DeMarree, and Petty (2007) proposed the active-self account as an organizing framework for priming effects, which incorporates findings that follow from the ideomotor account (Dijksterhuis & Bargh, 2001) as well as other prime-to-behavior effects. For instance, Wheeler et al. (2007) argue that perspective taking increases the likelihood of assimilation of primed concepts into one’s self-concept. In our research, prime participants took the perspective of a female college student. Thus, males may have assimilated information associated with female gender roles into their active-self, which guided their behavior, leading them to behave as (they believe) females would when taking a math test. Thus, consistent with Wheeler et al.’s (2007) model, assimilation processes can account for the impact of stereotype primes on non-stereotyped individuals.
Chapter 5: General Discussion

Realizing the goal of a truly egalitarian society requires minimizing differences between groups that confer advantages on one group at the expense of another. Although great strides towards this goal have been made in recent years, there is still much work to be done. Negative group stereotypes remain pervasive in human societies and can hinder stereotyped individuals’ ability to succeed in stereotyped domains. The current research focused on the impact of stereotype threat, or the concern about confirming negative group stereotypes (Steele & Aronson, 1995), on stigmatized individuals’ performance. A multitude of research has demonstrated that stereotype threat debilitates performance across a wide array of stereotypes in a variety of domains.

Recently, the focus of stereotype threat research has shifted to the study of what mechanism(s) mediate stereotype threat performance effects. Consistent with Steele et al.’s (2002) conceptualization of stereotype threat, researchers have found evidence for the role of cognition (Schmader et al., 2008), affect (Bosson et al., 2004), and motivation (Jamieson & Harkins, 2007) in producing stereotype threat performance effects. The current research focuses on the role of motivation.

Previous research on the role of motivation in stereotype threat examined the performance of individuals on an inhibition task, the antisaccade task (Jamieson & Harkins, 2007). However, most previous stereotype threat research on the gender/math stereotype has examined performance on quantitative problems from standardized tests.
such as the GRE and SAT, not on inhibition tasks. In the current research, we extended the mere effort analysis to performance on GRE problems.

Previous research suggests that when faced with quantitative problems, participants’ prepotent tendency is to take a solving approach. That is, participants attempt to employ previously learned equations, rules, algorithms to solve the problems presented to them. Gallagher and De Lisi (1994) found that their participants used this approach on 63% of the problems as opposed to some other method; Gallagher et al. (2000, Exp. 2) found that 56% did so; and Quinn and Spencer (2001; Exp. 2) found that 57% did so. These values compare favorably to our own findings from Experiment 3, which showed that when not given instruction, participants used the solving approach on 65% of the problems.

The GRE quantitative test used in the experiments presented here consisted of two types of problems: comparison and solve, which differ in the solution approach that tends to be most efficient. On comparison problems, the most efficient approach to problem solution tends to be simplifying the terms in each column and/or using logic, estimation, and/or intuition to find the correct answer. On solve problems, the most efficient approach tends to be representing the problem in an equation and solving it. Given these two types of problems, the predictions made by the mere effort account are straightforward. On solve problems, the prepotent response, solving, is correct and threat should facilitate performance, whereas on comparison problems, the prepotent response is incorrect, and performance should be debilitated.

In Experiment 1 we found that female participants subject to stereotype threat performed more poorly overall than participants in the other conditions, which did not
differ among themselves. However, when we looked at performance as a function of problem type, consistent with mere effort predictions, females subject to stereotype threat performed more poorly on comparison problems, but better on the solve problems than control participants.

In Experiment 2, we varied the difficulty of solve and comparison problems to determine whether the pattern of performance observed in Experiment 1 would remain consistent across difficulty level. Research by O’Brien and Crandall (2003) shows that stereotype threat facilitates performance on simple math problems but debilitates performance on difficult problems. This interpretation would suggest that if the math test used in Experiment 1 were made up of easier problems, then stereotype threat participants should perform better than controls. However, mere effort argues that it is problem type, not difficulty, which accounts for the pattern of performance observed in Experiment 1. That is, the pattern of performance should not differ as a function of difficulty level, but instead should depend on whether the prepotent response is correct or not. Consistent with this view, if the prepotent response was correct (solve problems), females subject to threat outperformed controls, but if this response was incorrect (comparison problems), they performed more poorly than controls at each level of difficulty.

Because the mere effort hypothesis argues that stereotype threat participants are more motivated to perform well, they should be motivated to correct a prepotent response when that response is incorrect, they recognize that it is incorrect, and they have the opportunity to correct it (Jamieson & Harkins 2007, Exp. 3). Experiments 1 and 2 of the current research demonstrate that threatened females perform more poorly on comparison
problems than females who are not under threat. In this case, the participants may not know that their approach is inappropriate, and even if they do recognize this, they do not know the appropriate approach to take.

In Experiment 3, we tested whether giving instructions that described the correct solution approach for each type of problem could improve threatened females’ performance. Consistent with this notion, when instructions were provided, females subject to threat did not differ from controls in the percentage of comparison problems that they answered correctly. Analysis of the participants’ scratch-paper showed that participants subject to threat used the solving approach on comparison problems less often when given instructions than when instructions were not provided. However, when not given instructions, females subject to threat applied the incorrect, but prepotent, solving approach on more comparison problems, and performed more poorly on these problems than no threat females.

Stone and McWhinnie (2008) have suggested that although blatant manipulations of threat (like the one used in Experiments 1-3) may affect performance though motivational processes, subtle manipulations of threat may work by producing working memory deficits. Such a deficit would be expected to debilitate performance on both solve and comparison problems (Logie et al., 1994). In Experiment 4, we used a subtle rather than a blatant manipulation of threat, but produced the same pattern of performance as was observed in Experiments 1-3. Overall, females subject to subtle threat performed more poorly on the GRE-Q test than no threat females and males in either threat condition. When performance was broken down by problem type females subject to subtle stereotype threat attempted more solve problems than no threat females,
but solved the same percentage of these problems correctly. In contrast, on comparison problems, females under threat attempted the same number of problems as control females, but solved a smaller percentage of these problems correctly.

Finally, Experiment 5 sought to differentiate stereotype threat and stereotype priming effects. The threat findings replicated those produced in the first four experiments of this research: Facilitation on solve type problems, debilitation on comparison type problems. This research also replicated Marx and Stapel’s (2006) priming finding: Overall, participants given the female prime performed more poorly than controls. However, priming and stereotype threat produced the same overall effect in different ways. Threat participants performed better on solve problems but worse on comparison problems than controls, whereas the female prime participants performed worse than controls on both problem types. Moreover, analysis of participants’ solution approaches indicates that debilitation under threat resulted from participants trying too hard, whereas priming effects resulted from participants trying too little.

Across the five experiments presented here, stereotype threat debilitated performance (lower % correct) on comparison problems (weighted mean $r = -0.48$, combined $Z = 8.79$, $p < .001$), but the experience of threat also led females to attempt more solve problems (weighted mean $r = 0.37$, combined $Z = 7.01$, $p < .001$). Without a deficit in accuracy on solve problems, the greater number of solve problems attempted by threat participants translated into a greater number correct for threat versus no threat participants (weighted mean $r = 0.27$, combined $Z = 4.92$, $p < .001$). This greater number correct shows that threat facilitates performance on math problems when participants’ prepotent solving approach is correct. Moreover, this finding is consistent with previous
threat research which found that the experience of stereotype threat facilitated performance on multiplication problems (O’Brien & Crandall, 2003) on which the prepotent solving approach is also correct.

Taken together, the experiments presented here are consistent with a mere effort interpretation (Harkins, 2006; Jamieson & Harkins, 2007) of the effects of stereotype threat on females’ performance on math problems. However, it could be argued that other findings in the threat literature are inconsistent with this account. For example, if females are motivated by stereotype threat, one could expect to find that they spend more time working on the problems, and/or attempt more problems than females not subject to threat. Yet no consistent effects have been reported on these measures. For example, Steele and Aronson (1995, Exp. 2) found that threat participants spent more time per problem than controls, whereas Spencer et al. (1999, Exp. 1) found no differences on this measure. Steele and Aronson (1995, Exp. 4) found that threat participants completed fewer problems than control participants, whereas Keller (2003) found no difference on this measure. However, there is no evidence to suggest that spending more time per problem or attempting more problems represents the prepotent response on this task. There is evidence that the participants’ prepotent response when faced with these problems is to use the solving approach, rather than simplifying terms or using logic, estimation, or intuition to find the correct answer (e.g., Gallagher, et al., 2000). And the mere effort account’s predictions concerning the effect of threat on the prepotent response were supported.

A motivational explanation may also seem to be at odds with research that has found that stereotype threat affects performance, but not self-reported effort (e.g., Brown
& Pinel, 2003; Keller, 2002; Steele & Aronson, 1995). However, it is not at all unusual to find a lack of correspondence between measures of self-reported effort and measures of performance. For example, in a meta-analysis of social loafing research, Karau and Williams (1993) found that the social loafing effect was robust, but the average effect size for self-reported effort in these experiments was not significantly different from zero. Apparently participants are unwilling or unable to acknowledge the fact that they loaf. In any event, the focus of the mere effort account is on task performance, not on self-reported effort, and we make no claims concerning the latter.

A motivational explanation may seem inconsistent with research that has shown that participants subject to threat take advantage of explanations that allow them to deflect responsibility for their performance (e.g., Ben-Zeev et al., 2005; Johns et al., 2005; Keller, 2002; Steele & Aronson, 1995; Stone, 2002). However, the mere effort account does not argue that stigmatized individuals are motivated to seek out situations to demonstrate their proficiency in stereotyped domains. It characterizes their behavior in situations that do not provide them with the opportunity to avoid judgment in the context of the stereotype. Consistent with this view, Harkins and his colleagues have demonstrated that although participants are more motivated to perform well when subject to experimenter evaluation than when they are not (e.g., Harkins & Szymanski, 1988; McFall et al., 2009; White, Kjelgaard & Harkins, 1995), when given a choice, they avoid subjecting themselves to the scrutiny of the experimenter (Szymanski & Harkins, 1993).

The fact that the prepotent response tends to be correct on one type of GRE-Q question, solve problems, but tends to be incorrect on another type, comparison problems is key to the mere effort account. However, researchers have found that a manipulation of
stereotype threat impairs females’ math performance on a test comprised entirely of solve-format GRE problems (e.g. Brown & Pinel, 2003; Johns et al., 2005; Schmader & Johns, 2003).

One could argue that the debilitation stems from fact that these researchers used a subtle manipulation of threat, which could produce a working memory deficit, whereas the blatant manipulation used in the current work impacted motivational processes (cf. Stone & McWhinnie, 2008). However, Experiment 4 suggests that the same pattern of results is produced by subtle and blatant manipulations of threat. In addition, Brown and Pinel (2003) used a blatant threat manipulation and found debilitation on a test composed of only solve format problems, leaving this inconsistency unexplained.

However, a closer look at solve-type problems suggests a possible resolution. In an analysis of 540 solve-format problems taken from actual GRE-Q tests, the prepotent solving approach was the most efficient solution approach on 75.56% of the problems. The unconventional approach (i.e., a “comparison” approach using logic, estimation, and/or intuition) was most efficient for the other 24.44%, though. Furthermore, the percentage of solve-format problems for which an unconventional approach is most efficient increases with difficulty level. For instance, for the subset of solve format problems with an average solution rate of less than or equal to 33%, the solving approach was the most efficient solution approach on 57% of the problems. As a result, researchers could have included “comparison” problems (unconventional problems that are most efficiently solved using logic, estimation, or intuition) that are presented in a “solve” format (word problems) in such numbers that the “solve”-only test ends up being functionally equivalent to a mixed-format test.
Consistent with this possibility, half of Brown and Pinel’s (2003) problems were most efficiently solved by taking a conventional approach (i.e., were solve-type problems) whereas the other half were most efficiently solved by taking an unconventional approach (i.e., were comparison-type problems). Of the problems used by Schmader and her colleagues, sixty percent of the items were most efficiently solved taking a conventional approach, whereas forty percent were most efficiently solved taking an unconventional approach. Thus, this analysis suggests the possibility that a motivational explanation may also apply to tests composed only of “solve” problems.

To test this hypothesis, we requested and were provided the data from Schmader and Johns (2003, Exp. 3) and Johns, Schmader, and Martens (2005). However, we used only the non-teaching conditions from Johns et al. (2005) because the teaching manipulation reduced the negative effect of stereotype threat on performance. We also replicated Schmader and Johns’ (2003, Exp. 3) study with a sample of 36 Northeastern undergraduates (all female).

We then used meta-analytic techniques to examine the effect of threat on the unconventional (prepotent response is incorrect) and conventional (prepotent response is correct) problems. First, across all three studies, we replicated that traditional stereotype threat finding: threat participants performed more poorly (lower % correct) than controls, weighted mean $r = -.32$, combined $Z = 3.44, p = .0006$). When the data were analyzed as a function of problem type (unconventional vs. conventional), the pattern was consistent with that observed in the current research. Across the three experiments, on unconventional (i.e. comparison) problems, participants subject to threat exhibited lower accuracy rates than controls, weighted mean $r = -.39$, combined $Z = 4.21, p < .0001$. In
addition, threat participants attempted more conventional problems than controls participants, weighted mean $r = .21$, combined $Z = 2.61, p = .009$, but did not differ in accuracy, weighted mean $r = -.12$, combined $Z = 1.36, p = .17$. Furthermore, consistent with the notion that threat differentially impacts performance as a function of problem type, the effect for percentage correct on comparison problems ($r = -.39$) was significantly stronger than the effect for percentage correct on solve problems ($r = -.12$), $Z = 2.52, p = .012$ (Meng et al., 1992). These findings suggest that it is possible that motivation could play a role in accounting for the effect of stereotype threat on math tests that are comprised only of “solve format” GRE problems.

Although the current work finds support for the role of motivation in producing the performance effects observed under threat, this research does not rule out the role of other processes in accounting for stereotype threat performance effects. Both affective (e.g. Bosson et al., 2004) and cognitive (Schmader et al., 2008) mechanisms also influence behavior under stereotype threat. In fact, it is quite possible that these other accounts could be elaborated to provide an explanation for the pattern of performance found in the current work. Although stereotype threat is likely a multiply mediated phenomenon (Steele et al., 2002), one strength of the motivation-based, mere effort explanation is that it can account for both facilitation and debilitation effects as a result of threat, whereas a single-process affective or cognitive impairment model cannot specifically account for the instances where stereotype threat facilitates performance (e.g., performance improvements on solve-type GRE problems in Exps. 1-5; Jamieson & Harkins, 2007; O’Brien & Crandall, 2003).
Additionally, the current work only examined the effects of subtle and blatant threat on performance, but other models of stereotype threat suggest other types of stereotype threat experiences that could be mediated by different mechanisms or combinations of mechanisms. For instance, rather than subtle and blatant threat, Shapiro and Neuberg’s (2007) Multi-Threat Framework argues for six qualitatively distinct stereotype threats, arising from the intersection of two dimensions—the target of the threat (the self vs. one’s group) and the source of the threat (the self vs. outgroup others vs. ingroup others).

In the current research, as in most other stereotype threat research, neither the blatant nor the subtle threat manipulation isolated the potentially unique influences of target and source factors. For instance, in the blatant threat manipulation used in Experiments 1-3, participants were aware that the experimenter could evaluate performance both at the level of the group (males vs. females) and at the level of the individual. Here, both target dimensions are activated by the threat manipulation. However, motivation may be produced by the activation of one (or both) of the targets. Although some initial research has begun isolating these distinct types of threat outlined by Shapiro and Neuberg (2007) (e.g., Inzlicht & Ben-Zeev, 2003; Wout, Danso, Jackson, & Spencer, 2008), future research is required to test potential mediating mechanism(s) of these multiple forms of stereotype threat.

Individual differences also limit the scope of stereotype threat research. Stereotype threat is experienced when stigmatized individuals are concerned with confirming negative group stereotypes. There can be no performance concerns if an individual does not view performance in the stereotyped domain as important. Consistent
with this view, research has indicated that to be threatened “a person probably needs to either care about having the ability or at least care about the social consequences of being seen as lacking the ability.” (Aronson, Lustina, Good, & Keough, 1999, p. 35). Similarly, if participants do not feel identified with the stereotyped group, they will not experience threat (Schmader, 2002). Thus, an individual not identified with the stereotyped domain or with the stereotyped group will not have any reason to be motivated to perform well in a “stereotype threat” situation. Thus, one must take care when generalizing the behavior of groups to the behavior of any one individual in that group.

It must also be acknowledged that to make predictions based on the mere effort account, one must be able to identify the prepotent response on the particular task under consideration. In some cases, prepotent responses will have been identified by previous research, as was the case for the quantitative GRE problems used in the current research. However, in other cases, preliminary work will be required for this purpose, as was the case for the RAT (Harkins, 2006). Even when the prepotent response for a given task has been identified and the findings for a given population of participants offer support for the mere effort account, caution must be exercised in generalizing the findings. For example, as noted previously, American students are typically taught to use the solving approach, rather than logic and estimation in solving math problems (e.g., Katz, et al., 2000; Stigler & Hiebert, 1999). However, students have to learn this approach, and so, we would not expect novices to display this tendency.

In addition, as students take more advanced courses, their approaches to solving math problems may change. For example, Spencer et al. (1999, Experiment 1) examined the effect of threat on the math performance of students majoring in math and physics
and found that these students exhibited no differences on quantitative GRE problems as a function of threat. However, Spencer and colleagues (1999) did find that stereotype threat impaired math and physics majors’ performance on problems taken from the GRE-Math subject test, which involves advanced calculus and knowledge of abstract algebra and real variable theory. Given these highly selected students’ experience and success in solving math problems with logic and estimation, it is unlikely that they would try to use the solving approach on abstract algebra problems. Thus, to generalize the mere effort account to highly selected populations, it would be necessary to first identify what approaches these students take to solving GRE Math problems.

Despite these caveats, the experiments presented here contribute to our understanding of the effects of stereotype threat by suggesting a route through which motivation can impact performance. Reducing the negative effects of stereotypes on the test performance of stigmatized group members is one goal of stereotype threat research and the current research suggests another means of achieving this goal. This research suggests that performance can be improved not only by breaking the link between stereotypes and performance (e.g., Johns et al., 2005), but also by assisting stigmatized individuals in channeling their efforts to best take advantage of their heightened motivational state.
References


Footnotes

1. McFall et al. (2009) found that the potential for evaluation produced the same pattern of results on the antisaccade task (Exp. 4).

2. In some cases, it may be possible to correctly answer solve problems by using logic and estimation, and to solve comparison problems by solving the equations. In fact, on some problems, the alternative approach is even more efficient. However, in the majority of cases, it is more efficient to solve the equations on solving problems and to use logic and estimation on comparison problems. For example, of the 30 problems that were used in the current research, it would have been possible to use logic and estimation on 5 of the 15 solve problems and to solve the equations on 6 of the 15 comparison problems. However, solving the equations was the most efficient approach for 14 of the 15 solve problems and using logic and estimation was the most efficient approach for 14 of the 15 comparison problems.

In addition to comparison and solve-type problems, the GRE-Quantitative test includes chart-type problems. These problems present test-takers with graphs, tables, and/or figures and ask questions regarding the information presented. GRE preparation books (see Lurie et al., 2004) note that chart problems focus on percentages and basic arithmetic processes. Thus, these problems generally do not require the knowledge of geometry, trigonometry, or algebra that is tested by the other problem types. Moreover, only 10-16% of the problems on quantitative GRE tests are chart-type problems.

3. Number correct was used in this analysis because it represents a single measure that allows a direct comparison of the relative strength of the debilitation and facilitation effects. Percentage solved could not serve this function because the facilitation of the threat participants’ performance on solve problems was reflected in the fact that they attempted more problems than no threat participants, but solved them at the same rate as these participants.

4. This measure may underestimate participants’ use of the solving approach. That is, participants could have used the solving approach but made the calculations in their head, rather than committing their work to paper. However, this underestimation should not impact group-level comparisons. There is no reason to believe that participants would be more or less likely to compute an answer in their head as a function of either stereotype threat or instruction. We did not estimate the participants’ use of the comparison approach from the scratch-paper analysis because in the great majority of cases, the use of logic and estimation does not require computation.

5. Number of times participants used the solving approach was used in the mediation analysis, rather than percentage, because the former provides a more valid indicator of the role that the solving approach played in comparison problem performance. Percentage assumes that whenever participants did not provide written evidence of using the solving approach, they were not using this method, even though they may have been doing so in their heads. On the other hand, although number could underestimate the number of times
the solving approach was used, it cannot misclassify problems as nonsolving when in fact they are solving.

In addition, we predicted the number of incorrect problems rather than number correct because our hypothesis is that using the solving approach on comparison problems directly leads to poor performance. However, not using the solving approach is only one requirement for success. To solve the problem correctly participants must also recognize that the comparison approach should be used, and implement it successfully. Thus, the most direct test of our hypothesis requires prediction of incorrect responses.

6. It should be noted that the average population solution rate does not differ across problem types on this test (Muncon. = 65.50%, SD = 8.34%; Mcon. = 64.72%, SD = 10.50%), p > .50.

7. Experiments 1 and 3 of this research are part of a manuscript currently in press at Personality and Social Psychology Bulletin (PSPB). Experiment 4 of this research is part of a larger study examining the role of motivation in producing subtle and blatant threat effects, and is part of a manuscript in preparation. Experiment 5 of this research is part of a manuscript under review at PSPB. Experiment 2 is currently not included as part of any manuscript in press, under review, or in preparation.
Appendix A

Examples of the problem types found on the quantitative GRE test. These problems appeared in the math tests used in this research.

Solve Type:

If the total surface area of a cube is 24, what is the volume of the cube?

a. 8  
b. 24  
c. 64  
d. $48\sqrt{6}$  
e. 216

For this problem, the individual must apply the formula for the volume of a cube, which is: length x width x height (all of which are the same value for a cube). To get the length of a side, the individual divides 24 by 6 (there are 6 faces on a cube) to obtain the area of one face, 4. The length of one side is 2 (area = length x width). To compute volume, the test-taker then cubes 2 to get the answer, 8. Thus, solve problems involve the application and computation of equations.

Comparison Type:

\[ n = (7)(19^3) \]

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of distinct positive factors of ( n )</td>
<td>10</td>
</tr>
</tbody>
</table>

a. The quantity in Column A is greater  
b. The quantity in Column B is greater  
c. The two quantities are equal  
d. The relationship cannot be determined from the information given
This problem can be solved by using intuition. First, the test-taker must realize that each number presented (7, 19, 3) is a prime number. Thus, the test-taker can logically deduce that the factors of the end product can only be multiples of 7 and 19. Thus, the factors of the final product are: 7, 7*19, 7*19^2, 19, 19^2, 19^3, plus the final product itself (7*19^3) and 1. Because the goal of the problem is not to compute the value of n, but simply to determine whether the number of positive factors of n is greater than, less than, or equal to 10, all the test-taker now needs to do is to add up the number of distinct positive factors (8) to find that Column B is greater than Column A. Thus, the correct answer choice is “b,” and only intuition and logic were used. No calculations were necessary.