ELICITING STEREOTYPE CHALLENGE AND STEREOTYPE THREAT EFFECTS
WITHIN THE CONTEXT OF WOMEN’S MATH PERFORMANCE

A thesis presented to
the faculty of
the College of Arts and Sciences of Ohio University

In partial fulfillment
of the requirements for the degree
Master of Science

Ronald A. Elizaga
November 2005
This thesis entitled

ELICITING STEREOTYPE CHALLENGE AND STEREOTYPE THREAT EFFECTS
WITHIN THE CONTEXT OF WOMEN’S MATH PERFORMANCE

by

Ronald A. Elizaga

has been approved for

the Department of Psychology

and the College of Arts and Sciences by

Keith D. Markman

Assistant Professor of Psychology

Benjamin M. Ogles

Interim Dean, College of Arts and Sciences
The present study examined stereotype challenge and threat responses through a manipulation of social comparison information. Female participants received feedback about an ostensibly matched partner who was either similar (i.e., female) or dissimilar (i.e., male) to themselves. After learning that they would be taking a math test that either has or has not demonstrated gender differences in the past, participants learned that their partner either performed very well or very poorly on the test. Subsequently, participants took the math test themselves, and their performance was evaluated. It was predicted that interacting with a peer whose performance was consistent with the negative stereotype would exacerbate stereotype threat effects, while interacting with a peer whose performance challenged the negative stereotype would eliminate stereotype threat effects. The results support these predictions, highlighting the manner whereby peer-group social comparisons can exert a powerful influence on women’s math test performance when negative stereotypes are salient.
Dedicated to my mom and my wife for their unending love and support,
and my dad for his angelic guidance.
Acknowledgments

I wish to express my extreme gratitude to Dr. Keith Markman for his patience and invaluable guidance through all stages of the research project. In addition, I wish to extend my thanks to the members of my thesis committee for their constructive feedback: Drs. Keith Markman (Chair), Mark Alicke, Dan Lassiter, and Kathi Heffner. Thanks also go to the many undergraduate research assistants that made this project possible: Jessie Wycoff, Jenny Janda, Shane Lantz, Cody Carson, and Nick Stewart.
Table of Contents

Abstract ...........................................................................................................................................3

Dedication .........................................................................................................................................4

Acknowledgments ............................................................................................................................5

List of Tables ......................................................................................................................................9

Overview ........................................................................................................................................10

Stereotype Threat ............................................................................................................................14

Empirical Support for Stereotype Threat Effects ............................................................................15

Moderation of Stereotype Threat ......................................................................................................18

Task difficulty ....................................................................................................................................18

Group identification .........................................................................................................................19

Domain identification .......................................................................................................................19

Stigma consciousness .......................................................................................................................19

Solo status ..........................................................................................................................................20

Mediation of Stereotype Threat ........................................................................................................20

Anxiety and evaluation apprehension ........................................................................................……20

Performance expectancies ..................................................................................................................21

Arousal ...............................................................................................................................................22

Reduced working memory capacity ..................................................................................................22

Effort ..................................................................................................................................................23
### Coping with the Consequences of Stereotype Threat

- **Self-handicapping** ....................................................... 24
- **Disidentification** .......................................................... 24
- **Domain avoidance** ....................................................... 25
- **Self-affirmation** .......................................................... 25
- **Identity bifurcation** ..................................................... 26

### Challenge and Threat Appraisals

- **Primary and secondary appraisals** ................................. 27
- **Evidence for challenge and threat appraisals** ...................... 28
- **Social facilitation as challenge and threat** .......................... 30

### Goal of the Present Research

- **Overview of Study** .......................................................... 31
- **Hypotheses and Expected Results** ..................................... 31

### Method

- **Participants** ................................................................. 34
- **Stimulus Material** .......................................................... 34
  - **Challenge-threat manipulation check** .............................. 34
  - **Math task** .................................................................. 35
- **Procedure** ..................................................................... 35
  - **Stereotype salience manipulation** ................................... 35
  - **Exemplar gender and exemplar performance manipulations** ........ 36
Results ...........................................................................................................................................37
Discussion ......................................................................................................................................40
References ....................................................................................................................................48
Appendix: Challenge-Threat Manipulation Check ........................................................................60
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Number of items answered correctly as a function of stereotype salience, exemplar gender, and exemplar performance</td>
<td>38</td>
</tr>
<tr>
<td>2.</td>
<td>Accuracy as a function of stereotype salience, exemplar gender, and exemplar performance</td>
<td>40</td>
</tr>
<tr>
<td>3.</td>
<td>Overall percentage as a function of stereotype salience, exemplar gender, and exemplar performance</td>
<td>40</td>
</tr>
</tbody>
</table>
Overview

In 1992, the American Association of University Women (AAUW) released a report entitled “How Schools Shortchange Girls.” In their report focusing on grade school- to high school-age children, the AAUW claimed that schools were damaging girls’ self-esteem, and that teachers were paying more attention to boys. According to the report, schools were discouraging girls from taking certain courses and thus handicapping them in areas in which they would need to compete in the future. In particular, they focused on how the educational system was not meeting girls’ needs, and how upon graduation from high school girls had fallen behind their male cohorts in key areas such as math and science. This was soon followed by a substantial increase in school-based gender-equity programs, as well as the appearance of numerous articles in various media outlets written about the maltreatment of girls in classrooms. It had been a long standing assumption that girls were inferior to men in math and science, but the AAUW report had instead portrayed girls as being victims of schools and their policies. However, this idea was not without its opponents.

In 2000, a report was filed by the Secretary of Education through the National Center for Education Statistics (NCES) in response to a Congressional mandate to examine the status of educational equity for girls and women in the United States. The report examined the extent to which males and females have access to the same educational opportunities, benefit from these opportunities, and succeed at the same rate.

---

1 The request by Congress was made under the Women’s Educational Equity Act provisions (§Title V, Part B) of the Elementary and Secondary Education Act of 1965 (as amended in 1994).
In contrast to the AAUW’s claims, the data presented by the NCES showed that in high schools and colleges females are performing academically just as well as males, if not better, and the gap in achievement that may have once existed between males and females has significantly decreased, if not been eliminated entirely.

Another blow to the AAUW came in the form of a review (Kleinfeld, 1998) which found little evidence that girls have lower self-esteem than boys, or that boys receive more attention in the classroom. Rather, it was argued, boys and girls start school on similar footing in terms of opportunity, girls have fewer problems in the early grades than do boys, and females consistently outperform males in reading and writing (NCES, 2000). Importantly, however, neither the AAUW nor its opponents could speak to the gap that still remains between males and females in the domains of mathematics and science achievement.

Arguably, expertise in science and math has become increasingly important in this day and age of technological advancement and complexity. Fortunately for women, the NCES reports that the gender gap in science proficiency has narrowed in recent years as students get older. Not so for mathematics, however. Between 1973 and 1994, gender differences in math scores did not exist for 9- and 13-year-olds (NCES, 2000). In fact, beginning only in the 1990’s did achievement scores for males in these age ranges exceed that of females. Among 17-year-olds, however, males have consistently achieved higher average math proficiency scores. The AAUW claimed that females’ lower math achievement was the result of schools discouraging women from taking math courses (AAUW, 1992). Contrary to this claim, the NCES report showed that, overall, females
and males take similarly challenging mathematic programs. Nevertheless, they found that male and female views about mathematics diverge as the students grow older. By the 12th grade, females as compared to males are less likely to enjoy math and to think that they are good at math. According to the data provided by the NCES, women have educational opportunities equal to men, but still do not succeed at the same rate. Moreover, achievement gaps in mathematics between men and women appear to be more related to attitudes than to course taking. This trend in mathematical inequity has helped fuel the development of the stereotype that women are inferior to men in mathematical ability.

Ironically, the AAUW’s efforts to point out the (now discredited) disadvantage women have in math achievement, along with the resulting media focus, appear to have contributed to the proliferation of the stereotype throughout society, even among those who do not endorse it. Knowledge of the stereotype is especially acquired by people who are targets of the stereotype, and when such individuals find themselves in situations where the stereotype is relevant, they run the risk of behaving in a manner that will result in their being evaluated in light of the stereotype (Spencer, Steele, & Quinn, 1999). Increased evaluation apprehension experienced by individuals in a stereotyped domain has been termed *stereotype threat* (Steele & Aronson, 1995). According to Steele and his colleagues, the activation of a negative stereotype prior to a stigmatized individual’s performance on a task can create enough anxiety (i.e., threat) to impair performance (e.g., Spencer, et al., 1999; Steele & Aronson, 1995). For instance, Spencer et al. (1999) found that women tend to underperform on math tests relative to men when they felt that they
were at risk of confirming the existing cultural stereotype regarding women’s math ability and thereby be evaluated negatively. Importantly, then, stereotype threat has been cited as a potential contributor to the gender gap between males and females in math achievement.

In a separate literature, Folkman and Lazarus (e.g., Folkman & Lazarus, 1985; Lazarus & Folkman, 1984) have highlighted the importance of cognitive appraisal processes for coping with stressful stimuli. Their Transactional Model of Stress and Coping (Folkman & Lazarus, 1985) emphasized that two kinds of stress appraisals are critical to performance: *threat* appraisals, in which the assessment of danger exceeds the assessment of abilities or resources to cope with the stressor, and *challenge* appraisals, in which the assessment of danger does not exceed the assessment of abilities or resources to cope.

The broader goal of this proposal is to integrate the stereotype threat and coping literatures in a manner that has not been previously attempted. Although past research has focused on stereotype threat effects, the possibility that stereotype *challenge* effects might also occur – whereby the activation of a negative stereotype leads to enhanced performance – has not been examined. Moreover, although recent research has extended the findings of Folkman and Lazurus by investigating the consequences of threat and challenge appraisals in a variety of performance situations (Drach-Zahavy & Erez, 2002; Haslam, Jetten, O’Brien, & Jacobs, 2004; Mendes, Blascovich, Major, & Seery, 2001; Tomaka & Blascovich, 1994; Tomaka, Blascovich, Kelsey, & Leitten, 1993; Tomaka,
Palacios, & Lovegrove, 1995), none have exclusively focused on stereotype threat as a potential stressor.

Overall, I intend to demonstrate both threat and challenge effects within the context of women who have been reminded of the stereotype that they are inferior to men in math ability. In so doing, I hope to gain a better understanding of the processes that contribute to the undermining of women’s performance in math domains, and elucidate means by which their performance can be improved. Two studies have been designed to demonstrate that reminding individuals of their stigmatized status in a relevant performance domain can serve to both impair and enhance effort and/or performance in that domain depending upon how the situation is appraised.

The literature review begins by providing evidence for stereotype threat effects on different targets within numerous domains, before focusing more specifically on women in the math domain. Explorations into the possible mechanisms behind stereotype threat effects are reviewed, followed by empirical attempts to minimize stereotype threat effects. A review of the challenge and threat literature then follows, focusing primarily on studies in the domain of instrumental problem solving. Empirical evidence for challenge and threat appraisals will be described, as well as their breakdown into primary and secondary appraisals.

Stereotype Threat

The mechanisms underlying stereotype threat effects are assumed to be tied to self-evaluation processes, as well as the social pressures elicited by awareness of relevant stereotypes. Allport (1954) described social prejudices and the deleterious consequences
for those who are targeted. Aware that their actions are susceptible to stereotypic interpretation, the stigmatized are left to manage their tainted identities (Goffman, 1963). Stereotype threat, then, is defined as the awareness that one can be evaluated with regard to a negative stereotype related to one’s group affiliation, as well as an acknowledgment that one might behave in a manner that inadvertently confirms it (Steele & Aronson, 1995; Steele, Spencer, & Aronson, 2002). Furthermore, stereotypic judgments need not be explicitly verbalized for their effects to influence the target. Rather, stereotype threat can be elicited by the mere possibility of discrimination.

Stereotype threat is situational. The experience of the threat does not necessitate the target’s belief in the stereotype, nor the forming of chronically low expectations from repeated exposures to the stereotype. Therefore, depending on the setting and circumstance, anyone who identifies with a group for which a negative stereotype exists may experience stereotype threat. Furthermore, the specificity of the stereotype determines the situation, the targets, and the consequences. For example, men and women are equally susceptible to stereotype threat, but it is the relevance of the setting that determines what type of threat will be experienced. So, a woman may feel stereotype threat in a math class, where her group stereotype is relevant, while a man does not. Conversely, on a test of emotional sensitivity, the stereotype now applies to the man, while the woman is no longer susceptible.

Empirical Support for Stereotype Threat Effects

Steele and Aronson (1995) conducted the seminal stereotype threat study that examined the performance of African-Americans on a standardized test of verbal
aptitude. Previous research had found underperformance by groups of students who shared a particular social identity (e.g., African-Americans, Native Americans, and Latinos) in certain domains when skill and preparation were held constant (Bowen & Bok, 1998). Steele and Aronson theorized that a “threat” from a situation-specific stigma may be in part responsible for their underperformance. In Studies 1 and 2, Black and White students were recruited to take a 30-minute section of the GRE-English subject exam. In the stereotype threat condition, participants were told that the test was a measure of verbal ability, whereas in the no-threat condition, participants were told that the test was simply a measure of “problem solving” that was not diagnostic of individual differences in verbal ability. Thus, the stereotype threat condition was framed to make the negative stereotype about African-Americans salient to Black participants, thereby placing them in a situation in which their performance on the exam might serve to confirm the stereotype. The no-threat condition, on the other hand, made the racial stereotype irrelevant to their performance and, thus, was not expected to activate the stereotype. After adjusting for ability, Blacks in the stereotype threat condition did indeed underperform on the exam relative to White participants, but there was no difference in performance between Black and White participants in the no-threat condition.

In Study 3, Steele and Aronson (1995) provided evidence that the racial stereotype was activated among participants in the diagnostic condition. Using a word-fragment completion task, they found that Black participants in the diagnostic condition produced more race-related completions than their Black counterparts in the non-
diagnostic condition, as well as participants in any other condition. Moreover, the authors also found evidence for the activation of stereotype-related self-doubts for Black participants in the diagnostic condition (i.e., when asked to rate their enjoyment of a number of activities, blacks in the diagnostic condition were most averse to conforming to stereotypic images of African-Americans). Finally, Study 4 found that merely indicating their race prior to taking an exam impaired Black participants’ performance on an ostensibly non-diagnostic exam relative to a condition where participants did not indicate their race.

This early study of the effects of stereotype threat on the performance of African-Americans on standardized tests has since been replicated (e.g., Aronson, Fried, & Good, 2002; Blascovich, Spencer, Quinn & Steele, 2001; Stricker, 1998). Findings were also extended to samples of Latino participants under the same threat of intellectual inferiority to White-Americans (Aronson & Salinas, 1997; Gonzales, Blanton & Williams, 2002). As mentioned earlier, the effects of stereotype threat on women’s math performance have also been well documented (e.g., Ambady, Shih, Kim, & Pittinsky, 2001; Inzlicht & Ben-Zeev, 2000; Keller, 2002; Keller & Dauenheimer, 2003; Quinn & Spencer, 2001; Shih, Pittinsky, & Ambady, 1999; Spencer et al., 1999), and will be described in more detail later. Other stigmatized groups who have been found to suffer decrements in performance due to stereotype threat include students of low-socioeconomic background on verbal ability tests (Croizet & Claire, 1998), elderly people on memory tasks (Levy, 1996), Black and White participants on a test of athletic aptitude (Stone, Lynch, Sjomeling, & Darley, 1999), men on tests of emotional sensitivity (Leyens, Désert,
Croizet, & Darcis, 2000), and White men on a math test when comparisons to Asian men were made salient (Aronson, Lustina, Good, & Keough, 1999).

In each of these contexts, the degree of stereotype threat an individual experiences will depend in part on the strength of the stereotype involved – some stereotypes will have a more negative impact on self-evaluations than will others. For example, a threat to one’s athletic ability may not be as damaging as a threat to one’s intelligence, and therefore may elicit a weaker experience of stereotype threat. Moreover, the degree of stereotype threat experienced will depend upon both the level of identification the stigmatized person has with the performance domain as well as the stereotyped identity. Thus, the more one cares about the domain in question and/or feels a part of the stigmatized group, the stronger will be the experience of stereotype threat.

*Moderation of Stereotype Threat*

Given that stereotype threat has been shown to influence a variety of targets in a host of different domains, the focus now shifts to different variables that might affect the extent to which an individual becomes more or less vulnerable to stereotype threat. In searching for moderators of the stereotype threat effect, the following variables have received the most attention. Most moderators are defined within the test performance domain.

*Task difficulty.* Research has indicated that stereotype threat effects are strongest when the test or task is difficult (Spencer et al., 1999). Of course, the task cannot be so difficult that it is beyond the capability of the target to perform well. Upon administering two sets of test – one that participants would find easy and one that they would find
difficult – Spencer and colleagues found that women performed as well as men on the easy test, but exhibited the typical stereotype threat effect on the difficult test. It was theorized that test difficulty elicits frustration which, when combined with the awareness that poor results may be interpreted as confirmation of an existing stereotype, becomes disruptive to the test-taker (Spencer et al., 1999; Steele et al., 2002).

*Group identification.* Another important moderator is the extent to which the target identifies with the stigmatized group. Schmader (2002) found that when women completed a math test under conditions of stereotype threat, those who were high in gender identification underperformed relative to women who indicated that gender was less important to their identity. Because stereotype threat is based on the concern that one might be evaluated on the basis of a negative stereotype about one’s group, women who feel that their gender is an important part of their self-definition are most susceptible to the threatening implications of such negative stereotypes.

*Domain identification.* Research has also shown that domain identification is a moderator of stereotype threat effects (Aronson et al., 1999; Stone et al., 1998). The more a target individual identifies with a particular domain, the more susceptible that individual is to stereotype threat effects. For example, women who identify math as being important to their self-definition are more likely to experience stereotype threat effects than women who do not find math to be integral to their definition of self.

*Stigma consciousness.* Members of stigmatized groups may differ in how chronically self-conscious they are of their stigmatized status. This individual difference has been termed “stigma consciousness” (Pinel, 1999). Brown and Pinel (2003) provided
evidence that differences in stigma consciousness moderated stereotype threat effects on women’s math performance. Specifically, women who scored high in stigma consciousness performed worse on a test designed to evoke stereotype threat effects than women who were low in stigma consciousness.

Solo status. Recent research has indicated that women expect to be more negatively stereotyped when they are the lone representative of their gender within a group setting (Cohen & Swim, 1995). Sekaquaptewa and Thompson (2003) found evidence for dual influences of solo status and stereotype threat on women’s math performance: both solo status and stereotype threat conditions negatively influenced women’s performance, with poorest performance resulting in conditions where both factors were present.

Mediation of Stereotype Threat

Recent work has also focused on identifying possible mediating mechanisms that might underlie stereotype threat. Although no definitive conclusions have been drawn regarding such mechanisms (see Steele et al., 2002), progress has been made in identifying a number of potential mediating variables. Due to the complex nature of interactions between individual differences and social situations, it may be that stereotype threat works through a variety of pathways in any number of combinations. What follows is a review of the potential mediators that have received the most attention in the stereotype threat literature.

Anxiety and evaluation apprehension. Stereotype threat is evoked by the awareness of the stigmatized individual that their actions may be evaluated in light of the
stereotype about their group. This heightened awareness increases anxiety and evaluation apprehension that, in turn, undermines their performance on difficult tests (Steele et al., 2002). Spencer et al. (1999) found through self-reports that women who believed they were about to take a math test that had demonstrated gender differences in the past were more anxious than their counterparts who believed the same test had not demonstrated gender differences. Similar findings were obtained by Osborne (2001), with African-Americans indicating higher levels of anxiety before taking a test that was ostensibly diagnostic of their academic ability. Beyond self-report measures, evidence has also been reported of heightened blood pressure (i.e., consistent with feelings of anxiety) among African-Americans in typical stereotype threat-eliciting conditions (Blascovich et al., 2001).

Performance expectancies. Another possible mediator of stereotype threat effects is performance expectancy. The theory contends that stereotypes activated in stereotype threat situations lower performance expectations, which then lead to poorer task performance. In a golf-putting task, Stone et al. (1999) found that performance on a prior hole created expectations about the next hole, thereby affecting performance in the stereotype threat conditions. Even more compelling evidence for expectancy as a mediator of performance was obtained by Cadinu, Maass, Frigerio, Impagliazzo, and Latinotti (2003). In their study, female participants were told that women either perform worse, equivalently, or better than men on logic tests. Prior to taking a test, they were asked to predict their estimated level of performance. The results indicated that participants who were high in domain identification and received negative in-group
information showed a decrease in performance and lower levels of expectations. In turn, expectancy strength was found to mediate the effect of stereotype threat on performance.

Arousal. Classic social facilitation theory suggests that arousal should enhance performance on easy tasks and impair performance on difficult tasks (Zajonc, 1965). Specifically, arousal leads to heightened activity primarily in the sympathetic nervous system characterized by increased heart rate, systolic blood pressure, and the release of epinephrine. Arousal, then, is nonspecific, nondirective, and enhances behavior toward the dominant response. Thus, when the task is difficult, arousal will increase the number of inappropriate responses, thereby creating a decrement in overall performance, whereas when the task is easy, arousal will increase the number of appropriate responses and thereby enhance overall performance. O’Brien and Crandall (2003) applied this theory of arousal to the stereotype domain by showing not only that threat from a negative stereotype impaired performance on a difficult task, but also that stereotype type threat enhanced performance on an easy task. They posited that fear and anxiety stemming from the possibility of confirming a negative stereotype, or being categorized as an exemplar of a negative stereotype, is sufficient to create high arousal, which then interferes with performance. According to their results, women under stereotype threat performed worse on a difficult math test, but better on an easy math test, whereas men showed no differences, suggesting arousal as an underlying mechanism.

Reduced working memory capacity. Schmader and Johns (2003) provided evidence that stereotype threat interferes with test performance by reducing the individual’s working memory capacity – the ability to focus attention on a given task
while keeping task-irrelevant thoughts out of awareness. In their study, working memory capacity was assessed by an operation-span task, in which participants calculated math questions while at the same time memorizing words for later recall (Turner & Engle, 1989). As predicted, women who believed the working memory task was a test of mathematical ability showed reduced working memory capacity relative to men and women in the no-threat condition. In other words, women in the stereotype threat condition may have channeled some of their mental resources into suppressing unwanted thoughts of the negative stereotype, thereby leaving less working memory for the task at hand.

**Effort.** Finally, it seems plausible that a stigmatized individual would give up or reduce effort when faced with a difficult task under conditions that evoke stereotype threat. However, previous research has failed to find evidence for effort mediating stereotype threat effects. Aronson and Salinas (2001) found that even when participants believed their effort was being physiologically monitored, and that they would have to repeat the test if they did not put forth enough effort, stereotype threat effects still emerged (as cited in Steele et al., 2002). Moreover, because most stereotype threat studies are concerned with individuals who are highly identified with the relevant domain, it is unlikely that reduced effort would account for task underperformance.

*Coping with the Consequences of Stereotype Threat*

Negative stereotypes about stigmatized individuals are widely known within a society, not only by those who are targets of the stereotypes (Steele, 1997), but even by those who do not personally endorse them (Devine, 1989). The threat of being negatively
evaluated by prejudiced and non-prejudiced persons alike places stigmatized individuals in a precarious position. Recent research has examined various means by which the stigmatized individual can cope with and even counter the effects of stereotype threat.

**Self-handicapping.** Self-handicapping is a strategy that allows the individual to establish external explanations for a potentially poor performance (Jones & Berglas, 1978). Steele and Aronson (1995) found increased self-handicapping under conditions of stereotype threat, in that Black participants who expected to take a diagnostic test of verbal aptitude reported less sleep before and less concentration during the test as compared to Blacks who thought the test was non-diagnostic. Reporting possible external factors that might hinder performance allows the individual to pre-empt any negative evaluations that might arise from knowledge of the stereotype. These findings were further extended by Keller (2002), who found that stereotype threat conditions increased self-handicapping in women about to take a difficult math test.

**Disidentification.** In order to defend themselves against the aversive consequences of stereotype threat, stigmatized individuals may choose to adapt by separating or psychologically disidentifying their self-esteem from stereotyped domains. For instance, excelling in other academic domains makes it easier for a woman to distance herself and her identity from the math domain. In a study conducted by Major, Spencer, Schmader, Wolfe, and Crocker (1998), Black and White participants received bogus feedback on an intelligence test and then completed measures of self-esteem. The results indicated that White participants had higher self-esteem after success than after failure. In contrast, because they believed that they were at risk of confirming a negative
stereotype on a possibly biased test, Black participants displayed similar levels of self-esteem, regardless of feedback. Similarly, Osborne (1997) found weak correlations between academic performance and self-esteem among Blacks as compared to Whites.

**Domain avoidance.** Another coping strategy stigmatized individuals may use to defend against the consequences of stereotype threat is to completely avoid all stereotype-relevant situations. Davies, Spencer, Quinn, and Gerhardstein (2003) found that women who watched gender-stereotypic commercials later indicated less interest in stereotypically threatening domains related to math, and more interest in stereotypically less threatening domains such as verbal ability. In addition, Steele and Aronson (1995) found that Black males distanced themselves from stereotypically Black activities, such as playing basketball or rap music, when faced with a diagnostic test, thereby disassociating themselves from activities that might link them to the negative stereotype.

**Self-affirmation.** Research on self-affirmation has provided evidence that threats to the self from negative feedback can be alleviated by accessing positive alternative self-identities that restore the self-image (Fein & Spencer, 1997; Steele, 1988; Steele, Spencer, & Lynch, 1993). To be self-affirming, the identity accessed must be important to one’s self-worth. If stereotype threat serves as a threat to the self, then self-affirmation should be able to alleviate its performance-depressing effects. Croizet, Désert, Dutrévis, and Leyens (2001) exposed women to a difficult math test, after which half of the participants completed a questionnaire that led them to think they were good students. When the stereotype-threatened women completed the questionnaire that affirmed their academic self-image, their scores on the math test were higher than those who did not
self-affirm. Barlow (2002) found similar results when stereotypically threatened women were asked to circle the value that was most important to them and write about why it was important prior to taking a math test. This served as an affirmation of an important self-concept of the participant’s choosing, and resulted in math scores similar to women not under stereotype threat.

Identity bifurcation. Finally, another strategy that can be used to counter the effects of stereotype threat is identity bifurcation. Bifurcation is defined as separating one’s self from the aspects of one’s in-group that are linked to criticism in the performance domain, while simultaneously continuing to identify with in-group characteristics that are seen as unrelated to the domain. For example, after encountering stereotype threat information, women highly identified with math rejected negative female characteristics perceived to be strongly associated with stereotypes about women’s math ability (e.g., flirtatious, gossipy), while still endorsing positive characteristics supposedly unrelated to the stereotype (e.g., empathy, nurturance) (Pronin, Steele, & Ross, 2004). Identity bifurcation can be distinguished from disidentification (Major et al., 1998) in that the individual is disconnecting from the group targeted by the stereotype, as opposed to the domain in which the threat applies.

Challenge and Threat Appraisals

Lazarus and Folkman (1984) suggested that an individual’s cognitive, affective, and behavioral responses to a potentially stressful situation depend upon the appraisal of that situation. They proposed that threat and challenge appraisals occur in anticipation of stressful situations that threaten one’s well-being. Threat appraisals occur when one’s
perception of harm or distress exceeds one’s perception of their abilities or resources to cope with the stressor, while challenge appraisals, in contrast, occur when one’s perception of the stressor does not exceed one’s perception of their abilities or resources to cope. Those who appraise a situation as a threat tend to focus on the potential for loss, whereas those who appraise the same situation as a challenge instead perceive the potential for gain (Lazarus & Folkman, 1984; Tomaka, Blascovich, Kelsey, & Leitten, 1993). According to the biopsychosocial model (Blascovich, Mendes, Hunter, Lickel, & Kowai-Bell, 2001; Blascovich, Mendes, Hunter, & Solomon, 1999; Blascovich, Mendes, Tomaka, Salomon, Seery, 2002; Blascovich & Tomaka, 1996), situational demand characteristics of threat appraisals invoke judgments of danger, uncertainty, and a calculation of required effort. In contrast, challenge appraisals invoke judgments of all relevant knowledge, abilities, dispositions, and support required to be successful in the situation.

Primary and Secondary Appraisals

Two major components of the appraisal process are primary and secondary appraisals. Primary appraisal refers to how the individual perceives the nature of the situation, and if there is any degree of risk (Lazarus & Folkman, 1984). In other words, the person judges what is at stake, and evaluates the situation as being an opportunity for either self-growth or risk. For example, Holmes and Houston (1974) examined the effect of different primary appraisals on participants experiencing an electric shock. Those participants who were instructed to appraise the situation as a new physiological sensation reported less anxiety than did participants who were not encouraged to do so.
Secondary appraisals, on the other hand, reflect perceptions of resources to cope with the situation (Lazarus & Folkman, 1984). In contrast to primary appraisal, secondary appraisal is concerned with the controllability of the situation. The individual determines if there are enough personal resources to cope with or manage the demands of the situation. Folkman, Lazarus, Dunkel-Schetter, DeLongis, and Gruen (1986) found that participants who appraised a stressful situation as controllable or changeable used more task-oriented problem-solving strategies, as opposed to those who thought the situation was unchangeable and instead used more emotion-oriented coping strategies (e.g., distancing, escape avoidance, see also Markman, Gavanski, Sherman, & McMullen, 1993).

**Evidence for Challenge and Threat Appraisals**

Evidence for challenge and threat appraisals is often obtained through physiological measures (Blascovich et al., 1999; Blascovich & Tomaka, 1996; Tomaka et al., 1993). Blascovich and colleagues argue that situations that evoke challenge and threat appraisals should provide correspondingly different physiological responses. In order to distinguish between challenge and threat responses, these researchers limit all performance situations to those that required more active, psychological demands (e.g., giving a speech, solving math problems, interpersonal negotiations). The index of threat and challenge developed by Blascovich et al. is based on patterns of cardiovascular reactivity. Typically, they examine pre-ejection period (PEP), or the time from left ventricle contraction until the aortic valve opens, as well as changes in cardiac output (CO), measured in liters of blood ejected from the left ventricle per minute. In addition,
they examine heart rate (HR), a rate-based measure of cardiac performance, and total peripheral resistance (TPR), which is an index of vascular resistance derived from blood pressure. Challenge situations, then, are characterized by enhanced cardiac performance, decreased vascular resistance, and little or no change in blood pressure. The cardiovascular performance found to occur during challenge situations is similar to that of aerobic exercise and is more energy efficient for coping (Blascovich et al., 1999). In threat situations, however, cardiac performance is also greater, but because vascular resistance is not decreased, there is an increase in blood pressure.

To illustrate, Tomaka et al. (1997) conducted a study that elicited challenge and threat appraisals among participants expecting to engage in a potentially stressful laboratory task. Participants performed a mental arithmetic task after hearing one of two sets of instructions: threat instructions, which emphasized accuracy, speed, and potential evaluation, and challenge instructions, which emphasized effort and trying one’s best. Appraisals were assessed after participants read the instructions, and physiological variables were measured during the task. Results showed that the participants who were given the threat instructions not only appraised the upcoming task as threatening, but also evidenced physiological response patterns consistent with threat appraisals (i.e., increased PEP, CO, and HR, and TPR). In contrast, participants who were given instructions emphasizing challenge appraised the task as challenging and evidenced physiological response patterns consistent with challenge appraisals.
Social Facilitation as Challenge and Threat

Blascovich et al. (1999) used their model of threat and challenge to explain classic social facilitation effects (Zajonc, 1965) by showing that the presence of others elicits arousal that, in turn, increases the likelihood of challenge and threat responses. They hypothesized that when individuals perform well-learned tasks in the presence of others they should exhibit physiological characteristics consistent with challenge appraisals, whereas those who perform unlearned tasks in the presence of others should exhibit physiological characteristics consistent with threat appraisals. Participants in their study performed a well-learned task or an unlearned task either alone or in the presence of another while cardiovascular responses were being recorded. Results indicated that the presence of others during well-learned tasks evidenced the challenge pattern of cardiovascular reactivity, whereas the presence of others during an unlearned task evidenced the threat pattern of cardiovascular reactivity. The results of their study provided support for a challenge and threat rationale for both the performance enhancement and impairment effects of social facilitation.

Goal of the Present Research

The threat and challenge literature has indicated that people may react to stressful situations either by performing worse or better. The goal of the present research is to apply this idea to the stereotype threat literature, which thus far has assumed that the only reaction to stressful situations is performance impairment. It is proposed here, however, that making individuals aware of existing cultural stereotypes may in fact allow stigmatized individuals to perform better or at least short-circuit the effects of stereotype
threat. In a related study, O’Brien and Crandall (2003) found that threat from a negative stereotype impaired performance on a difficult task, but enhanced performance on an easy task. However, enhanced performance in their study did not focus on stereotype challenge per se, but rather on the typical enhancing effects of arousal on the performance of easy tasks. In contrast, the present work focuses on challenge effects evoked by a motive to disconfirm the stereotype as applied to difficult tasks.

Overview of Study

The present study examined stereotype challenge and threat responses through a manipulation of social comparison information. Female participants received feedback about an ostensibly matched partner who was either similar (i.e., female) or dissimilar (i.e., male) to themselves. After learning that they will be taking a math task that either has (i.e., high stereotype salience) or has not (i.e., low stereotype salience) demonstrated gender differences in the past, participants learned that their partner performed very well or very poorly on the math task they were about to take. Subsequently, participants took the math task themselves, and their performance was evaluated.

Hypotheses and Expected Results

Social comparison theory (Festinger, 1954) posits that individuals compare themselves to others in order to evaluate their abilities or beliefs. In so doing, individuals provide each other with standards for self-evaluation. Social comparisons also function as a coping mechanism when facing negative situations (e.g., Taylor & Lobel, 1989; Wills, 1981; Wood, Taylor, & Lichtman, 1985). In such situations, Lockwood and her colleagues (e.g., Lockwood, 2002; Lockwood, Jordan, & Kunda, 2002; Lockwood &
Kunda, 1997) have found that comparison others can serve as either positive or negative role models when the comparison other is considered similar to oneself. Thus, Lockwood and Kunda (1997) found that a similar upward comparison target was more inspiring to participants than a dissimilar target, so long as the outcome in question was perceived as being attainable, whereas Lockwood (2002) found that although a similar downward comparison target exerted a deflating effect on self-perceptions to the extent that participants felt vulnerable to having a similar negative fate befall them, they nevertheless indicated that they were more motivated to work hard to avoid such an outcome in the future. More generally, research has shown that comparisons with similar others yield assimilation toward the standard of comparison, while comparisons with dissimilar others yield contrast away from the standard (Mussweiler, 2001; Mussweiler, 2003; see also Markman & McMullen, 2003).

Hypothesis 1: Based on social comparison theory, participants in the low stereotype salience condition will be more motivated when the comparison target is a similar female than when the comparison target is a dissimilar male.

Mendes et al. (2001) examined challenge and threat responses following downward and upward social comparisons. Participants assigned to the upward comparison condition received feedback that their interaction partner was ranked higher than they were on a word-finding task, whereas those in the downward comparison condition received feedback that their partner was ranked lower than they were on the task. Results indicated that participants who interacted with upward comparison partners
evaluated an ensuing cooperative word-finding task as more threatening than participants who interacted with downward comparison targets, who instead evaluated the task as more challenging. These findings, considered along side those of Lockwood and colleagues, lead to the generation of more complex predictions regarding social comparison effects on challenge and threat appraisals. In other words, comparisons with a similar or dissimilar partner should elicit challenge and threat effects depending on the direction (i.e., upward vs. downward) of the comparison. More specifically, a female who observes either a strong, stereotype-disconfirming performance by a (similar) female, or a poor, stereotype-disconfirming performance by a (dissimilar) male, should feel relatively more inspired and challenged, and thereby exhibit a relatively higher score when they take the math task themselves. On the other hand, a female who observes either a poor, stereotype-confirming performance by a (similar) female, or a strong, stereotype-confirming performance by a (dissimilar) male, should feel relatively less inspired and more threatened, and thereby exhibit a relatively lower score when they take the math task themselves.

Hypothesis 2: In the stigma-salient condition, participants will exhibit enhanced performance characteristic of challenge effects when the comparison target is either a strong-performing, similar exemplar or a poor-performing, dissimilar exemplar. In contrast, participants will exhibit impaired performance characteristic of threat effects when the comparison target is either a poor-performing, similar exemplar or a strong-performing, dissimilar exemplar.
Method

Participants

One hundred sixty-five female participants were recruited from introductory psychology courses at Ohio University in partial fulfillment of course requirement and randomly assigned to conditions of a 2 (Stereotype Salience: high salience vs. low salience) X 2 (Exemplar Gender: female vs. male) X 2 (Exemplar Performance: strong vs. poor) between-subjects factorial design.

Stimulus Materials

Challenge-threat manipulation check items. Perceptions of challenge and threat in response to the task were measured using a scale employed by Drach-Zahavy and Erez (2002). The measure consisted of 12 Likert-type items describing primary and secondary appraisals (see Appendix). Eight items meant to measure primary appraisal addressed the participants’ perceptions of the expected consequences of the situation. The challenge appraisal items were: “The task seems like a challenge to me,” “The task provides opportunities to exercise reasoning skills,” “The task provides opportunities to overcome obstacles,” and “The task provides opportunities to strengthen my self-esteem.” The threat appraisal items were: “The task seems like a threat to me,” “I’m worried that the task might reveal my weaknesses,” “The task seems long and tiresome,” and “I’m worried that the task might threaten my self-esteem.” Secondary appraisals were measured by four items addressing participants’ perceptions of controllability, their available resources, and their expectancies for success. The secondary challenge appraisals were: “Overall, I think I’ll succeed in carrying out the task,” and “I think that I
have the abilities necessary for successful performance.” The secondary threat appraisals were: “Overall, it seems that I cannot succeed in a task like this,” and “I’m worried that I lack the abilities to perform the task successfully.”

**Math task.** The math task comprised of 45 multiple-choice questions drawn from practice tests for the quantitative section of the Graduate Records Examination (GRE), and is a modified version of a test used by Schmader (2002). In order to correct for guessing when calculating the overall percentage, correct items received 1 point, items left blank received no points, and incorrect items received a deduction of 1 point divided by the number of response options for that item (e.g., Spencer et al., 1999).

**Procedure**

Participants arrived and waited in the lobby outside of the laboratory. A male experimenter greeted the participant, escorted her into a separate waiting room, and announced that the study examined “matching working styles in problem solving.”

**Stereotype salience manipulation.** Participants sat at a desk and read instructions indicating that they were about to engage in a math task that would be used to examine problem-solving styles. In the high salience condition, the instructions indicated that the math test had been shown to be diagnostic of math ability and had also been known in the past to produce gender differences, specifically, that women typically perform worse on the task than do men. In the low salience condition, participants simply learned that they would be taking a test that was “evaluated across a large group of students.” These instructions have proven in past research to be successful in manipulating stereotype threat conditions (see Spencer et al., 1999).
**Exemplar gender and performance manipulations.** In the female exemplar condition, the experimenter explained to the participant that she has been matched with another student judged to be very similar to her based on their responses to a prior prescreening questionnaire, whereas in the male exemplar condition, the experimenter explained that the participant has been matched with a male student judged to be very dissimilar to her on the basis of their responses to the prescreening questionnaire. To further enhance the strength of the manipulation, the participant learned that their partner was currently completing the experiment, and that the participant’s data were now needed in order to examine similarities and differences in working styles. The experimenter then left the room, ostensibly to check on the other student’s progress.

The experimenter returned five minutes later with either the male or female student with whom the participant supposedly had been paired. The student, a confederate of the experimenter, was told to wait while the experimenter checked to make sure that “all was well with the data.” When the experimenter left the room, the confederate casually initiated a conversation with the participant, asked if she had completed the math task yet, and proceeded to tell her how he (she) had performed. In the *strong performance* condition, the confederate mentioned that he (she) had done well on the math task, scoring in the 90th percentile among Ohio University undergraduates, whereas in the *weak performance* condition the confederate mentioned that he (she) had done poorly on the math task, scoring in only the 10th percentile of Ohio University undergraduates.
After this conversation ended, the experimenter returned to dismiss the confederate and escort the participant to a small computer cubicle. After completing the initial measure of threat and challenge appraisals (Drach-Zahavy & Erez, 2002), the participant was then given 25 minutes to work on the math task, presented on MediaLab (Jarvis, 2005) experimental software. Upon completion, participants were asked to report their scores on the quantitative section of the Scholastic Achievement Test (SAT), after which they were thoroughly debriefed and thanked for their participation.

Results

Data were collected from 165 participants, however data from seven participants were eliminated, four because they appeared to have answered the questions randomly, two because they chose not to complete the experiment, and one because she voiced suspicion after recognizing the confederate. Analyses were thus performed on data provided by the 158 participants that remained. Fifty of these 158 participants reported their scores on the quantitative section of the SAT, and analyses revealed no differences as a function of stereotype salience, exemplar gender, or exemplar performance. Because only a minority of participants reported their SAT scores, these scores were not used as a covariate in any subsequent analyses.

To examine the effect of downward and upward comparisons to stereotypic or counter-stereotypic exemplars on women’s math performance, a 2 (Stereotype Salience) X 2 (Exemplar Gender) X 2 (Exemplar Performance) analysis of variance (ANOVA) was conducted on the number of test items answered correctly. Results revealed a main effect for stereotype salience, $F(1, 150) = 4.49$, $p = .04$, indicating that participants in the high
salient condition answered fewer items correctly ($M = 8.27$) than did participants in the low salience condition ($M = 9.44$). In particular, the analysis also revealed a significant Salience X Gender X Performance interaction, $F(1, 150) = 7.31$, $p = .008$ (see Table 1).

Table 1: Number of Items Answered Correctly As a Function of Stereotype Salience, Exemplar Gender, and Exemplar Performance

<table>
<thead>
<tr>
<th>Stereotype Salience</th>
<th>Low Salience</th>
<th>High Salience</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Poor</td>
<td>Strong</td>
</tr>
<tr>
<td>Female Exemplar</td>
<td>10.50 (4.08)</td>
<td>9.11 (4.61)</td>
</tr>
<tr>
<td>Male Exemplar</td>
<td>8.86 (4.53)</td>
<td>9.43 (2.63)</td>
</tr>
</tbody>
</table>

$F(1, 150) = 7.31$, $p = .008$

Note: Cell means refer to the number of items answered correctly (out of 45). Standard deviations appear in parentheses.

To explore the nature of the three-way interaction, separate analyses were conducted on the low and high salient conditions. To address more closely the predictions made in Hypothesis 1, analysis within the low stereotype salience condition showed that participants in the female exemplar condition ($M = 9.78$, $SD = 4.35$) on the average scored slightly better on the math test than those in the male exemplar condition ($M = 9.15$, $SD = 3.62$), although the test failed to show any significant main effect.

Similar tendencies were found in the data when considering the effects of comparisons to female exemplars in contrast to male exemplars on accuracy ($M = 41.49$ and $M = 35.89$, respectively) and overall percentage$^2$ ($M = 15.08$ and $M = 12.28$, respectively).

Addressing the predictions made in Hypothesis 2, analysis within the high stereotype salience condition yielded a significant Gender X Performance interaction,

$^2$ The overall percentage calculation included a correction for guessing as specified in the Methods section.
$F(1, 73) = 8.34, p = .005$. Consistent with predictions, participants who interacted with a strong-performing, similar exemplar ($M = 9.40$) answered more items correct compared to those who interacted with a poor-performing, similar exemplar ($M = 7.00$), $t(71) = 2.06, p = .02$ (one-tailed). Likewise, participants who interacted with a poor-performing, dissimilar exemplar ($M = 9.25$) answered more items correct compared to those who interacted with a strong-performing, dissimilar exemplar ($M = 7.26$), $t(71) = 1.67, p = .02$ (one-tailed). Analyses performed on accuracy and overall percentage revealed similar 3-way interaction patterns, $F(1, 150) = 7.03, p = .009$, and $F(1, 150) = 7.06, p = .009$, respectively (see Table 2 and Table 3).

To examine the measure of threat and challenge appraisals, a total appraisal score was computed by summing the challenge scores and reversed threat scores. Higher scores indicated high challenge appraisal. A 3-way ANOVA was conducted on the total appraisal score. Although no main effects were found for this measure, the Salience X Gender interaction was marginally significant, $F(1, 150) = 3.09, p = .08$. In other words, participants in the high stereotype salience condition showed higher challenge appraisals when the interaction partner was female ($M = 58.0$) as opposed to male ($M = 54.7$). However, this relationship reversed in the low stereotype salience condition resulting in higher challenge appraisals when the interaction partner was male ($M = 57.7$) as opposed to a female ($M = 55.1$). A 3-way ANOVA conducted on only the challenge appraisal items revealed a significant Gender main effect, $F(1, 150) = 4.16, p = .04$. Female participants appraised the math task as more of a challenge when they interacted with a female partner ($M = 8.60$) as opposed to a male partner ($M = 6.15$).
Table 2: Accuracy As a Function of Stereotype Salience, Exemplar Gender, and Exemplar Performance

<table>
<thead>
<tr>
<th>Exemplar Performance</th>
<th>Stereotype Salience</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Salience</td>
<td>High Salience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>Strong</td>
<td>Poor</td>
<td>Strong</td>
<td></td>
</tr>
<tr>
<td>Female Exemplar</td>
<td>42.82 (17.5)</td>
<td>40.21 (19.6)</td>
<td>29.63 (11.7)</td>
<td>47.46 (24.9)</td>
<td></td>
</tr>
<tr>
<td>Male Exemplar</td>
<td>30.75 (17.3)</td>
<td>40.59 (24.3)</td>
<td>36.31 (17.7)</td>
<td>33.38 (18.8)</td>
<td></td>
</tr>
</tbody>
</table>

F(1, 150) = 7.03, p = .009

Note: Accuracy was calculated by dividing the number of correct answers by the total number of questions answered and multiplying by 100%. Standard deviations appear in parentheses.

Table 3: Overall Percentage As a Function of Stereotype Salience, Exemplar Gender, and Exemplar Performance

<table>
<thead>
<tr>
<th>Exemplar Performance</th>
<th>Stereotype Salience</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low Salience</td>
<td>High Salience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>Strong</td>
<td>Poor</td>
<td>Strong</td>
<td></td>
</tr>
<tr>
<td>Female Exemplar</td>
<td>16.53 (11.6)</td>
<td>13.71 (11.7)</td>
<td>7.34 (7.8)</td>
<td>14.86 (8.10)</td>
<td></td>
</tr>
<tr>
<td>Male Exemplar</td>
<td>11.03 (11.5)</td>
<td>13.42 (7.19)</td>
<td>12.65 (10.2)</td>
<td>8.93 (8.47)</td>
<td></td>
</tr>
</tbody>
</table>

F(1, 150) = 7.06, p = .009

Note: When calculating the overall percentage, correct items received 1 point, items left blank received no points, and incorrect items received a deduction of 1 point divided by the number of response options for that item. Standard deviations appear in parentheses.

Discussion

Previous research in the area of social comparisons uncovered the effects that similar or dissimilar exemplars can have on an individual. In contrast to dissimilar exemplars, similar upward comparison targets can be inspirational while similar downward comparison targets serve as motivation to avoid negative outcomes in the future (Lockwood, 2002; Lockwood, Jordan, & Kunda, 2002; Lockwood and Kunda, 1997). More generally, assimilation towards the standard of comparison results from
comparisons with similar others, while contrast away from the standard results from comparisons with dissimilar others (Mussweiler, 2001, 2003; see also Markman & McMullen, 2003). The goal of the present research was to apply this idea of assimilation towards and contrast away from similar or dissimilar others to the stereotype threat literature in hopes of eliciting stereotype challenge effects (i.e. performance enhancement) to the already established stereotype threat findings (i.e., performance impairment).

It was predicted that exposure to a similar or dissimilar partner whose performance on a stereotype relevant task disconfirmed the negative stereotype would exhibit higher scores indicative of a challenge response, whereas exposure to a similar or dissimilar partner whose performance confirmed the negative stereotype would exhibit lower scores indicative of a threat response. Results indicated that participants in the low stereotype salience condition had higher math scores when the comparison target was a similar female than when the comparison target was a dissimilar male. Furthermore, in the high stereotype salience condition, participants exhibited impaired performance characteristic of threat effects when the comparison target was either a poor-performing, similar exemplar or a strong-performing, dissimilar exemplar. In contrast, although performance was not significantly enhanced as predicted, threat effects were diminished as evidenced by the superior scores exhibited by participants who interacted with either a strong-performing, similar exemplar or a poor-performing, dissimilar exemplar. In other words, comparisons with counter-stereotypic exemplars in light of a negative stigma seemed to “short-circuit” the negative effects of stereotype threat on performance.
In addition, the results of the math test do indeed show a pattern that could be predicted from a difference in challenge and threat appraisals. However, the particular challenge-threat measure used in this study could not shed more light on the matter. The challenge-threat scale, originally used by Drach-Zahavy and Erez (2002), was initially intended to examine the effects of stress and goals and their interaction on the perceptions of threat and challenge. Perhaps a more subtle (albeit intrusive) measurement, such as evaluating physiological responses, would suit our purposes better. Nevertheless, when considering the challenge items separately, participants judged the math task as more of a challenge when interacting with a female partner relative to a male partner, as predicted. A similar pattern was found when considering the threat items separately, but the difference did not quite reach significance.

These results raise some interesting theoretical questions in light of the perspective recently extended by Stapel and Koomen (2001). According to their theory, when a person’s social self becomes accessible, that person shifts into a “we” frame of mind which emphasizes their in-group and stresses comparisons with similar others. In contrast, when an individual’s personal self is more accessible, that individual shifts into an “I” frame of mind which emphasizes distinctness and focuses on differences with comparison others. Extending this perspective, perhaps participants in the present study shift into a “we” frame of mind when interacting with a similar female. The focus on their social self may lead to increased comfort and perhaps camaraderie amongst their valued in-group which, in turn, reduces anxiety or apprehension usually present during stereotype threat situations. Conversely, an interaction with a dissimilar male may cause
a shift into a more individuating “I” frame of mind, which along with the impending task decreases feelings of comfort while increasing appraisals of threat.

Moreover, research by Sekaquaptewa and Thompson (2003) showed that, under stereotype threat conditions, women performed worse on a math test when they were the only woman present compared to when they were in same-gender groups, suggesting that women experienced greater apprehension about stereotypic evaluations when they were the lone representative of their gender. In the present study, perhaps similar apprehensions arose when women interacted with a male partner instead of another woman, further exacerbating the “I” frame of mind.

Marx, Stapel, and Muller (2005) found further evidence that a collective self-construal orientation (Aron, Aron, & Smollan, 1992; Brewer & Gardner, 1996) is activated for those individuals targeted by stereotype threat. Therefore, stereotype threat conditions heighten the accessibility of the collective self, enhance feelings of “we-ness” which, in turn, pull for assimilative responses to in-group comparison information. In contrast, while in this “we” frame of mind, exposure to out-group comparison information under stereotype threat conditions activates a competing “they” representation.

Wills (1991) argued that downward comparisons with dissimilar partners resulted in more positive reactions relative to similar partners because the latter increase discomfort. In other words, Wills claims that discomfort occurs when one is superior to a member of their in-group, but does not occur when one is superior to an out-group member. In the present study, this could explain in part, the assimilation effects found
when participants interacted with a poor performing similar exemplar under high stereotype salience and the contrast effects when participants interacted with a poor performing dissimilar exemplar. In turn, upward comparisons with dissimilar others evoke implicit validation of the dissimilar other’s superiority, also causing discomfort. Mendes et al. (2001) provide further evidence that upward and downward social comparisons effect challenge and threat appraisals. They found that upward comparisons with dissimilar partners, relative to similar partners, resulted in greater threat evaluations, stronger cardiovascular threat responses, and greater reported discomfort, whereas downward comparisons with dissimilar partners engendered more challenge evaluations, stronger cardiovascular challenge responses, and less discomfort.

Upon further examination of the complete challenge-threat measure, it may be worthwhile to note the interaction effect between stereotype salience and exemplar gender. Participants in the high stereotype salience condition again showed higher challenge appraisals when the interaction partner was female as opposed to male. However, this relationship reversed in the low stereotype salience condition resulting in higher challenge appraisals when the interaction partner was male. Although this interaction was not quite significant, it is possible that feelings of challenge when faced with a male partner were evoked in a different way. Specifically, awareness of the negative stigma in a low stereotype threat situation may instead arouse anger, thereby motivating one to try to disconfirm the stereotype. Pekrun, Goetz, Titz, and Perry (2002) developed a taxonomy of academic emotions by classifying them in terms of emotions (positive vs. negative) and behavior effects (activating vs. deactivating). From the four
groups that can be produced by crossing the two dimensions, anger, along with anxiety and shame, is classified as a “negative activating emotion” that induces a strong motivation to cope with negative events, overcome obstacles, and avoid failures. Furthermore, negative activating emotions may elicit beneficial effects when learning hinges on rehearsal strategies, effort, and attention. Therefore, a female interaction partner in a low stereotype salience situation would not elicit any feelings of challenge or competition because together they share the same stigma. However, a male interaction partner in conjunction with an impending math test may be enough to evoke the female participant’s negative stigma. At this point, the less threatening test in the low salience condition arouses anger and challenge, whereas the more foreboding test in the high salience condition arouses anxiety and threat.

Social psychological research is rich in its attempts to disconfirm stereotypes and alleviate prejudice. Most strategies, however, require controlled attitude change or stereotype suppression within an individual’s awareness (Brewer & Miller, 1988; Devine, 1989; Gaertner & Dovidio, 1986; Macrae, Bodenhausen, Milne, & Jetten, 1994; Monteith, 1993). In fact, stereotypic attitudes were proposed to be automatic and inescapable despite attempts to ignore them (Bargh, 1999; Devine, 1989), and persisted because of its cognitive pragmatism (Fiske & Stevens, 1993). Nevertheless, other research proposed that stereotypic attitudes are actually malleable, and that effortful processing may not be necessary in stereotype disconfirmation, but instead change as a function of the social context (Schwarz & Bless, 1992) or the perceiver’s goals and expectations (Macrae, Bodenhausen, Milne, Thorn, & Castelli, 1997). The idea of
malleable stereotypic attitudes is relevant to the present study because stereotype disconfirmation must occur when faced with a disconfirming exemplar in order for the target to view a stereotypically threatening situation as challenging and to eventually “short-circuit” the stereotype threat. Dasgupta and Greenwald (2001) showed evidence for attitude change after participants were exposed to admired versus disliked exemplars. For example, bias was reduced when admired African Americans (e.g. Martin Luther King) and disliked European Americans (e.g., Ted Bundy) were made salient. A similar mechanism may underlie the results of the present study. Exposure to an exemplar will make relevant information accessible, especially the perceiver’s evaluation of him or her, which in turn is applied to the situation at hand. Therefore, evaluation of the exemplar and whether the exemplar performs stereotypically or counter to the stereotype is critical in the appraisal of a stereotypically threatening situation.

Regardless of the specific underlying process that may be at work here, the present study is successful in demonstrating that people can look to similar or dissimilar others alike as sources of inspiration in the context of a negative stereotype. One question that arises from this research is the degree to which the phenomenon generalizes from the laboratory to applied settings, such as admissions testing for higher education or skills analysis in employment. Gender differences on tests that measure mathematical intelligence have long been observed, and much research has been devoted to understanding the cause of the differences, and addressing the underlying issues. In light of the present study, female students would do well to not only attend to evidence that confirms the stereotype (i.e., strong-performing male peers and poor-performing female
peers), but also to evidence that disconfirms the stereotype (i.e., strong-performing female peers and poor-performing male peers). Test takers and makers alike would benefit from further research that explored the pros and cons of same-sex testing environments. Future work that investigates under what circumstances in-group and out-group comparisons and interactions lead to exacerbating or alleviating stereotype threat related underperformance would be well received. Also, it may be useful to know whether some people are more “susceptible” targets than others. Of the differences between man and woman Tolstoy remarked, “Woman is more impressionable than man.” Nevertheless, the present study shows that a reduction in the math performance gap between men and women may actually lie in how the impression is managed.
References


Appendix

Please take a moment and answer each question the best you can by circling the appropriate number.

1. The task seems like a challenge to me.

   1--------2--------3--------4--------5--------6--------7
   Not At All     Very Much

2. The task provides opportunities to exercise reasoning skills

   1--------2--------3--------4--------5--------6--------7
   Not At All     Very Much

3. The task provides opportunities to overcome obstacles.

   1--------2--------3--------4--------5--------6--------7
   Not At All     Very Much

4. The task provides opportunities to strengthen my self-esteem.

   1--------2--------3--------4--------5--------6--------7
   Not At All     Very Much

5. The task seems like a threat to me.

   1--------2--------3--------4--------5--------6--------7
   Not At All     Very Much

6. I’m worried that the task might reveal my weaknesses.

   1--------2--------3--------4--------5--------6--------7
   Not At All     Very Much

7. The task seems long and tiresome.

   1--------2--------3--------4--------5--------6--------7
   Not At All     Very Much
Appendix

8. I’m worried that the task might threaten my self-esteem.

1--------2--------3--------4--------5--------6--------7
Not At All     Very Much

9. Overall, I think I’ll succeed in carrying out the task.

1--------2--------3--------4--------5--------6--------7
Not At All     Very Much

10. I think that I have the abilities necessary for successful performance.

1--------2--------3--------4--------5--------6--------7
Not At All     Very Much

11. Overall, it seems that I cannot succeed in a task like this.

1--------2--------3--------4--------5--------6--------7
Not At All     Very Much

12. I’m worried that I lack the abilities to perform the task successfully.

1--------2--------3--------4--------5--------6--------7
Not At All     Very Much

Back to text