STEREOTYPE THREAT, MENTAL ARITHMETIC, AND THE MERE EFFORT ACCOUNT

A dissertation presented

by

Allison E. Seitchik

to
The Department of Psychology

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

in the field of
Psychology

Northeastern University
Boston, Massachusetts
August 5, 2013
STEREOTYPE THREAT, MENTAL ARITHMETIC, AND THE MERE EFFORT ACCOUNT

by

Allison E. Seitchik

ABSTRACT OF DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology in the College of Science of Northeastern University, August 5, 2013
Abstract

Stereotype threat (i.e., 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, abstract)) has been shown to negatively impact performance. A working memory interference account for this effect suggests that cognitive resources that could be devoted to task performance are instead expended on processing the information resulting from the activation of the negative stereotype. It is this reduction in working memory capacity that produces the performance debilitation. More specifically, Beilock et al. (2007) suggest that these worries occupy the phonological loop of Baddeley’s (1986; 2000; Baddeley & Logie, 1999) multicomponent model of working memory. According to Trbovich and LeFevre (2003), participants solving mental arithmetic problems in a horizontal format rely on the phonological loop to keep track of intermediate values. Thus, Beilock et al. suggests that when threatened participants are presented with horizontal mental arithmetic problems that involve borrowing (e.g., 57 – 28), performance should be debilitated. To test this hypothesis, females completed horizontal or vertical modular arithmetic problems (subtracting the second number from the first, dividing by the mod, and then reporting, true or false, whether the result was a whole number) under threat and no threat. Consistent with the prediction, the performance of threatened females was debilitated on horizontal problems.

Although Beilock et al.’s (2007) findings support the working memory account, they do not rule out the possibility of a motivational explanation: the mere effort account (Harkins, 2006). The mere effort account argues that stereotype threat motivates stigmatized participants to want to perform well, potentiating the prepotent response on the given task (Jamieson & Harkins, 2007; 2009). If the prepotent response is correct, stereotype threat facilitates performance for
stigmatized individuals. If the prepotent response is incorrect and participants do not know, or lack the knowledge or time required for correction, performance is debilitated.

In the current studies, I examined how working memory and mere effort contribute to stereotype threat’s effects on horizontal arithmetic. In Experiment 1, replicating Beilock et al. (2007) in a paradigm sensitive to motivational effects, I found that threatened females did worse on horizontal modular arithmetic problems, but, consistent with a motivational account, they performed better on vertical problems than non-threatened females. Experiment 2 showed that threat facilitated division performance, suggesting that division alone does not account for the debilitation in modular arithmetic. Additionally, Beilock et al.’s account itself focuses on the subtraction component of the modular arithmetic problems. Focusing on subtraction, Experiments 3 and 4 identified a possible prepotent response for horizontal subtraction, termed the method of adjustment (e.g., adjust the second number to the nearest 10, subtract the two numbers, and then add the adjustment). Experiment 3 showed that participants prefer to work horizontal mental arithmetic problems from left-to-right and report using the method of adjustment. Consistent with the mere effort account, Experiment 4 found that stereotype threat potentiated the prepotent approach, the method of adjustment. Experiment 5 pitted the mere effort account against the working memory account. Working memory predicts debilitation effects on horizontal subtraction problems when participants need to use the phonological loop (i.e., entering answers from left-to-right), whereas mere effort predicts that the potentiated use of the method of adjustment should facilitate performance when answers must be entered from left-to-right. Results supported the mere effort account. Overall, these experiments support a motivational, mere effort, account of stereotype threat’s effects on horizontal subtraction performance.
Acknowledgments

First and foremost, I would like to extend my sincerest gratitude to my mentor, Dr. Steve Harkins, without whom my completion of this degree would not have been possible. You have taught me not only how to be a good experimental psychologist, but also how to better approach teaching, and life in general. You gave me the freedom to explore my interests and the guidance necessary to be a productive student. You have been an amazing mentor, and I cannot thank you enough for all that you have done for me.

I must also thank my committee members, Drs. Dave DeSteno and Nancy Kim, for all their helpful comments and revisions on my dissertation. Both of your insightful contributions helped me to think about my dissertation more thoroughly, which led to a more polished and comprehensive dissertation. As a research mentor, I would also like to thank Dr. Judy Hall for her guidance on both research and navigating life within academia. Your advice has helped me to keep an optimistic outlook for my future career path.

I would also like to thank my lab mates, Dr. Jeremy Jamieson and Adam Brown, for their help on not only this project but many others. I would especially like to thank Jeremy for his assistance throughout my graduate career with helping me gain a better understanding of stereotype threat. I would also like to thank my classmates, Martha Caffrey, Amy DiBattista, Sarah Gunnery, and Krista Hill, without whom my graduate experience would not be the same. We have celebrated every up and down together, and you have all offered me a shoulder to cry on or to hug in celebration. I cannot thank you all enough for every memory we made together. I especially want to thank Krista and Sarah for their unbelievable support and friendship. It has helped ease the transition into our next career moves knowing that you are both in my corner as friends and colleagues. To Jolie Baumann, Leah Dickens, Mollie Ruben, and Danielle Hartigan,
thank you for the laughs and talks of encouragement. You all gave me a great environment to work in every day.

I would like to extend my deepest gratitude to my family and friends who have been with me every step of the way. To my parents, Michael and Geri, you have been my loudest cheerleaders throughout all my walks of life. Whether I was training endless hours as an ice dancer or away at college and graduate school, you were always a call, email, or drive away. You taught me to work hard enough to attain my goals, and to relax enough to enjoy reaching them. To my sister, Jessica, you are a close second for being my loudest cheerleader. Throughout the years you have always been there for me and I cannot thank you enough for all of your love and support. Your video calls kept me going throughout my graduate career and offered several pick-me-ups when I needed them most.

Finally, I would like to thank my loving and supportive husband, Eric Correia. Toward the end of this process you have been my rock more than ever before. You have kept me calm and told me to stop worrying about things I could not control. It has been amazing to go through all of the milestones we have together and I cannot wait for the ones to come. Thank you for all your support throughout these last five years, especially these past few months when I know it could not have been easy to live with me. I love you.
## Table of Contents

Abstract 2

Acknowledgments 5

Table of Contents 7

List of Figures 8

Introduction 9

Chapter 1: Replicating Beilock et al.’s (2007) Effect 21

   Experiment 1 22

Chapter 2: Stereotype Threat and Division 28

   Experiment 2 28

Chapter 3: A Prepotent Response in Horizontal Subtraction 32

   Experiment 3 33

   Experiment 4 37

Chapter 4: Stereotype Threat and Subtraction 42

   Experiment 5 42

Chapter 5: General Discussion 49

References 58

Footnotes 66
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Experiment 1, number of problems correct</td>
<td>67</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Experiment 2, number of problems correct</td>
<td>68</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Experiment 3, percent use of traditional method</td>
<td>69</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Experiment 4, percent use of traditional method</td>
<td>70</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Experiment 5, predications</td>
<td>71</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Experiment 5, horizontal number of problems correct</td>
<td>72</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Experiment 5, vertical number of problems correct</td>
<td>73</td>
</tr>
</tbody>
</table>
Introduction

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, abstract). Stereotype threat has been shown to negatively impact the performance of stigmatized individuals in a variety of domains. For instance, past studies have examined stereotypes related to African-American’s underperformance on standardized tests (Aronson, Fried, & Good, 2002; Blascovich, Spencer, Steele, & Quinn, 2001; Steele & Aronson, 1995); White males’ athletic inferiority (Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Stone, 2002; Stone, Lynch, Sjomeling, & Darley, 1999); and females’s lack of ability in math and science domains (Beilock, Rydell, & McConnell, 2007; 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, 2004; Schmader & Johns, 2003; Sekaquaptewa & Thompson, 2003; Spencer, Steele, & Quinn, 1999). In each case, the research has shown that concern about confirming the relevant stereotype negatively impacts the performance of stigmatized individuals.

Even though the fact that stereotype threat negatively impacts stigmatized individuals’ performance is widely accepted, the mechanism(s) responsible for these effects is (are) less widely agreed upon. A number of explanations have been proposed, including anxiety (e.g., Bosson, Haymovitz, & Pinel, 2004; Spencer et al., 1999; Steele & Aronson, 1995); expectancy (e.g., Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003); arousal (e.g., Ben-Zeev et al., 2005; Blascovich et al., 2001; O’Brien & Crandall, 2003); cognitive load (e.g., Croizet, Després,
One explanation, working memory interference, has garnered significant attention in the literature as of late. This explanation conceptualizes stereotype threat as “a stressor in that a negative social stereotype that is primed in a performance situation poses a threat to one’s social identity (Schmader, 2002)” (Schmader & Johns, 2003, p. 442). Cognitive resources that could be devoted to task performance are instead expended on processing the information resulting from the activation of the negative stereotype. It is this reduction in working memory capacity that produces the performance debilitation reported in the stereotype threat literature.

To test this working memory deficit hypothesis, Schmader and Johns (2003, Experiment 3) asked females to complete two unrelated tasks under stereotype or no stereotype threat conditions: a working memory task and a standardized math test. The working memory task required participants to determine how many vowels were contained in a sentence. Following each sentence, a word was presented on the screen for 2 seconds. Upon completion of a set (i.e., four, five, or six sentence/word trials), participants were asked to recall as many words as they could in that set. After participants completed the working memory task, they took a 30-item multiple-choice standardized math test from the quantitative section of the Graduate Record Exam (GRE).

Supporting the working memory account, Schmader and Johns (2003) found that under stereotype threat conditions, females recalled fewer words than no stereotype threat females on the working memory task. In addition, stereotype threat females performed more poorly on the math task, as compared to no stereotype threat conditions. Finally, Schmader and Johns ran a mediational analysis and found that working memory mediated the relationship between
stereotype threat and math performance.

Beilock and her colleagues (e.g., Beilock et al., 2007) have also conducted research on the working memory account, attempting to spell out in greater detail exactly how this process might work. To do so, they used Baddeley’s (1986; 2000; Baddeley & Logie, 1999) multicomponent model of working memory. According to Baddeley (1986; 2000; Baddeley & Logie, 1999), there are four components or subsystems that are responsible for temporarily storing and processing different types of information, including: (1) a limited-capacity central executive, which coordinates the actions of the phonological loop and visuospatial sketchpad; (2) the phonological loop, which rehearses verbal information; (3) a visuospatial sketchpad, which serves as a mental blackboard and temporarily stores visual and spatial information; and (4) a multimodal episodic buffer, which connects information from the phonological loop, visuospatial sketchpad, and long-term memory into one episodic representation.

Beilock et al. (2007) argue that although the pressure-induced worries produced by stereotype threat may have some impact on central executive resources, these worries “rely more heavily on the phonological aspect of working memory, which is though to support inner speech and thinking in the service of complex cognitive activities” (p. 257). To support the argument that worries occupy the phonological loop, Beilock et al. (2007) cite work on the effect of anxiety on math performance in which thoughts and worries are considered a secondary verbal task (i.e., a verbal load on the phonological loop) (e.g., Darke, 1988; Markham & Darke, 1991). For example, Markham and Darke found that on a verbal task that required phonological resources, high anxiety participants took longer than low anxiety participants to complete the task but did not differ from low anxiety participants in performance on a visual task that utilized visual and spatial resources. Markham and Darke argued that the worries and anxieties
experienced by the high anxiety participants occupied the phonological loop, disrupting performance on a verbal task that required the loop but not affecting performance on the visual task that did not.

Beilock et al. (2007) tested the hypothesis that the worries produced by stereotype threat impact performance through their effect on the phonological loop by examining performance on “two types of math problems that are equally dependent on central executive process (e.g., because they require the same algorithmic computation; Baddeley & Logie, 1999), but are differentially dependent on phonological resources (e.g., because the maintenance and rehearsal of intermediate steps are represented in different forms)” (p. 258). They argue that work by Trbovich and LeFevre (2003) shows that participants presented mental arithmetic problems in a horizontal format rely more heavily on the phonological loop (remembering intermediate values in verbal form), whereas participants presented mental arithmetic problems in a vertical format rely more heavily on the visuospatial sketchpad (using a spatial mental work space in the same way that such problems are solved on paper). For example, the paper-and-pencil solution approach for vertical problems is usually to process each column of numbers sequentially from right-to-left, utilizing more visual than verbal (phonological) working memory resources (Trbovich & LeFevre). On the other hand, horizontal problems evoke strategies that involve adjusting or decomposing the problem (e.g., $46 + 9$ into $46 + 4 + 5$) which are more likely to involve verbal codes and, thus, the phonological loop (Imbo & LeFevre, 2010; Trbovich & LeFevre).

As a result, Beilock et al. (2007) argue that performance on horizontal problems, which require the use of the phonological loop, should be disrupted by stereotype threat, whereas performance on vertical problems, which depend more on the visuospatial sketchpad should not.
To test this hypothesis, Beilock et al. (2007, Experiment 1) asked females to complete modular arithmetic problems in a horizontal orientation under stereotype threat or no stereotype threat conditions. For example, participants saw: $52 = 24 \pmod{3}$. They were to subtract the second number from the first, divide by the mod, and then report, true or false, whether the result was a whole number. The problems were either high-demand problems, requiring a double-digit borrow subtraction operation, or low-demand problems, requiring a double-digit no borrow subtraction operation. All participants completed two blocks of 24 problems, with each block containing 8 low-demand problems, 8 high-demands problems, and 8 fillers. The first block served as the baseline (control) condition and participants were told that the purpose of the study was to examine how individuals learned a new math skill. Before the second block, half of participants received the stereotype threat manipulation whereas the other half was told that the researchers were interested in “why some people are better at math than others are” (Beilock et al., 2007, p. 259). To manipulate stereotype threat, participants were told that the study was “investigating why men are generally better than women at math” (Beilock et al., 2007, p. 259). Beilock et al.’s (2007) results supported their hypothesis. When females were threatened, they performed significantly worse on horizontal/high-demand problems compared to their baseline (no threat) performance. There was no performance debilitation in any other condition. In fact, performance slightly improved in the stereotype threat/horizontal/low-demand condition compared to baseline performance.

In another experiment, Beilock et al. (2007, Experiment 3A) had females complete low- or high-demand modular arithmetic problems in either a horizontal or vertical orientation. All participants completed two blocks of 20 problems, 10 low-demand and 10 high-demand problems. Similar to Experiment 1, the first block served as the baseline, no threat comparison
condition. After completing the first block, the stereotype threat manipulation used in Experiment 1 was implemented for all participants (Beilock et al., 2007, Experiment 3A). Thus, all participants performed under stereotype threat and served as their own baseline comparison. Results again supported their hypothesis; stereotype threat did not make a difference on vertical problems regardless of problem demand (i.e., high or low), whereas threat did make a difference on horizontal problems (Beilock et al., 2007, Experiment 3A). More specifically, only horizontal/high-demand problems showed a significant performance decrement under stereotype threat as compared to baseline (no threat).

Although these findings are consistent with Beilock et al.’s (2007) working memory account, they do not rule out the possibility of a motivational explanation: the mere effort account (Harkins, 2006), which argues that stigmatized participants are motivated to disprove the stereotype, and it is this effort, in and of itself, that debilitates performance. The mere effort account was suggested by Harkins’s (2006) attempt to isolate the processes that mediate the effect of evaluation on complex task performance. This approach suggests that the potential for evaluation, like stereotype threat, arouses participants’ concern about their ability to perform well on the task. In fact, many of the processes proposed to mediate stereotype threat performance effects (e.g., processing interference, withdrawal of effort, and arousal) have also been proposed to explain performance differences in the evaluation-performance domain, and, like the stereotype threat literature, researchers have not come to a consensus as to what the mediating mechanisms are.

Harkins (2006) suggested that the mediating process could be identified through a molecular analysis of performance on a specific task. To this end, Harkins (2006) examined the effect of evaluation on the performance of the remote associates task (RAT), which requires
participants to look at a set of three words (e.g., *lapse*, *elephant*, and *vivid*) and to generate a fourth word that is related to each word in the given triad (in this case *memory*). Harkins (2001) has shown that the potential for evaluation produces the typical pattern of performance on this task: Participants who anticipate evaluation by the experimenter solve more simple triads than no-evaluation participants, whereas participants who anticipate experimenter evaluation solve fewer difficult triads than no-evaluation participants.

Harkins’s (2006) analysis showed that these findings result from the fact that the potential for evaluation motivates participants to want to do well, which potentiates whatever response is prepotent, or most likely, on the given task. Cast in Hullian terms, on any given task there are a variety of responses that can be made, and these responses can be arranged in a hierarchy according to their likelihood of occurrence. Whatever the participant is most inclined to do in that situation is at the top of the hierarchy and is thus, the “dominant” or prepotent response (Zajonc, 1965). For example, when performing the Stroop Color-Word Test (Stroop, 1935), color words (e.g., green) are presented that are printed in some other color (e.g., red), and participants are asked to call out the color (e.g., red). On this task, the prepotent response is to read the word (e.g., green).

The mere effort account argues that this dominant or prepotent response is potentiated by the potential for evaluation. On the RAT, Harkins’s (2006) analysis showed that the prepotent response is to generate words that are closely related to one of the triad members. Because on simple items the correct answers tend to be a close associate of at least one of the triad members, the greater effort on the part of participants subject to evaluation leads to the production of more close associates and to better performance.
On the other hand, on the complex items, the associations between the triad members and the correct answer are much weaker (i.e., the associates are more remote), and the participants are extremely unlikely to produce the solution by generating associates for the individual triad members. For example, if presented with the triad member note, a participant would be extremely unlikely to produce the associate, “bank.” Nonetheless, when the participant considers the word note, the solution, “bank,” is weakly activated. Likewise, the solution “bank” is also weakly activated when the participant considers the other two triad members, river and blood. If this were the only process operating, this weak activation should accumulate over time, leading to the emergence of the correct answer. However, when participants actively test close associates as solutions for the triads, these associates are highly activated, and they strongly inhibit the activation of the remote (weak) associates. Thus, generating close associates, the same behavior that facilitates the performance of participants subject to experimenter evaluation on simple items, debilitates that performance on complex items.

Zajonc’s (1965) drive theory account of social facilitation effects also accords a central role to prepotent or dominant responses. Drive theory contends that the presence of others produces arousal, which increases drive. Increased drive enhances the probability of the emission of dominant, or prepotent, responses, which are likely to be correct on simple tasks but incorrect on difficult ones. In fact, Cottrell (1972) argued that this drive was the result of the participants’ apprehension about the fact that they would be evaluated.

Thus, both mere effort and Cottrell’s (1972) evaluation apprehension account of social facilitation effects predict that the potential for evaluation will potentiate dominant or prepotent responses. However, in the case of mere effort, this potentiation results from the motivation to perform well, which should also lead to an effort to correct the incorrect response if the
participant recognizes that his or her response is incorrect, knows the correct response, and has
the opportunity to make it. In contrast, Cottrell’s (1968, 1972) modification of Zajonc’s drive
theory suggests only that the positive or negative anticipations produced by the presence of
others nonselectively energize individual performance (i.e., potentiate the dominant response).
Of course, on a task like the RAT, they are unable to distinguish between mere effort and
evaluation apprehension accounts because even if the participants know that the response is
incorrect, they do not know how to correct it.

However, an inhibition task, like the Stroop, does allow us to see the effect of the
motivation to correct. On this task, the correct answer is quite obvious. So, although the initial
tendency for participants subject to evaluation to read the color will be stronger than that of their
no-evaluation counterparts, it will be quite clear to these participants that this response is
incorrect. Given enough time to counter the effect of response potentiation, the mere effort
account predicts that the heightened motivation of the evaluation participants to do well will lead
them to produce the correct answer more quickly than the no-evaluation participants, whereas
drive theory or evaluation apprehension predicts only response potentiation. Consistent with the
mere effort prediction, McFall, Jamieson, and Harkins (2009) found that if required to produce a
response in a brief time (1 s or 750 ms), then participants subject to evaluation made more errors
than no-evaluation participants, but when given up to 2 s to respond, participants subject to
evaluation responded more quickly than no-evaluation participants with no difference in
accuracy.

The mere effort account of stereotype threat performance effects argues that stereotype
threat should operate like the potential for evaluation in that threat will motivate participants to
want to perform well on the task. In fact, in this case the participant’s performance will not only
reflect on him or her, but also on his or her group. To test the mere effort account in relation to stereotype threat, Jamieson and Harkins (2007) used the antisaccade task (Hallett, 1978), an inhibition task like the Stroop. This task requires participants to respond to a target presented on either side of a display screen after a cue appears on the opposite side of the display. The objective of the task is to accurately describe the orientation of the target arrow (i.e., whether it is pointing up, left, or right). Participants are told not to look at the cue, but instead to look at the opposite side of the display where the target will be presented. However, participants have a reflexive-like prepotent tendency (i.e., reflexive saccade) to look at the cue that must be inhibited or corrected in order to perform well.

This task was framed as a measure of visuospatial capacity that was highly related to ability in math. Stereotype threat was manipulated by telling females either that gender differences had been found on this task (stereotype threat) or that such differences had not been found (control). The mere effort account argues that stereotype threat should produce the same basic pattern of findings on the antisaccade task as is produced by the potential for evaluation on other inhibition tasks like the Stroop. When given insufficient time to correct for the prepotent tendency (i.e., at a brief display time; e.g., 150 ms), the more motivated threat participants should be less accurate than controls in their ability to correctly identify target orientation. However, when the display time is increased enough to allow enough time for correction (e.g., 250 ms), stereotype threat participants should be able to respond to the target more quickly than controls, as a result of increased motivation to perform well, and this is exactly what Jamieson and Harkins (2007; Experiments 1 & 2) found.

Jamieson and Harkins (2007; Experiment 3) also used an eye tracker to conduct a more fine-grained analysis of performance on this antisaccade task at a display time that permitted correction.
Under these conditions, the mere effort account predicted that the participants under threat would look the wrong direction, toward the cue, more often than would participants in the control group, because the motivation to perform well potentiates the prepotent response. At this point, if the participants have failed to inhibit the reflexive saccade, their eyes are at the cue and they must launch a corrective saccade to get to the target site. If they have successfully inhibited the saccade, they must launch a correct saccade to the target site from the fixation point. Because correct and corrective saccades are each an “extreme example of a voluntary saccade” (Sereno, 1992, p. 92), the motivation to correct should reduce the latency to launch each type of saccade, and, as a result, the threatened participants should launch these saccades faster than control participants. Finally, after the participants’ eyes arrive at the target area, the participant must determine the target’s orientation and press the appropriate response key. When the participants see the target, the mere effort account predicts that the greater motivation of participants subject to stereotype threat would lead them to respond more quickly than would participants in the control condition. Jamieson and Harkins (2007) found support for each of these predictions.

After finding this initial support, Jamieson and Harkins (2009) then tested the mere effort account using a math task: the quantitative section of the Graduate Record Exam (GRE). The quantitative section of the GRE is comprised mainly of two types of problems: solve problems and comparison problems. Solve problems are word problems that tend to require the application of an equation or algorithm. Comparison problems require the individual to compare quantities from one column to another. These problems are usually solved by simplifying the terms in each column or using logic, intuition, and/or estimation. For standardized test math problems, the prepotent response is to apply an equation or algorithm (e.g., Gallagher & De Lisi, 1994; Gallagher, De Lisi, Holst, McGillicuddy-De Lisi, Morely, & Cahalan, 2000; Quinn & Spencer,
2001). For example, Gallagher et al. (2000, Experiment 2) found that on math problems on the Scholastic Aptitude Test (SAT), participants used a conventional (i.e., solving; applying an equation) approach 55.5% of the time, whereas they used the unconventional (i.e., comparison, logic, intuition, and/or estimation) approach 10% of the time. They found this pattern of findings to hold true regardless of the type of problem being solved. More specifically, on solve problems participants used the conventional, solving approach 66% of the time and the unconventional, comparison approach 9% of the time with the remainder comprising guesses (13%), omissions (6%), and unknowns (6%). On comparison problems, participants used the solving approach 45% of the time and the comparison approach 11% of the time, with the remainder comprising guesses (20%), omissions (12%), and unknowns (12%) (Gallagher et al., 2000, Experiment 2).

Therefore, according to the mere effort account, on GRE solve problems, the prepotent response is correct and stereotype threat should potentiate this response leading to better performance than no stereotype threat. Of course, this assumes the individual knows the correct equation or algorithm to apply to the problem. On the other hand, on GRE comparison problems the prepotent response is incorrect and inefficient, and may often lead to an incorrect answer. In addition, unlike the antisaccade task, it is unlikely that participants will be able to recognize that the prepotent response is incorrect on these comparison problems. Thus, on these problems, the potentiation of the prepotent response should debilitate performance for females under stereotype threat. This was exactly what Jamieson and Harkin (2009) found; when females were subjected to stereotype threat, they performed better on solve problems but worse on comparison problems than females under no stereotype threat.
Chapter 1: Replicating Beilock et al.’s (2007) Effect

As noted above, Beilock et al.’s (2007) results support the working memory account but do not rule out the possibility of the motivational, mere effort explanation. The purpose of the present experiments was to determine whether mere effort could provide an account of the effect of stereotype threat on the performance of Beilock et al.’s (2007) horizontal problems, while also testing the working memory account. As a first step, I sought to determine whether Beilock et al.’s (2007) findings could be replicated in a paradigm in which it is possible to see motivational effects, if they are present. These effects cannot be seen in Beilock et al.’s (2007) paradigm because participants were given problems in blocks of only 24, and were given as much time as necessary to solve these problems. To see motivational effects, it is necessary to give participants more problems than they could be expected to solve in the amount of time provided. Under these circumstances one can see the effects of facilitation (cf., Jamieson & Harkins, 2007; 2009) as well as debilitation.

However, there is no reason to expect that this difference in method should change Beilock et al.’s (2007) basic finding. In this paradigm, the effect of worries occupying the phonological loop on horizontal problems would be reflected in the finding that threatened participants solve fewer of the horizontal problems than control participants, rather than solve each problem more slowly as in Beilock et al.’s (2007) paradigm. In addition, Beilock et al. (2007) suggest that vertical problems are worked on the visuospatial sketchpad in a way that is similar to the way that they are solved on paper, which would be minimally impacted by the worries produced by stereotype threat. As a result, if the threatened participants are more motivated than the no threat controls, in a paradigm that is sensitive to motivational effects, threatened participants may outperform non-threatened ones. Of course, this facilitation effect
would not be predicted by the working memory account, nor can it be seen in Beilock et al.’s (2007) paradigm, but it would be consistent with the mere effort account.

**Experiment 1**

In the first experiment, participants were asked to complete Beilock et al.’s (2007) modular arithmetic problems presented horizontally or vertically in a paradigm that was sensitive to motivational effects. In a paper and pencil format, participants completed either horizontal or vertical problems and were given 10 minutes to complete as many problems as accurately as possible. Beilock et al.’s (2007) research suggested that participants required approximately 7 seconds to solve high-demand modular arithmetic problems. Thus, participants should not be able to complete the entire 120 problem set in the allotted 10 minutes. By giving participants a set amount of time and an unrealistic number of problems to finish, I could see the effect of threat on the performance of horizontal and vertical problems in a paradigm that is sensitive to the effect of motivation.

**Method**

**Participants**

Eighty undergraduates (54% female) from a northeastern university participated in the study for course credit.

**Procedures**

Participants came into the lab individually and were told that they would be solving math problems during the session. Participants were then told about the math task they were about to complete.

I adopted my methods from Beilock et al. (2007), and used modular arithmetic (Gauss, 1801, as cited in Beilock et al., 2007) as my math task. The object of modular arithmetic is to
judge the validity of problems such as $51 = 19 \ (\text{mod} \ 4)$. Participants were to subtract the second number from the first (e.g., $51 - 19$), divide by the mod (e.g., $32 \div 4$), and then report whether or not the result was a whole number. If the dividend was a whole number, the problem was “true.” Unlike Beilock et al. (2007), I used all high-demand math problems that involved numbers greater than 10 and included a borrow operation in the subtraction. Beilock et al. (2007) found a difference between stereotype threat and no stereotype threat performance on high-demand problems only; thus only high-demand problems were used in the present study. These high-demand problems require a sequence of steps and the maintenance of multiple pieces of information, and are thus more taxing on working memory systems than problems that do not employ borrow operations (e.g., Ashcraft, 1992; Beilock et al., 2007; DeStefano & LeFevre, 2004).

Consistent with the methodology reported by Beilock et al. (2007), half of the problems were “true” problem and each “true” problem had a “false” correlate that only differed as a function of the number involved in the division (mod) statement. For instance, if the “true” problem $51 = 19 \ (\text{mod} \ 4)$ was presented, then a “false” correlate problem $51 = 19 \ (\text{mod} \ 3)$ was also presented at some point in the problem set. This pairing was conducted to equate “true” and “false” problems as much as possible in terms of the specific numbers used in each equation.

Because all “false” problems were correlates of a “true” problem, I sought to reduce any effect familiarity had on performance. Problems were randomized throughout the test with the constraint that if a particular problem appeared on a page, then its correlate could not appear until the participant had completed at least 10 problems in between. One randomized version of the problems was made and given to all participants in each of the orientation conditions.

The paper and pencil problem sets included a total of 120 modular arithmetic problems,
with 10 problems per page. All problems were presented either horizontally or vertically in a problem set. After the task was described, participants were asked to complete 8 practice problems to familiarize themselves with the task. Once the practice problems were completed, the stereotype threat manipulation was induced. Stereotype threat participants were told,

The task you are about to complete is a test of math ability. As you know, there has been some controversy about whether there are gender differences in math and spatial ability. Previous research has demonstrated that gender differences exist on some tasks, but not on others. In our lab, we examine performance on both kinds of tasks. The task on which you are about to participate has been shown to produce gender differences.

The no stereotype threat manipulation was identical to the stereotype threat manipulation except that the last sentence read, “The task on which you are about to participate has not been shown to produce gender differences.” Previous research using a similar manipulation to the one used in the present study has shown that it produces stereotype threat effects (e.g., Brown & Pinel, 2003; Keller & Dauenheimer, 2003; Jamieson & Harkins, 2007; O’Brien & Crandall, 2003; Spencer et al., 1999). There was no specific mention as to whether males outperformed females or vice versa, only that gender differences did or did not exist on the task. It was expected that participants would infer that females would perform more poorly than men on the basis of the societal stereotype that men are superior to females in mathematical ability.

Upon completion of the math task, all participants completed a post-task questionnaire that included manipulation check items. Two questions allowed us to examine 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).

Results

Manipulation Checks

Responses to the questions asking to what extent gender differences exist on the task and which gender performs better on the task were analyzed in a 2 (threat condition: stereotype threat vs. no stereotype threat) x 2 (problem orientation: horizontal vs. vertical) x 2 (gender: male vs. female) between subjects ANOVA. Main effects for stereotype threat condition were found for both questions but no other reliable effects were found. Participants subject to stereotype threat reported that gender differences existed to a greater extent ($M = 6.33, SD = 2.25$) than no stereotype threat participants ($M = 2.78, SD = 2.02$), $F(1, 72) = 53.93, p < .001, d = 1.73$. Stereotype threat participants also reported that males perform better on the task than females to a greater extent ($M = 3.74, SD = 1.70$) than no stereotype threat participants ($M = 6.02, SD = .91$), $F(1, 72) = 52.41, p < .001, d = 1.71$. These results indicate that the stereotype threat manipulation used was successful. Participants in the threat condition were aware of the negative group stereotype, and females were expected to perform more poorly than males.

Task Performance

Consistent with Beilock et al.’s (2007) practice, data were excluded from analyses if a participant solved less than 75% of the problems correctly. However, all participants solved at least 75% of the problems they attempted correctly and, thus, no participants were excluded from analyses.

Math performance was analyzed in a 2 (threat condition: stereotype threat vs. no stereotype threat) x 2 (problem orientation: horizontal vs. vertical) x 2 (gender: male vs. female) between subjects ANOVA. There were no reliable effects for percentage correct and, thus, I
analyzed number of problems correct. Results revealed a main effect for problem orientation, 
\( F(1, 72) = 11.21, p < .001, d = .79 \), such that participants answered fewer problems correctly 
when the problems were horizontally oriented (\( M = 81.00, SD = 24.87 \)) than when vertically 
oriented (\( M = 68.46, SD = 13.85 \)). A stereotype threat x problem orientation interaction, \( F(1, 72) 
= 15.31, p < .001, d = .92 \), was also found. However, the main effect and two-way interaction 
must be interpreted in the context of the three-way interaction found between stereotype threat, 
problem orientation, and gender, \( F(1, 72) = 6.65, p = .012, d = .61 \) (see Figure 1). No other main 
effects or interactions were found. To decompose the three-way interaction, two two-way 
ANOVAs were conducted, one for females and one for males.

A 2 (stereotype threat condition: stereotype threat vs. no stereotype threat) x 2 (problem 
orientation: horizontal vs. vertical) ANOVA for females revealed a main effect for problem 
orientation, \( F(1, 72) = 7.40, p = .008, d = .64 \), such that females solved significantly more 
vertically oriented problems (\( M = 79.73, SD = 26.66 \)) than horizontally oriented problems (\( M 
= 66.86, SD = 14.06 \)). There was no main effect for stereotype threat. However, a Threat x Problem 
Orientation interaction was revealed for females, \( F(1, 72) = 22.73, p < .001, d = 1.12 \). 
Consistent with Beilock et al.’s (2007) findings, a planned contrast (Kirk, 1995) showed that 
stereotype threat females answered significantly fewer horizontal problems correctly (\( M = 58.50, 
SD = 10.61 \)) than no stereotype threat females (\( M = 74.45, SD = 12.68 \)), \( F(1, 72) = 4.17, p < .05, 
d = .48 \). A second planned contrast showed that females who were subject to threat answered 
significantly more vertical problems correctly (\( M = 99.50, SD = 17.80 \)) than no stereotype threat 
females (\( M = 63.25, SD = 21.12 \)), \( F(1, 72) = 22.56, p < .001, d = 1.12 \).

A 2 (stereotype threat condition: stereotype threat vs. no stereotype threat) x 2 (problem 
orientation: horizontal vs. vertical) ANOVA for males revealed a main effect for problem
orientation, $F(1, 72) = 4.17, p < .05, d = .48$, such that males presented vertical problems solved more of them correctly ($M = 82.47, SD = 23.26$) than males presented horizontal problems ($M = 70.33, SD = 13.76$). There was no reliable main effect for stereotype threat nor an interaction between threat and problem orientation.

**Discussion**

The results from Experiment 1 replicated the debilitation effect that Beilock et al. (2007) found for horizontal problems: females under stereotype threat performed significantly worse on horizontal problems than females not under stereotype threat. I also found that on vertical problems, stereotype threat females performed significantly better than no stereotype threat females. This effect is consistent with a motivational account, like mere effort, but would not be predicted by the working memory account.

As in past stereotype threat research (e.g., Jamieson & Harkins, 2007 (Experiment 1); Schmader & Johns 2003), I also found that males were not affected by the threat manipulation. Additionally, I found that males performed better overall on vertical problems than on horizontal problems. This finding is consistent with the findings of past research on the effects of problem orientation (e.g., Trbovich & LeFevre, 2003).
Chapter 2: Stereotype Threat and Division

Experiment 1 replicated Beilock et al.’s (2007) finding that stereotype threat debilitates female performance on horizontal modular arithmetic problems using a paradigm that is sensitive to motivational effects. However, these modular arithmetic problems require both subtraction and division. Trbovich and LeFevre (2003) have shown that horizontal subtraction by itself draws more on phonological resources than vertical subtraction. The division component is not required to produce these effects. In addition, Beilock et al.’s (2007) account itself focuses on the subtraction component of the modular arithmetic problems, and has nothing to say about the division portion (see also Beilock, Kulp, Holt, & Carr, 2004). In fact, it is unlikely that the participants rely on working memory much to solve the division portion of the problems. It is more likely that the division answers are simply recalled as math facts. For instance, students learn and memorize their multiplication tables (up to 12 x 12), which then become arithmetic facts, thus allowing multiplication to become routinized (e.g., Imbo & Vandierendonck, 2007a; 2007b). It has been shown that retrieving information, such as math or arithmetic facts, from long-term memory utilizes less working memory resources than non-retrieval strategies (e.g., Fürst & Hitch, 2000; Beilock et al., 2004). In fact, to the extent that threatened females are motivated to perform well, as suggested by the mere effort account, I could find that females subject to threat will outperform the no threat controls on this rote task. This hypothesis was tested in the next experiment.

Experiment 2

To look at the division component, the subtractions on Beilock et al.’s (2007) problems were performed (e.g., (57 = 28) mod 3 would be presented as 19 mod 3) and then participants under threat or no threat were asked to solve as many of the division problems as possible in 10
minutes (e.g., 19 mod 3, answer is false). Because the division problems require little, if any, working memory resources, Beilock et al.’s (2007) working memory account would predict no differences as a function of threat. To the extent that the division portion of the problems is solved by recalling math facts, it would be consistent with a motivational account, like mere effort, to find that threatened participants outperform the no threat controls.

Method

Participants

Thirty undergraduate females from a northeastern university participated in the study for course credit.¹

Procedure

The total number of problems was increased to 180, all of the subtraction portions of the problems were performed, and participants were asked to divide the resulting number by the mod (e.g., 32 ÷ 4), and then to report whether or not the result was a whole number. If the dividend was a whole number, the problem was “true.” All problems were presented horizontally and included half “true” problems and half “false” problems.

The procedure was identical to that in Experiment 1. After the task was described, participants were asked to complete 8 practice problems to familiarize themselves with the task. Once the practice problems were completed, the stereotype threat manipulation from Experiment 1 was implemented. Participants were given 10 minutes to complete as many of the division problems as they could and then completed the manipulation check items from Experiment 1.

Results

Manipulation Checks
Responses to the question asking to what extent gender differences exist on the task and which gender performs better on the task were analyzed in a one-way ANOVA with stereotype threat as the independent variable. Effects for stereotype threat condition were found for both questions. Participants subject to stereotype threat reported that gender differences existed to a greater extent ($M = 6.47, SD = 2.45$) than no stereotype threat participants ($M = 2.33, SD = 1.54$), $F (1, 28) = 30.65, p < .001, d = 2.09$. Stereotype threat participants also reported that males perform better on the task than females to a greater extent ($M = 3.87, SD = 1.60$) than no stereotype threat participants ($M = 5.73, SD = .80$), $F (1, 28) = 16.38, p < .001, d = 1.53$. These results indicate that the stereotype threat manipulation was successful. Participants in the threat condition were aware of the negative group stereotype, and females were expected to perform more poorly than males.

**Task Performance**

Consistent with Beilock et al.’s (2007) practice, data were excluded from analyses if a participant solved less than 75% of the problems correctly. However, all participants correctly solved at least 75% of the problems they attempted and, thus, no participants were excluded from the analyses.

Math performance was analyzed in a one-way between subjects ANOVA with stereotype threat as the independent variable. There was no reliable effect for percentage correct and, so, the number of problems correct was analyzed. Results revealed a main effect for stereotype threat, $F (1, 28) = 6.68, p = .015, d = .98$. As shown in Figure 2, stereotype threat females answered significantly more problems correctly ($M = 127.87, SD = 36.62$) than no stereotype threat females ($M = 93.87, SD = 35.44$).

**Discussion**
Experiment 2 showed that stereotype threat does not debilitate performance on division problems. In fact, threatened females solved more division problems than females who were not threatened. Thus far, the results have shown that performance on both vertical problems (Experiment 1) and on the division component of the modular problems (Experiment 2) are facilitated by threat. These findings are consistent with a motivational account, like mere effort, but are not predicted by the working memory account, although they are also not necessarily inconsistent with it either.

Of course, Experiment 2 only shows that stereotype threat does not debilitate performance on division problems taken alone. It is possible that the subtraction and division components of Beilock et al.’s (2007) horizontal modular problems interact in some way to produce the debilitation effect. However, the core of Beilock et al.’s (2007) argument is that threat debilitates performance on the modular horizontal problems because solving the subtraction portion requires phonological resources (e.g., remembering intermediate values) that are taken up by the thoughts and worries produced by threat. Thus, in the next experiments, I attempted to develop and test a mere effort account for performance on the subtraction component of the modular arithmetic problems.
Chapter 3: A Prepotent Response in Horizontal Subtraction

Key to the development of a mere effort account is the identification of the prepotent response. As a first step in doing so, I compared how participants approach the subtraction component of Beilock et al.’s (2007) modular arithmetic problems when they are presented vertically and horizontally without stereotype threat.

As noted previously, Trbovich and LeFevre (2003) argue that although both vertical and horizontal math problems require the central executive, problems presented in a vertical format rely more heavily on the visuospatial sketchpad, whereas problems presented in a horizontal format rely more heavily on the phonological loop. They suggest that participants use the visuospatial sketchpad to approach the vertical problems in the “traditional” manner, as though they are working them on paper, starting on the right with the units place and then moving to the decades (i.e., units to decades). In contrast, when problems are presented in a horizontal format, Trbovich and LeFevre suggest that participants may solve the problems by making adjustments that produce intermediate values (e.g., participants adjust the second number to the nearest 10, subtract the two numbers, and then add the adjustment), which involve verbal codes that require the phonological loop, an approach that I will call the method of adjustment.

This analysis suggests that if participants were allowed to solve vertical subtraction problems and could enter the answers from right-to-left or from left-to-right, they would tend to enter the answers from right-to-left, because they calculate the units digit first followed by the decades digit. However, on horizontal problems, the entire answer is produced in the solution procedure, and, therefore, they should tend to enter the answers from left-to-right. I tested this hypothesis in Experiment 3. If supported, it suggests the approach that participants take to
horizontal problems, which would then provide a candidate for the prepotent response on this problem type.

**Experiment 3**

Female participants were presented subtraction problems taken from the modular arithmetic problem set used in Experiment 1 in either a horizontal or a vertical format on a computer screen but without the manipulation of stereotype threat. For both horizontal and vertical problems, participants were told that they could enter the answers starting with the left digit (i.e., from left-to-right) or starting with the right digit (i.e., from right-to-left).

To the extent that participants solve horizontal subtraction problems using the method of adjustment (Trbovich & LeFevre, 2003), I should find that the answers to these problems are entered from left-to-right because using the method of adjustment produces the entire answer. For example, to solve the problem 57 - 28 using the method of adjustment, one rounds 28 to 30, then subtracts 30 from 57, and adds 2 to 27, yielding the whole answer, 29. One would then enter this answer from left-to-right. In contrast, if one uses the visuospatial sketchpad to solve vertical problems, the “traditional” method would be used and answers would be entered from right-to-left. That is, to solve 57 – 28, one would borrow a ten from the decades column to perform the subtraction 17 – 8, which equals 9. One could then enter the 9 in the right hand box, followed by the decade operation, 4 – 2 = 2, which would be entered in the left hand box.

In addition to finding that horizontal problems are entered from left-to-right, and vertical problems are entered from right-to-left, I could also replicate the typical finding that more vertical problems are solved than horizontal ones (Trbovich & LeFevre, 2003). However, it should be noted that in this paradigm, participants are self-selecting the direction in which they enter the digits, which could affect their performance.
Method

Participants

Twenty-nine females from a northeastern university participated in the study for course credit.

Procedure

Participants came into the lab individually and were told that they would be solving math problems. The task was then described to them. The task included 160 high-demand subtraction problems. As described by Beilock et al. (2007), these high-demand problems were two-digit subtraction problems that involved numbers in the problems that were greater than 10 and included a borrow operation. Problems were presented either horizontally or vertically.

Participants completed the problems on a computer. For the participants who saw horizontal problems, two boxes were presented to the right of the problem (e.g., 57 − 28 = □ □). For participants who saw vertical problems, the boxes were beneath the numbers. Problems were presented in random order in the middle of the screen and remained until the participant answered. The interval between problems was 750 ms. Participants were told that the program would tell them when they had finished the task. In actuality, the program always stopped participants after 10 minutes.

There was no stereotype threat manipulation in this experiment. Instead, all participants were given the following instructions:

The math problems are all subtraction and they all have two-digit answers. There will be two boxes where you will enter your two-digit answer. You can enter it however you want, from right-to-left or left-to-right. You indicate what digit you want to enter first by clicking on that box first. For example, if you want to enter the right digit first, click on
the right box. For the left digit first, click on the left box. After you enter the first digit, the program automatically goes to the next digit and you do not need to click on it. You just need to click on the first box you want to enter.

After completing the task, participants responded to the same threat manipulation check items that were used in Experiments 1 and 2. In addition, they were asked about the approach(es) that they took to solving the problems. They read: “In working on the problems (e.g., 57 – 38), you may have followed a traditional approach (e.g., started on the right; to subtract the 8 from the 7, you borrowed one from the 5 in the tens column making the 7 into 17; 17 – 8 = 9; subtracted the 3 from the 4 in the tens column yielding a 1; and then put the number together, 19). Or you could have used some other approach. For example, you could have added 2 to 38 (40), subtracted 40 from 57, and then added back the 2 to get the final answer, 19. Or you could have used some variant of this approach. On what percentage of the problems did you use the traditional approach? _____% On what percentage of the problems did you use some other approach? ______% The two percentages should sum to 100%. If you used a non-traditional approach, please describe the approach or approaches that you used.”

Results

Manipulation Checks

Responses to the question asking to what extent gender differences exist on the task and which gender performs better on the task were analyzed in one-way ANOVAs with orientation (horizontal vs. vertical) as the independent variable. These analyses revealed no significant effects. The overall mean for the extent to which gender differences exist on this task was 2.99 (where 1 is no gender difference), and for which gender performs better, the overall mean was 5.69 (where 6 is no gender difference).
Task Performance

Once again, data were excluded from analysis if a participant solved less than 75% of the problems correctly but all of the participants met this criterion.

An analysis of the proportion of digits entered from left-to-right in a one-way ANOVA with orientation as the independent variable yielded a significant effect, $F(1, 27) = 4.15, p = .05, d = .78$. Participants entered a greater proportion of horizontal problems from left-to-right ($M = .62, SD = .48$) than vertical problems ($M = .28, SD = .43$). Of course, the right-to-left analysis is a mirror image of left-to-right. Thus, participants tended to enter horizontal problems from left-to-right and vertical problems from right-to-left.

Math performance was analyzed in the same one-way ANOVA with orientation as the independent variable. There was no reliable effect for percentage correct. The analysis of number of problems correctly solved yielded a significant effect for orientation, $F(1, 27) = 5.52, p < .05, d = .90$. As in Experiment 1, participants solved more problems presented in a vertical orientation ($M = 78.71, SD = 24.14$) than in a horizontal orientation ($M = 61.53, SD = 14.33$).

Ancillary Measure

The percentage of problems on which the participants reported using the traditional method was analyzed in a one-way ANOVA. This analysis yielded an effect for orientation, $F(1, 27) = 4.57, p < .05, d = .82$. Participants reported using the traditional method on a smaller proportion of horizontal problems ($M = .31, SD = .44$) than participants working on vertical problems ($M = .66, SD = .45$) (see Figure 3). In other words, when working on horizontal problems, participants reported using the method of adjustment on 69% of the problems, whereas when working on vertical problems, they reported using the traditional method on 66% of them.

Discussion
Trbovich and LeFevre (2003) argue that vertical problems are represented on the visuospatial sketchpad, and are solved using the same traditional, unit-to-decade algorithm that is used on paper. Consistent with this argument, I found that participants tended to enter the answers for vertical problems from right-to-left, and reported using the traditional method. Thus, they borrowed a ten from the decades column, performed the unit calculation, and entered the unit digit. They then performed the digit calculation and entered this value.

However, on horizontal problems, the participants tended to enter the numbers from left-to-right and reported that they used the method of adjustment, consistent with the argument that they calculated the entire answer and then entered it. For example, when presented $57 - 38$, they reported adding 2 to 38 (40), subtracted 40 from 57, and then added back the 2 to get the final answer, 19. Or they reported some variant of this approach in which they adjusted one of the numbers up or down, and then performed the subtraction, followed by taking into account the adjustment.\(^2\) This finding is consistent with Trbovich and LeFevre’s (2003) proposal that the participants’ solution approach on horizontal problems makes use of the phonological loop to store intermediate values, as suggested by Beilock et al. (2007). In terms of performance, replicating the findings in Experiment 1, participants solved more vertical problems correctly than horizontal problems.

As far as the mere effort account is concerned, the approach findings discussed above suggest that using the method of adjustment is the approach that most participants take to solving the horizontal subtraction problems. Given that this is the prepotent response, the mere effort account would argue that this tendency would be potentiated when female participants are subject to threat. I examined this hypothesis in the next experiment.

**Experiment 4**
To test the potentiation hypothesis, I used female participants in the same paradigm as in the previous experiment but with the addition of the stereotype threat manipulation. Thus, females were either threatened or not and then solved horizontal or vertical subtraction problems for which they could enter the answer either from left-to-right or right-to-left.

To the extent that using method of adjustment is the prepotent response for horizontal subtraction problems, I should find that threatened females enter the answers from left-to-right more than unthreatened participants. Once again, when using the method of adjustment, the approach yields the whole answer, which is then most easily entered from left-to-right. In contrast, when one uses the traditional method, the problem is worked from right-to-left, and the most efficient mode of entry would be from right-to-left. In this experiment as in Experiment 3, I primarily focused on the approach the participants take to solving the problems, rather than on performance. Once again, participants self-selected their approach to solving the problems, which could have impacted their task performance.

**Method**

**Participants**

Seventy-four female undergraduates from a northeastern university participated in the study for course credit.

**Procedure**

The procedure and measures were the same as those used in Experiment 3 with the addition of the stereotype threat manipulation used in Experiments 1 and 2. After participants completed the 5 practice problems, the stereotype threat manipulation was implemented. Following task completion, all participants filled out the same manipulation check items as those used in Experiments 1 and 2, and the same questions about the approaches they took to solving
the problems as those used in Experiment 3.

**Results**

**Manipulation Checks**

Responses to the question asking to what extent gender differences exist on the task and which gender performs better on the task were analyzed in two-way ANOVAs with threat (stereotype threat vs. no stereotype threat) and orientation (horizontal vs. vertical) as independent variables. Participants in the threat condition reported that gender differences existed to a greater extent ($M = 5.85, SD = 1.93$) than no stereotype threat participants ($M = 2.82, SD = 2.19$), $F(1, 68) = 39.13, p < .001, d = 1.52$. Stereotype threat participants also reported that males perform better on the task than females to a greater extent ($M = 4.39, SD = 1.73$) than no stereotype threat participants ($M = 6.03, SD = 1.13$), $F(1, 68) = 21.50, p < .001, d = 1.12$. These results indicated that the stereotype threat manipulation was successful.

**Task Performance**

Once again, data were excluded from analyses if a participant solved less than 75% of the problems correctly. Two participants failed to reach this criterion.

The proportion of digits entered from left-to-right was analyzed in a two-way ANOVA with threat and orientation as the independent variables. To test the hypothesis that threat would potentiate the tendency to enter digits from left-to-right (use the method of adjustment), I conducted a planned contrast (Kirk, 1995) of the proportion of horizontal problems entered from left-to-right for threat and no threat participants, which was significant, $F(1, 68) = 7.20, p < .05, d = .65$. Threat participants entered a greater proportion of problems from left-to-right ($M = .87, SD = .32$) than no threat participants ($M = .51, SD = .50$). The same contrast for vertical problems was not reliable, $p > .40$. Threatened participants entered problems from left-to-right to
the same extent ($M = .28, SD = .43$) as participants who were not threatened ($M = .31, SD = .46$).

The interaction suggested by this pattern of means was significant, $F (1, 68) = 3.88, p = .05, d = .48$. This analysis also revealed a main effect for orientation, $F (1, 68) = 15.57, p < .001, d = .96$.

As in Experiment 3, a greater proportion of horizontal problems was entered from left-to-right ($M = .71, SD = .44$) than vertical problems ($M = .29, SD = .44$).

Math performance was analyzed in the same two-way ANOVA with threat and orientation as the independent variables. There was no reliable effect for percentage correct. The analysis of number of problems correctly solved yielded a Threat x Orientation interaction, $F (1, 68) = 9.63, p < .01, d = .75$. Planned contrasts (Kirk, 1995) showed that threatened participants solved fewer horizontal problems ($M = 64.46, SD = 15.54$) than non-threatened participants ($M = 80.41, SD = 30.20$), $F (1, 68) = 4.92, p < .05, d = .54$. However, threatened participants solved more vertical problems ($M = 99.65, SD = 26.07$) than participants who were not threatened ($M = 82.75, SD = 14.69$), $F (1, 68) = 4.74, p < .05, d = .53$. There was also a main effect for orientation, $F (1, 68) = 12.57, p < .01, d = .86$, as a result of the fact that participants correctly solved more vertical problems ($M = 91.46, SD = 22.68$) than horizontal problems ($M = 71.41, SD = 24.12$).

**Ancillary Measure**

The percentage of problems on which the participants reported using the traditional method was analyzed in a two-way ANOVA. A planned contrast (Kirk, 1995) of the proportion of horizontal problems on which participants reported using the traditional method for threat and no threat participants was significant, $F (1, 68) = 6.86, p < .05, d = .64$. As seen in Figure 4, threat participants reported using the traditional method on a smaller proportion of horizontal problems ($M = .18, SD = .32$) than no threat participants ($M = .51, SD = .45$). The same contrast
for vertical problems was not reliable, $p > .40$. Threatened participants reported using the traditional method to the same extent on vertical problems ($M = .74, SD = .40$) as participants who were not threatened ($M = .71, SD = .42$). The interaction suggested by this pattern of means was significant, $F(1, 68) = 3.89$, $p = .05$, $d = .48$. This analysis also revealed a main effect for orientation, $F(1, 68) = 16.64$, $p < .001$, $d = .99$. As in Experiment 3, participants who solved horizontal problems reported using the traditional method less ($M = .32, SD = .41$) than participants who worked on vertical problems ($M = .72, SD = .40$).

**Discussion**

Replicating Experiment 3, and consistent with the argument that using the method of adjustment is the prepotent approach to solving horizontal subtraction problems, I found that participants tended to work from left-to-right using the method of adjustment on horizontal problems, but to work from right-to-left using the traditional method on vertical problems.

Consistent with the mere effort account, I found that the tendency to use the method of adjustment was potentiated for participants under threat. I also found that threatened participants solved fewer horizontal subtraction problems correctly, but more vertical problems correctly than no threat controls.

These findings are consistent with the notion that the method of adjustment is the prepotent response to horizontal subtraction problems, and show that stereotype threat potentiates this response.
Chapter 4: Stereotype Threat and Subtraction

As stated above, the core of Beilock et al.’s (2007) argument is that threat debilitates performance on the modular horizontal problems because solving the subtraction portion requires phonological resources (e.g., remembering intermediate values) that are taken up by the thoughts and worries produced by threat. In Experiment 5, I directly pitted the mere effort account against this working memory account.

Experiment 5

Females were randomly assigned to either a stereotype threat or no threat condition. Crossed with this manipulation, participants were asked to solve two digit subtraction problems that were presented either vertically or horizontally on a computer screen. Crossed with these two manipulations, participants were required to enter the answers to these problems either from right-to-left or from left-to-right.

In this paradigm, Beilock et al.’s (2007) working memory account would predict that, on horizontal problems, allowing participants to enter their answers from right-to-left could reduce the debilitation produced by stereotype threat. Participants do not need to use the phonological loop for remembering the intermediate value (Imbo, Vandierendonck, & Vergauwe, 2007). They can simply perform the units subtraction and enter the value (e.g., 57 - 38 = 9). As a result, the fact that the phonological loop is occupied by concerns about fulfilling the threat should have diminished or no impact on performance (see Figure 5). However, this shortcut is not available when answers are entered from left-to-right. Because the whole answer is entered, the phonological loop is needed to keep track of the intermediate values. According to Beilock et al.’s (2007) account, thoughts about fulfilling the stereotype should interfere with remembering these values, resulting in debilitated performance (see Figure 5). Supporting this claim, Hitch
(1978) found that when participants used the traditional method, writing down the answer in reverse order (i.e., left-to-right) produced more errors than writing the answer in the same order (i.e., right-to-left) due to the delay in writing down some of the digits.

The mere effort account would predict that on horizontal problems, stereotype threat potentiates the prepotent response of using the method of adjustment. As seen in Experiments 3 and 4, entering the answer from left-to-right is consistent with the method of adjustment. Thus, the fact that the threat participants are motivated to want to perform well should lead them to outperform no threat participants on horizontal problems when they enter the answers from left-to-right (see Figure 5). On the other hand, using the method of adjustment should interfere and debilitate performance when the answer is entered from right-to-left. One must remember the adjustments to reach the final answer and then take the right hand digit of the answer and enter it followed by the left hand digit. Therefore, the potentiated use of the method of adjustment should lead threatened participants to perform more poorly than no threat participants when they enter the answers from right-to-left (see Figure 5).

To the extent that participants rely more heavily on visuospatial resources to solve vertically presented problems, verbal working memory resources (i.e., the phonological loop) are not necessary (Beilock et al., 2007). As a result, the working memory account would not make a strong prediction for vertical problems in the current experimental paradigm. However, it would be consistent with a motivational account, like mere effort, to find that threatened participants perform better on vertical problems than no threat participants, as was found on modular vertical problems in Experiment 1 and the vertical subtraction problems in Experiments 3 and 4.

**Method**

**Participants**
One hundred and sixteen female undergraduates from a northeastern university participated in the study for course credit.

**Procedure**

The same 160 subtraction problems used in Experiments 3 and 4 were used in the current experiment. As in Experiments 3 and 4, for the participants who saw horizontal problems, two boxes were presented to the right of the problem (e.g., $57 - 28 = \square \square$). For participants who saw vertical problems, the boxes were beneath the numbers.

The procedures were identical to those used in Experiment 4 with one exception. Participants were told to enter their answers from left-to-right or from right-to-left. In order to manipulate the direction in which participants had to enter their response, the computer cursor was programmed to always be on the appropriate box. For example, if participants were asked to enter their answer from right-to-left, then the cursor was always on the right box and the participant had to enter that digit before it moved on to the left box. Those participants required to enter their answers from left-to-right were told the following:

> The math problems you are about to solve are all subtraction problems that require two-digit answers. There will be two boxes where you will enter your two-digit answer. You are to enter the left-hand digit first, followed by the right-hand digit. The program automatically goes to the next box and you do not need to click on it.

The right-to-left manipulation was identical to the left-to-right manipulation except it read, “You are to enter the right-hand digit first, followed by the left-hand digit.”

Participants were given 5 practice problems to become familiar with the task. After the practice problems, the same stereotype threat manipulation as the one used in Experiments 1, 2, and 4 was implemented.
After completing the task, participants responded to the same threat manipulation check items that were used in Experiments 1, 2, and 4, and answered the same questions about the approaches they took to solving the problems as were used in Experiments 3 and 4.

**Results**

**Manipulation Checks**

Responses to the questions asking to what extent gender differences exist on the task and which gender performs better on the task were analyzed in a 2 (threat condition: stereotype threat vs. no stereotype threat) x 2 (problem orientation: horizontal vs. vertical) x 2 (left-to-right vs. right-to-left) between subjects ANOVA. Main effects for stereotype threat condition were found for both questions but no other reliable effects were found. Participants subject to stereotype threat reported that gender differences existed to a greater extent ($M = 6.37$, $SD = 2.44$) than no stereotype threat participants ($M = 2.43$, $SD = 1.98$), $F(1, 106) = 88.33$, $p < .001$, $d = 1.83$. Stereotype threat participants also reported that males perform better on the task than females to a greater extent ($M = 4.55$, $SD = 1.94$) than no stereotype threat participants ($M = 5.98$, $SD = .96$), $F(1, 106) = 22.20$, $p < .001$, $d = .92$. These results indicate that the stereotype threat manipulation used was successful. Participants in the threat condition were aware of the negative group stereotype, and females were expected to perform more poorly than males.

**Task Performance**

Consistent with Beilock et al. ’s (2007) practice, two participants were excluded because they solved less than 75% of the problems correctly.

Math performance was analyzed in a 2 (threat condition: stereotype threat vs. no stereotype threat) x 2 (problem orientation: horizontal vs. vertical) x 2 (direction: left-to-right vs. right-to-left) between subjects ANOVA. There were no reliable effects for percentage correct
and, thus, I analyzed number of problems correct. Results revealed a three-way interaction, \( F(1, 106) = 5.21, p < .05, d = .44. \) To decompose the three-way interaction, two two-way ANOVAs were conducted, one for horizontal problems and one for vertical problems.

Analysis of the horizontal problems in a 2 (threat vs. no threat) x 2 (left-to-right vs. right-to-left) ANOVA revealed only a Threat x Direction interaction, \( F(1, 106) = 9.29, p < .01, d = .59. \) As seen in Figure 6, planned contrasts (Kirk, 1995) showed that threatened participants solved more horizontal problems correctly (\( M = 76.61, SD = 24.17 \)) than non-threatened participants (\( M = 59.85, SD = 14.49 \)) when the problems were entered from left-to-right, \( F(1, 106) = 4.38, p < .05, d = .41. \) However, non-threatened participants solved more problems correctly (\( M = 83.79, SD = 23.26 \)) than threatened participants (\( M = 65.67, SD = 21.45 \)) when they were entered from right-to-left, \( F(1, 106) = 4.91, p < .05, d = .43. \)

Analysis of the vertical problems in a 2 (threat vs. no threat) x 2 (left-to-right vs. right-to-left) ANOVA revealed only a main effect for threat, \( F(1, 106) = 6.30, p < .05, d = .49. \)Participants subject to stereotype threat solved more vertical problems correctly (\( M = 93.74, SD = 24.51 \)) than participants who were not subject to threat (\( M = 78.63, SD = 19.48 \)) (see Figure 7).

**Ancillary Measure**

The proportion of problems on which the participants reported using the traditional approach (units-to-decades) to solving these problems (i.e., work them from right-to-left) was analyzed in a 2 (threat) x 2 (orientation) x 2 (direction) ANOVA, which produced three main effects. Consistent with Experiment 4, threatened participants reported using the traditional method on a smaller proportion of problems (\( M = .34, SD = .44 \)) than non-threatened participants (\( M = .60, SD = .43 \)), \( F(1, 106) = 9.93, p < .05, d = .61. \) Also consistent with Experiments 3 and 4, participants working on horizontal problems reported using the traditional method on a smaller
proportion of problems \((M = .34, SD = .44)\) than participants who worked on vertical problems \((M = .60, SD = .46)\), \(F(1, 106) = 10.07, p < .05, d = .62\). And participants reported using the traditional method on a smaller proportion of problems entered from left-to-right \((M = .36, SD = .42)\) than participants who entered their answers from right-to-left \((M = .57, SD = .47)\), \(F(1, 106) = 6.47, p < .05, d = .49\).

**Discussion**

In Experiment 5, threatened and non-threatened participants were required to enter the answers to horizontal or vertical subtraction problems from left-to-right or right-to-left. According to Beilock et al.’s (2007) working memory account, entering the answers to horizontal subtraction problems from right-to-left could represent a benefit to the threatened participants because the units digit can simply be entered and does not have to be remembered. As a result, it need not compete with thoughts about poor performance for space on the phonological loop. However, I found that threatened participants performed more poorly than no threat control participants when entering the answers from right-to-left.

When horizontal subtraction problems are entered from left-to-right, one must remember intermediate values. As a result, Beilock et al.’s (2007) working memory account would predict that threatened participants would perform more poorly than non-threatened participants because these intermediate values are rehearsed on the phonological loop, which is also occupied by the negative thoughts produced by the threat manipulation. Instead, the results showed that threatened participants outperformed participants who were not under threat. These findings do not support Beilock et al.’s (2007) working memory account.

Experiment 3 suggested the method of adjustment as the prepotent response to horizontal subtraction problems. Experiment 4 showed that stereotype threat potentiates this response.
Requiring the participants to enter the numbers from left-to-right is consistent with the method of adjustment, because this approach produces the whole answer. As a result, the fact that the threat participants are motivated to perform well should lead them to outperform control participants on horizontal problems when they enter the problems from left-to-right. On the other hand, using the method of adjustment should interfere with performance when the numbers must be entered from right-to-left. One must remember the adjustments to reach the final answer and then take the right hand number of that answer and enter it followed by the left hand number. Thus, the potentiated use of the method of adjustment should lead threatened participants to perform more poorly than control participants when answers must be entered from right-to-left. Of course, this is exactly what was found in Experiment 5.

The working memory account does not make strong predictions for performance on the vertical problems because Beilock et al. (2007) argue that these problems are primarily solved with the visuospatial sketchpad, and would be minimally impacted by the worries occupying the phonological loop. However, consistent with a motivational account, like mere effort, and also with the findings for vertical modular arithmetic problems in Experiments 1 and vertical subtraction problems in Experiment 4, I found that threat participants outperformed no threat controls on vertical problems whether entered from left-to-right or right-to-left.

Overall, Experiment 5 provided support for a motivational account of the effects of stereotype threat on both horizontal and vertical subtraction performance. The working memory account only predicted performance on horizontal problems (i.e., when phonological resources were supposed to be depleted), which was not supported.
Chapter 5: General Discussion

In recent years, researchers have focused on the mechanisms involved in stereotype threat performance effects (e.g., Beilock et al., 2007; Ben-Zeev et al., 2005; Blascovich et al., 2001; Bosson et al., 2004; Cadinu et al., 2003; Croizet et al., 2004; Jamieson & Harkins, 2007, 2009; O’Brien & Crandall, 2003; Schmader & Johns, 2003; Seitchik, Jamieson, & Harkins, in press). Stereotype threat has repeatedly been found to hinder stigmatized individual’s performance (e.g., Aronson et al., 2002; Beilock et al., 2006; Beilock et al., 2007; Blascovich et al.; Ben-Zeev et al., 2005; Brown & Josephs, 1999; Brown & Pinel, 2003; Davies et al., 2002; Jamieson & Harkins, 2007; Johns et al., 2005; Keller & Dauenheimer, 2003; O’Brien & Crandall; Pronin et al., 2004; Schmader & Johns; Sekaquaptewa & Thompson, 2003; Seitchik et al.; Spencer et al., 1999). However, there is no agreement about the mechanism(s) responsible for this debilitation in performance due to stereotype threat. The current experiments examined the most widely accepted explanation, working memory, against a motivational explanation, mere effort.

One mechanism proposed to account for stereotype threat effects is working memory interference. According to the working memory account, concern over confirming the stereotype produces thoughts and worries that utilize cognitive resources that could be devoted to task performance and it is this reduction in working memory capacity that produces the performance debilitation (e.g., Schmader, 2002; Schmader & Johns, 2003). Using Baddeley’s (1986; 2000; Baddeley & Logie, 1999) multicomponent model of working memory, Beilock et al. (2007) found support for this account by showing exactly how stereotype threat undermines performance on horizontal modular arithmetic problems. Beilock et al. (2007) argue that the thoughts and worries produced by stereotype threat occupy the phonological loop, which is also required to store the intermediate values that are produced when participants solve horizontal
subtraction problems (Trbovich & LeFevre, 2003). It is this reduced capacity that debilitates performance.

However, Beilock et al.’s (2007) findings do not rule out an alternative account, mere effort (Harkins, 2006). According to the mere effort account, stereotype threat motivates stigmatized individuals to not confirm the stereotype, which potentiates the prepotent response on the given task (Jamieson & Harkins, 2007; 2009; Seitchik et al., in press). If the prepotent response is correct, stereotype threat facilitates performance for stigmatized individuals. 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 have the knowledge that their prepotent tendencies are incorrect and are given the opportunity to correct them, performance can be facilitated (Jamieson & Harkins, 2007; 2009). Therefore, the mere effort account proposes that motivation plays a direct role in producing stereotype threat performance effects. The current research focused on this motivational account in the context of Beilock et al.’s (2007) working memory interpretation of the subtraction component of modular arithmetic problems.

In Experiment 1, I replicated Beilock et al.’s (2007) finding that stereotype threat led to poorer performance on horizontal modular arithmetic problems, but in a paradigm that was sensitive to motivational effects. I also found that threatened participants outperformed non-threatened participants on vertical modular arithmetic, an effect that is consistent with a motivational account, like mere effort, but that would not be predicted by a working memory account.

In Experiment 2, I looked at the effect of threat on the performance of the division component of the modular arithmetic problems. Beilock et al.’s (2007) account would not predict
that stereotype threat would debilitate performance on these simple division problems because solving them does not require the phonological loop, and I did not find debilitation. Instead, threatened participants solved more division problems than non-threatened participants, a finding that is, once again, consistent with a motivational account.

Beilock et al.’s (2007) explanation focuses on performance of the subtraction component of the modular arithmetic problems, and so, I looked at performance on horizontal and vertical subtraction problems in the remaining experiments. In Experiment 3, in an effort to identify the prepotent response I examined the approaches participants take to solving horizontal and vertical subtraction problems. Participants were presented subtraction problems either horizontally or vertically and were allowed to enter their answers from left-to-right or from right-to-left with no stereotype threat manipulation. Research by Trbovich and LeFevre (2003) suggests that problems presented in a horizontal format depend heavily on phonological resources as a result of solution procedures that require the verbal maintenance of intermediate steps in memory. For example, faced with a problem like 64 – 37, participants may adjust the 37 to 40, subtract 40 from 64 (24) and then add in the adjustment, 24 + 3 = 27. I termed this approach the method of adjustment. If participants are using this approach on horizontal problems, it should yield the whole answer, and participants should tend to enter the answers from left-to-right.

In contrast, Trbovich and LeFevre (2003) suggested that “math problems presented in a vertical format rely more heavily on visuospatial resources as individuals tend to solve vertical problems in a spatial mental work space similar to how such problems are solved on paper” (Beilock et al., 2007, p. 260). This approach is termed the traditional or units-to-decades approach. Participants using this approach should enter the units digit first followed by the decade digit (i.e., right-to-left).
Consistent with Trbovich and LeFevre’s (2003) analysis, in Experiment 3, I found that participants tended to enter answers to the horizontal problems from left-to-right and reported using the method of adjustment. On vertical problems, participants tended to enter the answers from right-to-left and reported using the traditional method. These results suggest that the method of adjustment is the prepotent response to horizontal subtraction problems. The mere effort account would suggest that stereotype threat should potentiate this response, a hypothesis that I tested in Experiment 4.

In Experiment 4, participants went through the same procedures as those used in Experiment 3 with the exception that the stereotype threat manipulation was implemented. I found that, as proposed by the mere effort account, the tendency to use the method of adjustment was potentiated by stereotype threat: threatened participants entered a greater proportion of answers from left-to-right than non-threatened participants. In addition, participants under stereotype threat reported using the method of adjustment on around 80% of horizontal subtraction problems, whereas participants not under stereotype threat reported using it on about 50% of horizontal problems. I also found that performance on horizontal subtraction problems was debilitated by stereotype threat, but, once again, performance on vertical problems was facilitated. These results are consistent with the argument that the method of adjustment is the prepotent response on horizontal subtraction problems, and that this response is potentiated by stereotype threat.

In Experiment 5, I pitted this motivational account against the working memory account. I required participants to enter the answers from left-to-right or from right-to-left. When number entry is from left-to-right, the working memory account makes a strong prediction. In this case, the participants must use the phonological loop to rehearse intermediate values, which, for
threatened participants, should be occupied by negative thoughts about the potential for failure on the task, leading to debilitated performance. However, I did not find debilitated performance. In fact, threatened participants solved more problems than non-threatened participants. When digit entry is from right-to-left, the working memory account would not predict debilitation. In fact, the phonological load should be diminished because the participants can enter the units digit rather than having to remember it (e.g., Hitch, 1978; Imbo, Vandierendonck, & De Rammelaere, 2007; Imbo et al., 2007). However, I found debilitation.

In contrast, if, as I propose, the method of adjustment is the prepotent response to horizontal problems, the mere effort account would predict that stereotype threat would potentiate this response. Entering the numbers from left-to-right is consistent with the method of adjustment. As a result, the fact that the threat participants are motivated to perform well should lead them to outperform control participants on horizontal problems when they enter the problems from left-to-right, and they do. On the other hand, using the method of adjustment should interfere with performance when the numbers are entered from right-to-left. One must remember the adjustments to reach the final answer and then take the right hand number of that answer and enter it followed by the left hand number. Thus, the potentiated use of the method of adjustment should lead threatened participants to perform more poorly than control participants when answers are entered from right-to-left, and they do.

To the extent that participants use visuospatial resources to solve vertically presented problems, verbal resources are not needed as much (Beilock et al., 2007). Consequently, the left-to-right versus right-to-left manipulation should not make a difference. As a result, the fact that threatened participants are motivated to perform well should lead them to outperform control participants, as I also found.
Taken together, the findings from this line of work are consistent with the mere effort account but not with Beilock et al.’s (2007) working memory account of performance on horizontal subtraction problems. However, these findings may not explain why threatened participants perform more poorly than non-threatened participants on Beilock et al.’s (2007) horizontal modular arithmetic problems. In both Experiments 4 and 5 there were conditions under which non-threatened participants outperformed threatened participants on horizontal problems as is found on horizontal modular arithmetic problems. But, in each case, a structural advantage favored the non-threatened participants that is not available when working modular arithmetic problems.

For example, in Experiment 4, I found that threatened participants performed more poorly on the horizontal arithmetic problems than non-threatened participants. But, in this experiment, participants could choose to enter the numbers from right-to-left or left-to-right, and I found that threatened participants had a stronger tendency to enter them from left-to-right than non-threatened participants. The fact that non-threatened participants were more likely to work the problems from right-to-left gave them an advantage because they could enter the units number as soon as they calculated it, eliminating the need to remember it at all. In Experiment 5, non-threatened participants outperformed threatened participants when they were required to enter the problems from right-to-left. In this case, the non-threatened participants had the same advantage as in Experiment 4. That is, they could just perform the units subtraction and enter the value in the right-hand box.

However, this shortcut is not available to the non-threatened participants when they solve horizontal modular arithmetic problems. When they use the traditional, units-to-decade method, working from right-to-left, they must remember the units digit when they move to the decade
calculation because they cannot enter the units digit. Even so, this method may be more efficient than the method of adjustment. For example, when using the traditional method, the steps may be more easily routinized than is the case with the method of adjustment. One subtracts the units digits, borrowing a ten (e.g., \(57 - 28; 17 - 8 = 9\)), and then decrements the decade minuend by one and subtracts the subtrahend (\(4 - 2 = 2\)), yielding the answer 29. To use this method effectively, one need only know the subtraction facts for numbers under 20, and then remember the units digit while decrementing the decades value. Using the method of adjustment, for each problem one must determine how much to increment the subtrahend (\(57 - 28; \text{plus 2, 28 to 30}\)), then subtract this value from the minuend (\(57 - 30 = 27\)), and then add the adjustment (\(27 + 2 = 29\)). The amount to adjust the subtrahend will change with each problem, as will the size of the difference between the minuend and the adjusted subtrahend. Finally, one must remember to make the adjustment, as well as remember what it is. Thus, it is possible that the traditional method is more efficient than the method of adjustment even without the advantage conferred by the opportunity to enter the units digit first. If this were the case, because the non-threatened participants use the traditional method more, it could help to account for why they perform better than threatened participants on horizontal modular problems. In a future study, this possible explanation for why performance differs in modular arithmetic should be explored.

If further research suggests that the approach does cause the performance gap between threatened and non-threatened participants on horizontal modular arithmetic problems, then an intervention strategy can be used to overcome this performance gap. In previous research, my lab has found that when participants are subject to stereotype threat and have the opportunity and knowledge necessary to make the appropriate response, the debilitation effect is eliminated (e.g., Jamieson & Harkins, 2009) or reversed (e.g., Jamieson & Harkins, 2007). For example, Jamieson
and Harkins (2009) found that the difference in performance between threatened and non-threatened participants on GRE comparison problems was eliminated when a solution strategy for solving comparison problems (use logic and/or estimation) was suggested. In an experiment where all participants are required to use the same approach (method of adjustment, traditional method), holding constant the method used to solve the horizontal arithmetic problems may not only eliminate the debilitation effect but reverse it. That is, given that the participants take the same approach, the fact that threatened participants are more motivated than non-threatened participants may lead them to outperform the non-threatened participants on these relatively simple problems, as has been found on the antisaccade task (Jamieson & Harkins, 2007) and the Stroop (McFall et al., 2009; Jamieson & Harkins, 2011). This possible intervention will also need to be tested in future research.

Overall, the present experiments supported a motivational, mere effort, explanation for stereotype threat performance effects. The current and previous research (e.g., Jamieson & Harkins, 2007; 2009; Kray, Thompson, & Galinsky, 2001) suggests that females subject to threat do not simply accept their fate and resign themselves to performing poorly. Although they do temper expectations based on their awareness of stereotypes (e.g., Stone & McWhinnie, 2008), females also try very hard to disprove negative stereotypes directed at their group. Thus, the motivational processes argued for in this research offer a source of encouragement because the findings suggest that females react against being negatively typecast as poor performers in math and science.

Determining the mechanism(s) responsible for stereotype threat’s influence on performance has practical implications. More specifically, stereotype threat in the domain of mathematics has been shown to reduce interest in science, technology, engineering, and
mathematics (STEM) careers (e.g., Shapiro & Williams, 2012; Shapiro, Williams, & Hambarchyan, 2013; Smeding, 2012). This reduction in performance and interest has influenced the gender gap in STEM-related majors and careers. In order to reduce this gender difference, it is important to understand how stereotype threat influences performance and create interventions to reduce this performance deficit. These interventions may then assist in reducing the gender gap in STEM-related fields.

For example, Jamieson and Harkins (2009, Experiment 2) attempted to improve the performance of stereotype threat participants on GRE math problems by providing instructions designed to reduce their reliance on the solving approach (i.e., the prepotent response on GRE math problems) 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. A mediation analysis showed that threat participants’ improvement on comparison problems resulted from the fact that threatened females took the instructions into account and reduced their reliance on the solving approach when they tried to solve comparison problems. Stereotype threat participants are motivated to perform well, and as long as they have the requisite ability and the opportunity to make the response, simple instructions are enough to make up for the fact that participants do not know that the prepotent response is incorrect. Increasing performance may then serve as positive reinforcement in STEM courses and increase interest in STEM careers. Future research should focus on more motivational interventions to overcome stereotype threat performance effects.
References


Footnotes

1. In Experiment 1, as in previous stereotype threat research, I found that threat did not affect the performance of male participants. Given this finding, in Experiment 2 and in the subsequent experiments I used only female participants.

2. Although participants sometimes described other methods (e.g., go up by tens from the lower number until you get within units of the upper number and then adjust up or down) or shortcuts that could be used on particular problems (e.g., true if the divisor is five and the unit digit after subtraction equals five or zero, and false otherwise), the great majority of the participants described using the traditional method and/or the method of adjustment. As a result, I focused on these methods.
Figure 1. Experiment 1: Mean number of problems correct as a function of gender, stereotype threat condition, and problem orientation. Note: Error bars = +/- standard error of the mean.
Figure 2. Experiment 2: Mean number of division problems correct as a function of stereotype threat condition.
Note: Error bars = +/- standard error of the mean.
Figure 3. Experiment 3: Mean percent of problems participants reported using the traditional approach as a function of problem orientation. Note: Error bars = +/- standard error of the mean.
Figure 4. Experiment 4: Mean percent of problems participants reported using the traditional approach as a function of problem orientation and stereotype threat condition. Note: Error bars = +/- standard error of the mean.
Figure 5. Experiment 5: Predictions for the working memory and mere effort accounts for horizontal subtraction performance when participants were forced to enter their answers from left-to-right or right-to-left.
Figure 6. Experiment 5: Mean number of problems correct on Horizontal Problems as a function of stereotype threat condition and direction of answer entry. Note: Error bars = +/- standard error of the mean.
Figure 7. Experiment 5: Mean number of problems correct on Vertical Problems as a function of stereotype threat condition and direction of answer entry. Note: Error bars = +/- standard error of the mean.