VISUAL-SPATIAL SKILLS: A BIOPSYCHOSOCIAL

APPROACH TO PEFORMANCE DIFFERENCES

By

Jill B. Fancher

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
Department of Psychology

AUGUST 2008

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of JILL B FANCHER find it satisfactory and recommend that it be accepted.

____
Chair

ACKNOWLEDGMENTS

My daughter, Emma, said it best when she said, “When I am with you, I am not alone.” This project was not completed in isolation, indeed it was completed with the most thoughtful and supportive mentorship and guidance. Elizabeth Soliday, Ph.D., took me on as her student after my M.S. project. She asked me a simple question, “What are you interested in?” This question spawned a series of conversations that resulted in this dissertation. Some mentors are research advisors, some guide careers, and some mentors provide general support; Dr. Soliday does all three. Through her guidance, patience, and amazingly efficient editing she brought me from idea to write-up in the midst of my internship year. Thank you for *everything*!

A team effort is more inclusive than just advisor-advisee. My committee, Maureen Schmitter-Edgecombe, Ph.D, Heidi Hamann, Ph.D., and Lisa Fournier, Ph.D., came together, each providing a unique set of skills and experience, to help form a cohesive, feasible, and well-constructed project. I am ever grateful for their time and effort working through the proposal plans and providing their insights. In addition, I wish to thank the two people responsible for starting me on the graduate education path, Paul and Diana Whitney. Without them I would not have known the inspiring and fulfilling work of helping others through clinical practice.

This project represents the culmination of a lifelong passion to understand people, the way we think, feel and behave. From a very young age I was wonderfully encouraged to seek out answers to my questions; from library trips to scale models, I was never without opportunities to indulge my curiosities. For that I wish to thank: my parents, Joan and Bernie Burreson, who provided me the encouragement and materials to explore; my older sister, Jenny, and twin sister, Joy, who collaboratively planned “experiments”; and my teachers who gave a little extra and inspired me to go further. This project truly could not have happened without the love and support of my husband, Jay. Since high school we have encouraged and supported each other as we have pursued our dual dream of completing doctorates. This dream has brought many highs and lows, and through it all brought us closer. Because of Jay, and his understanding and encouragement, I was able to spend precious evening hours, usually devoted to our daughter, intensely focused on studying, analyzing, writing and re-writing. He kept our family afloat through the worst of times and celebrated every little occasion with a smile! He gave Emma the attention she needed and answered all her questions with patience. Emma, herself, deserves sincere thanks. For a child of four she has shown great patience and an understanding that far exceeded my expectations. She has great questions and an unending curiosity that I hope to continue to foster, as it was fostered for me. Lastly, I wish to thank all of Jay’s family, who have been a central part of my life since he and I met in high school. They all have been immensely understanding and I am so grateful for all of their support.

VISUAL-SPATIAL SKILLS: A BIOPSYCHOSOCIAL

APPROACH TO PERFORMANCE DIFFERENCES

Abstract

by Jill B. Fancher, Ph.D.
Washington State University
August 2008

Chair: Elizabeth Soliday

Performance differences between men and women on measures of visual-spatial skills are well-documented and have resulted in an extensive literature on types of visual-spatial skills, variations in the observed effect sizes, and methods to moderate the effects. Of particular note, a large and consistent sex differences effect is observed for one particular visual-spatial test, the Vandenberg and Kuse (1978) Mental Rotations Test, where men obtain higher scores than women. For decades the particular source of this effect has been examined with an argument that spans the history of psychological research: Is it nature or nurture? To evaluate this effect from a more integrative perspective, this study manipulated the environmental demands and attempted to control for variations in sex-related biological state (menstrual phase). Results indicate that of six visual-spatial measures spanning three types of tasks, four demonstrated expected sex differences, and one of these tasks (MRT) was negatively affected specifically by inducing stereotype threat in male participants. Unexpectedly, women were not affected by the negative stereotype threat (i.e., explicit instruction of the stereotype), but may have been affected by social experience (female-examiner for female-participant) instead.

TABLE OF CONTENTS

Introduction

Increasing attention has been given to group differences in cognitive tests, particularly related to underperformance by members of traditionally disadvantaged groups, such as women and ethnic minorities, with African-Americans and women being the most commonly studied. Until recently, little was known about the reasons for the discrepancies. Arguments of racial or gender inferiority were popular in the early 1900's and were used in the eugenic "science" of the time (Hothersall, 2004). With the advent of desegregation and women's liberation movements, arguments related more to discrepancies in early environment and available resources. However, the pendulum of the nature and nurture debate continues to swing. As recently as 2005, the Harvard University President at an academic conference attributed the disproportionate number of men compared to women in science and engineering careers wholly to "innate" abilities (Summers, 2005).

While Harvard's former President may have taken one side of the debate based on one set of literature, he overlooked the empirically supported discoveries in social psychology. In the 1990's a new theory emerged from the field of social cognition, documenting underperformance of marginalized groups depending on the way the test materials were presented. Thus a variety of social constructs, including demographic composition of groups being tested (i.e., same ethnicity vs. mixed ethnicity), salience of the individual's minority status (i.e., one woman in a group of men or one woman among other women), and explicit awareness of the negative stereotype, were examined as potential sources of variance for the documented group differences. Steele and Aronson (1995) identified discrepant performance patterns for African-American participants under specific testing conditions. The authors argued that stereotypes can be internalized by the targeted group – in this case, African-

Americans, and lead to a reduction in performance when the stereotype is activated (Steele & Aronson). In their study, Steele and Aronson "activated" the negative stereotype by highlighting the testing situation as a "diagnostic [test] of intellectual ability," which according to their hypothesis would put African-American students at risk of confirming negative societal views that African-Americans are less intelligent than other races. By phrasing the situation as a test of intelligence and the other "non-diagnostic" conditions as "problem-solving" or a "challenge" the difference is in whether the testing situation threatens to support and continue a negative stereotype or is irrelevant to the stereotype.

Congruent with Harvard's President's biological position, psychoneuroendocrinologists have in the past several decades conducted studies focusing on the impact of the biological states, demonstrating the effects of serum hormone levels on specific cognitive functioning. The psychoneuroendocrinologists focused on sex differences and menstrual phase differences.

Early Stereotype Threat Research

In their landmark publication, Steele and Aronson (1995) put forth the theory of stereotype threat and evaluated performance on a verbal task in African-American students compared with Caucasian students. In their first two studies they manipulated the explicit instructions to students creating diagnostic condition, non-diagnostic-only and non-diagnostic challenge groups. The diagnostic condition referred to the depiction of the verbal test as evaluating "personal factors involved in performance" and telling the participants they would receive feedback from this "*genuine test of [their] verbal abilities and limitations*…" (p. 799). Non-diagnostic conditions depicted the task as a method to mentally challenge "highly verbal" individuals and feedback would be provided to create awareness of the types of problems they may encounter in the future. The non-diagnostic challenge condition also encouraged participants to take the task seriously despite the fact they were not being evaluated.

Results from the first study demonstrated that African-American participants had significantly reduced scores in the diagnostic condition compared to the non-diagnostic conditions. Study 2 expanded on these results and examined the role of state anxiety as a mediating variable and how this and the activation of stereotype threat changes self-reported effort, cognitive interference, and time on items. Results paralleled study 1 and identified that African-Americans in the diagnostic condition spent more time on each item than any other group or condition. No effects for condition or ethnicity were found on anxiety, effort or cognitive interference. Together, the Steele and Aronson (1995) studies presented a socially based explanation for group differences on cognitive tasks from variations in task presentation.

This study spurred a new era of research on stereotype threat evaluating the mechanisms, affected groups, and treatments for any domain for which a selected group appears marginalized. This literature has evaluated a variety of cognitive skills as well as non-cognitive domains, such as sports performance. While the Steele and Aronson (1995) study focused on African-Americans and language tests, other historically disadvantaged groups such as women have garnered significant attention in this literature.

Stereotype Threat: Intervening Variables and Mechanisms of Action

Stereotype threat has been evaluated with a variety of mediating and moderating variables. These generally include a social self that ascribes to the negative stereotype (Marx & Stapel, 2006a), importance of the social group membership (Schmader, 2002), personality variables (i.e., agreeableness; Sawyer & Hollis-Sawyer, 2005), working self-concept (Steele & Ambady, 2006) or self-relevance that can be manipulated through first person descriptions of the negative stereotype (Marx & Stapel, 2006c). In addition, high self-monitoring moderates performance resulting in increased performance as a minority in a stereotype condition rather than decreased performance (Inzlicht, Aronson, Good & McKay, 2006).

Different arguments also exist regarding the mechanisms of action for stereotype threat. These arguments include apprehension due to activation of the negative stereotype (Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995), sympathetic nervous system arousal (Ben-Zeev, Fein, & Inzlicht, 2005; O'Brien & Candall, 2003), negative impact on working memory (Beilock, Rydell, & McConnell, 2007; Schmader & Johns, 2003), cognitive load from self-referencing during task completion (Croizet et al., 2004), explicit monitoring of skills that are typically automatic (Beilock, Jellison, Rydell, McConnell & Carr, 2006), and vigilance due to a prevention focus (Seibt & Förster, 2004). A general association with anxiety has been difficult to identify. Bosson, Haymovitz, and Pinel (2004) noted that non-verbal anxiety predicted the stereotype threat effect significantly better than a verbal measure. In addition, the timing of anxiety measurement is important, with pre-testing anxiety measurement demonstrating a stronger association with stereotype threat than post-testing (Marx & Stapel, 2006b). Physiological markers of anxiety have been evaluated, demonstrating changes in skin conductance, blood pressure and skin temperature consistent with increased anxiety during the stereotype threat condition (Osborne, 2007). Aronson et al. (1999) demonstrated that awareness of a negative stereotype is necessary and sufficient to produce the stereotype threat effect. The strength of this effect is moderated by personal identification with the domain assessed. That is, individuals who identify strongly with the domain assessed (e.g., math) will have the strongest reduced performance under stereotype threat.

Stereotype Threat and Women

Stereotype threat has presented a significant explanation for the cognitive differences observed on standardized tests by marginalized groups. With regards to women, the focus has included math and visual-spatial skills. It has been repeatedly demonstrated that women in stereotype threat conditions have lower scores than women in non-stereotype threat conditions on difficult math tests (Ben-Zeev et al., 2005; O'Brien & Crandall, 2003; Spencer et al., 1999). These stereotype threat conditions can either be in the form of being the only woman present among men or in the way the materials are presented (Seqkaquaptewa & Thompson, 2003). When the degree of internalization of the negative stereotype (stigma consciousness) is measured, it is clear that greater stigma consciousness is associated with lower math scores (Brown & Pinel, 2003). The effect of reduced performance on math tests is present regardless of whether the participant expects their scores to be public or private (Inzlicht & Ben-Zeev, 2003). The effect for math performance is most apparent when women place greater importance on gender identity (Schmader, 2002).

The recognized negative stereotype for women and math performance appears to extend to the spatial reasoning domain as argued by McGlone and Aronson (2006) and Martens et al. (2006). When primed with gender as identity, women had lower average scores on the Mental Rotations Test (MRT; Vandenberg & Kuse, 1978) than women in the control condition (McGlone & Aronson, 2006). Priming in these studies refers to focusing the attention of the participant to an aspect of their identity (e.g., gender). Priming is not intended to be undetectable or below the participants' awareness. In addition, when explicitly instructed of the negative stereotype and asked to identify their gender on the answer form, women's scores on the MRT were significantly lower than those obtained by men (Martens et al., 2006). In a study conducted in Turkey, a negative stereotype was not detected for this type of task; thus, it is important to determine and/or create awareness of the negative stereotype before assessing effects of a stereotype threat condition in any given sample as group stereotypes are not consistent across cultures (Halpern & Tan, 2001).

Treatment of Stereotype Threat (Methods to Reduce Underperformance)

Researchers have sought methods to reduce or reverse the effects of stereotype threat. When women were primed to academic success (private school attendance) they performed significantly better than women primed with gender when completing the Vandenberg and Kuse Mental Rotations Test (McGlone & Aronson, 2006) and when completing math problems (McGlone & Aronson, 2007). The term "priming," as noted above, is used to indicate that the participants were explicitly guided to focus on some aspect of their identity. In these studies, priming of gender was accomplished through the use of six probe questions per condition that were modeled after a questionnaire designed by Shih, Pittinsky, and Ambady (1999; e.g., List three reasons why one might prefer living on a coed floor in a dormitory). Using intrinsic (personal identity characteristics rather than social identity characteristics) self- affirmations attenuated the stereotype threat and improved performance for women on math tests (Schimel, Arndt, Banko & Cook, 2004). Martens et al. (2006) demonstrated that the self-affirmations can attenuate the stereotype threat effect for women in the domain of visual-spatial skills. In addition, they demonstrated that using self-affirmations does not universally increase visual-spatial performance because the scores obtained by men did not significantly increase with the self-affirmations.

Psychoneuroendocrinological Theories of Visual-Spatial Task Performance

Sex differences on specific cognitive tasks have been demonstrated consistently over the past several decades and have been reviewed on multiple occasions. One domain of cognitive skills frequently reviewed is the visual-spatial skill domain, consisting of visual perception and mental manipulation (rotation, alignment, closure, assembly, etc.) of objects.

Social psychology studies related to women and visual-spatial skills employed one of the dependent measures included in the trait biological effects (gender differences) and state biological effects (menstrual cycle variations) literature: the Mental Rotations Test (Vandenberg & Kuse, 1978). The MRT is a visual-spatial task that asks the participant to view a three-dimensional arrangement of cubes and imagine it being rotated in three-dimensions. To the right of the stimulus are four choices. Two of these choices are accurate representations of the object rotated the other two are objects that are similar, but not the same as the original stimulus object because of variations in design not rotation. In a meta-analytic review of sex differences in visual-spatial skills, Linn and Petersen (1985) noted that the strength of the sex differences effect on visual-spatial skills varies by type of skill. However, the direction of the effect consistently favors men. Voyer, Voyer, and Bryden (1995) replicated this sub-typing and found that the strongest differences were in mental rotation type tasks, whereas, the most inconsistent data are from spatial visualization type tasks.

Data from meta-analyses of sex differences in visual-spatial skills have supported the concept that the term “visual-spatial” is too broad and that further specifying the concept results in clearer similarities of effects across studies. Specifically, dividing visual-spatial skills into three conceptual categories, then analyzing these results by age groups results in significant and near-significant homogeneity of effects (Voyer et al., 1995; Linn & Petersen, 1985). These categories as defined by Linn and Peterson were mental rotations, spatial visualization, and spatial perception (see Table 1).

Mental rotation was defined as the use of imagination to quickly and accurately rotate figures in three dimensional space. In the college population, the effect size of mental rotation type tasks between sexes was d = .66 (Voyer et al., 1995), with scores favoring male participants. Effects varied across specific measures used. Specifically, the estimated effects size of the Card Rotation Test (Ekstrom et al., 1976) between sexes is d = .31 (Voyer et al.). In contrast, the Mental Rotations Test when scored using the 20 point maximum procedure results in an effect of d = .94 (Voyer et al.). The difference in effect sizes within the domain of mental rotations has been a source of debate with theories ranging from difficulty of 3D tasks over 2D tasks to the Hunter-Gatherer theory as posited by Silverman and Eals (1994), which argues that 3D visual-spatial processing is selected for in male populations because their survival has been more closely tied to life or death decisions regarding the predator or prey status of an object.

The spatial perception tests are those measures that require determination of spatial relationships (i.e. orientation) when provided distracting information (Linn & Petersen, 1985). The estimated sex differences effect size of college student performance on measures of spatial perception is d = .48 (Voyer et al.). This is relatively consistent across types of spatial perception measures (e.g., Water Level Test and Rod-and-Frame Test).

Lastly, Linn and Petersen (1985) described spatial visualization tests as a complex, multi-step manipulation of spatial cues. The estimated sex differences effect size of the college student performance on these measures was small d = .23 (Voyer et al.). Evaluation of the effect size for particular spatial visualization measures demonstrates minimal effects between sexes (Block Design, *d* = .17; Embedded Figures Test, *d* = .18; Voyer et al.). One possible explanation for the small overall effect of this class of visuospatial skill is that results on performance of these tasks are highly inconsistent.

In addition to the trait theories of visuospatial performance differences between men and women, psychoneuroendocrinology also lends a state theory of fluctuating hormones influencing performance. While general results from this literature are inconsistent, women tend to perform better on Mental Rotations Tasks during specific hormone phases, specifically during menstruation (see Fancher, unpublished manuscript for a review). It has been argued that sexually dimorphic cognitive skills are more likely to show hormone related effects than non-sexually dimorphic skills (Hampson, 1990). Thus, at times of low estrogen and progesterone (menstruation), women may be expected to perform better on visual-spatial tasks than at times of high estrogen and/or progesterone (luteal phase).

Social Psychology and Psychoneuroendocrinology: Complementary Theories?

Biopsychosocial theory is an underutilized explanation in the sex differences effects for visual-spatial tasks. Social cognition research tends to view the sex differences effect through the lens of socially based stereotype threat and the various intervening variables and cognitive mechanisms of action. In contrast, psychoneuroendocrinologists tend to view the effect through the lens of trait and state biological effects. Dichotimizing interactional processes, such as nature and nurture, results in a skewed understanding of a process where cause and effect are circular (Halpern & Tan, 2001). Thus, integrating both theoretical perspectives may provide greater explanatory power. In addition, gaining greater insight into the role of the sex of the test administrator in individual sessions, the constitution of group administered tests, and individually held negative stereotypes would illuminate a possible source of the inconsistencies on spatial visualization tasks (e.g., Hidden Figures) identified in the hormone literature. To extend previous research and test an integrated theoretical model on sex differences in visual-spatial skills, the following research questions and hypotheses were developed:

Research Questions

Question 1- Mental Rotations and Spatial Perception Tasks. Does knowledge of a negative stereotype (stereotype threat activation) contribute significantly to the variance on the mental rotations and spatial perception types of tasks after statistically accounting for trait biology (sex of the participant) and methodologically holding state biology constant (menstrual status in women)? Given that prior research documents some effect of fluctuating hormones on performance, holding menstrual state constant may reduce error and improve interpretation of the variables of interest. Thus, this study intends to isolate the variances due to the social construct of stereotype threat and the biological trait of sex of the participant.

Hypothesis 1.1 - Interaction Effects for Mental Rotations. Given the previous research on stereotype threat and a mental rotations type task it is hypothesized that women in the stereotype threat condition will have significantly lower scores as compared to men and women in the control stereotype threat condition. To date, the literature has not shown men to be negatively affected by stereotype threat conditions for visual-spatial or math skills, when compared to women. Susceptibility to stereotype threat appears to require some basis in known stereotypes. This was demonstrated when Caucasian men had lower math test scores in the stereotype threat condition as compared to scores of Caucasian men in the control condition when the comparison group of Asian men was brought to awareness (Aronson et al., 1999). Thus, it is expected that men will not demonstrate a stereotype threat effect.

Hypothesis 1.2 - Interaction Effects for Spatial Perception. Given the consistent documentation of sex differences on these types of tasks with a moderate effect size it is hypothesized that, similar to mental rotations tasks, stereotype threat will account for significant variance. Therefore, with stereotype threat activation it is expected that women will have significantly lower scores when compared to men and women in the control group.

1.3 - Main Effect of Sex. Given the long-standing literature and supporting meta-analyses it is expected that the main effect for sex will be significant with men performing significantly better than women.

1.31 - Effect Size of the Main Effect of Gender. Given the meta-analyses documenting the various effect sizes for the three classifications of visual-spatial skills it is expected that the effect size (R^2) will be larger for the mental rotations tasks than the spatial perception tasks.

1.4 - No Main Effect of Condition. Because an interaction effect is hypothesized and a main effect of sex is hypothesized, no main effect of condition is expected. Therefore, it is expected that across men and women there will be no global effect of condition.

Research Question 2 - Spatial Visualization (Exploratory). Does stereotype threat contribute significantly to the variance of the spatial visualization type tasks after accounting for gender of the participant?

There are considerable inconsistencies in the literature relating sex differences and spatial visualization tasks, such as Hidden Figures (Ekstrom et al., 1976). An effect, if any is considered “small”. This type of task has also not been evaluated with stereotype threat. The question remains whether there are state or trait biological effects for this type of visual-spatial skill task and/or is there a stereotype threat effect for this type of task. Therefore, tests on these measures are considered exploratory with the goal of detecting any interaction effects, and/or main effects of sex.

Methods

Purpose and Study Design

The purpose of this study was to evaluate the multiple sources of variance contributing to the visual-spatial sex differences effect. Recognizing and estimating the contributions from biological, psychological, and social constructs provides more depth of interpretation than single source studies (i.e., either biology or social psychology). A 2 (sex; male or female) x 2 (condition; control or stereotype threat) factorial design was employed. Given the potential for other independent variables to mediate/moderate the outcomes, several potential covariates were examined (i.e., anxiety, depression, estimated intelligence, and number of years in school). To address the biopsychosocial construct and test the relative contributions by the factors (sex, stereotype threat condition, and the interaction of sex and stereotype threat), a hierarchical linear regression analytic plan was used.

Procedures

Participants and recruitment. Ninety-one participants were recruited through classroom announcements in courses at Washington State University Vancouver. This study was reviewed and approved by the Institutional Review Board at Washington State University. Participants were compensated with extra credit by their instructors and $5 upon the completion of the 1.5 hour testing session. If the course did not offer extra credit (primarily non-psychology courses), students were compensated with a $15 gift card to a local department store instead (N = 14). There were no significant differences on any outcomes depending on the compensation (independent samples t-tests; *p*‘s >.10). Women signed up to participate at a time they expected to be menstruating. Menstruation was confirmed via self-report at the time of testing. Due to various factors (including irregular menstruation, menopause, and difficulty recalling the first day of last month's period) many women were not menstruating when they presented for testing appointments. Therefore, hierarchical regression analyses were conducted to determine whether menstruating women vs. those not menstruating at the time of testing performed differently on dependent measures. Menstruation was not a significant predictor of the visual-spatial outcome measures, nor did menstruation significantly interact with condition (p's >.10; see Preliminary Analysis Section for further detail.)

Testing sessions and stereotype threat induction. All participants were scheduled with a test administrator of the opposite sex. This is consistent with prior studies (e.g., Aronson & Steele, 1995) that controlled the social environment, including the identity (i.e., ethnicity or gender) of the test administrators. While control or measurement of the social environment is a key component in social psychological design, few studies reported the sex of the test administrator. However, in a study that closely matches the design of the current study, male test administrators (for mixed-sex groups) were employed. Due to a greater number of available female participants and a limited number of male volunteers, we tested a small number of women with female administrators (N = 12). The data from these participants were not included in the primary analyses due to this variation in procedure which reduced the overall sample size for analysis to 79 (32 men and 47 women). Participants were asked to report gender, age, and education level (number of years of college education). Participants were screened for state depression and anxiety symptoms using the Clinical Epidemiological Scale for Depression (Radloff, 1977) and Beck Anxiety Inventory (Beck & Steer, 1993).

All procedures were conducted in individual sessions consisting of one examiner and one participant. This procedure is a deviation from previous stereotype threat studies in which tests were administered in groups. Utilizing one-to-one testing sessions was necessary to follow standardized administration for clinical neuropsychological measures. In addition, this design standardizes the social experience across participants. Furthermore, group administration does not permit interpretation of the impact of the same-sex participants, opposite-sex participants, and/or the sex of the administrator. Following informed consent procedures, which included a description of the measures, the time commitment, confidentiality, and compensation, all participants were administered a measure to estimate general intelligence, followed by the screening measures for depression and anxiety and a brief questionnaire asking if participants were currently menstruating. Demographic information was collected as part of the “general instruction” form that is used to establish the control and stereotype threat conditions.

Participants were randomly assigned to either the stereotype threat condition or the control condition, using the publicly available Research Randomizer (Urbaniak & Plous, 2008). Control and stereotype threat conditions were established after collecting baseline measures using one of two “general instruction” forms, each of which contained information related to the participant’s assigned condition. The description of the purpose of the study and the follow-up task varied by condition (see Table 2 for the phrasing of each instruction page). Participants were asked if they had any questions about this form before proceeding. After reading the general instruction form, the examiners presented each of six visual-spatial measures by individually reading the instructions aloud. The test order of the visual-spatial measures was counterbalanced using the 3 x 3 Latin Square approach to account for order effects. After completing the measures, the examiner debriefed the subjects by reading a script. Prior to the verbal debriefing all participants were asked what they believed was the hypothesis of the current study. No participant accurately identified the stereotype threat purpose. However most individuals in the stereotype threat conditions identified the purpose as stated in their respective forms (e.g., "to find out why women do better on these tasks"). Participants in the control condition had a wide variety of responses (e.g., "to find out how people do on tests when they are being timed"). The examiner then proceeded to debrief the participant by using a script.

Control condition. To create consistent testing pressure between stereotype threat and control conditions, this condition began with a general instructions form. The general instruction form identified the purpose of the study in technical terms, instructed the participant that the tasks would be challenging, and identified the test as a direct measure of their visual-spatial intelligence (see Table 2). This form also requested basic demographic information (gender, age, and education level). In addition to the general instructions, the participants were asked to do a short written task which instructed them to list reasons why the region they live is different than another region of the United States. This latter statement is consistent with activities of the other conditions and does not focus the participant to gender. A similar procedure was used in McGlone and Aronson (2006).

Stereotype threat condition. To create the stereotype threat condition this study followed standard procedures to activate stereotype threat (e.g., Steele & Aronson, 1995; Martens et al., 2006; McGlone & Aronson, 2007). The general instructions discussed for the control condition were amended to include a statement that the opposite sex generally performs better on these measures. The use of explicit information on the stereotype has been used successfully in prior research to ensure the participants are aware of the negative stereotype (Martens et al., 2006). To prime their gender status, the participants were asked to do a short written task instructing them to list reasons why men and women are different, and describe what it is like to either be a man or a woman.

Measures

Estimate of intellectual functioning. To test for potential confounding effect of general intelligence, participants were administered the North American Adult Reading Test (NAART; Blair & Spreen, 1989). This test consists of reading aloud 50 irregularly spelled words (e.g., debt, debris); total errors were calculated and entered into an equation to predict full scale IQ. The equation, 127.8-.78(NAART errors) = eFSIQ, was developed by Blair and Spreen (1989) for comparison with the WAIS-R. The total time to administer this test is approximately five minutes. To minimize interference from the instruction forms to the start of the actual visual-spatial measures, the NAART was administered before administration of the “general instruction” form and the visual-spatial measures. According to Spreen and Strauss (1998) the NAART is “among the most reliable tests in clinical use”. Internal consistency as measured by Chronbach’s α is high and recently measured at .93 (Uttl, 2002). In addition, Uttl tested the construct validity of the NAART and identified a high (r = .78) correlation between the NAART and a measure of verbal intelligence (WAIS-R Vocabulary Subtest).

Mental Rotations measures. The Vandenberg & Kuse Mental Rotations Test (1978) is a paper-and-pencil test that presents a stimulus 3D figure on the left column of the page and four similar options to the right. The participant was asked to choose two figures from the right that can be rotations (not mirror images) of the stimulus on the left. The MRT has two parts each with a maximum time limit of 3 minutes. The MRT is the most widely used measure of mental rotation skills in the hormone literature and recently in the stereotype threat literature. The score is based on the total number correct minus 25% of the number incorrect to account for guessing. Internal consistency as measured by Kuder-Richardson 20, a measure of inter-item consistency, is .88 (Wilson et al., 1975). In addition, Vandenberg and Kuse tested the construct validity of the MRT and identified higher correlations with other mental rotations tests (i.e., CRT; $r = .58$) and lower correlations with verbal measures such as Vocabulary ($r = .00$) and Verbal Reasoning ($r = .07$).

The Card Rotations Test (CRT; Ekstrom et al, 1976.) is similar to the MRT in that the participant has stimulus figures on the left and rotations of the figures on the right. The participant was asked to determine if each of the eight figures on the right is the same or different from the stimulus. To be the same, it must not be shaped differently or be a mirror image. The CRT has two parts, each with a three minute time limit. The score is the number right minus the number incorrect. Internal consistency for the CRT is .80 (Ekstrom et al.).

Spatial Perception measures. The Water Level Test (Piaget & Inhelder, 1956) was presented as eight bottles in various orientations, each on a separate piece of paper. Participants were instructed to draw a line representing the surface of the water in each bottle. Participants were shown a demonstration of the bottle at 0°. The angles of presentation were those used by Thomas, Lohaus and Kessler (1999) and Thomas and Lohaus (1993), which included a randomized presentation of 30°, 60°, 120°, 150°, 210°, 240°, 300°, and 330°. Thus, the Water Level Test asked the participant to identify a line orientation in spite of visually conflicting orientation of the angle of the bottle. Lines were scored using a protractor to measure the number of degrees their line deviates from the horizontal axis. Scoring utilized by

Vasta, Rosenberg, Knott, and Gaze (1997) was to sum the total number of degrees from horizontal for all items. This task requires approximately 5 minutes to administer. Standard reliability and validity data are not available for this measure as it is constructed by each laboratory based on the previously published guidelines. Regardless, it is a widely used measure, identified as a measure of spatial perception in over 30 studies (Voyer et al., 1995).

The Judgment of Line Orientation (JLO; Benton, Hamsher, Varney, & Spreen, 1983) is a clinical measure of spatial perception. According to Strauss, Sherman and Spreen (2006), the JLO has been used and accurately identified impairment with clinical populations (i.e., individuals with dementia of the Alzheimer Type, Turner Syndrome, Parkinson's Disease and/or visual neglect). The participant was presented one page at a time with two lines at various orientations and lengths that correspond to two of the lines on a 180° array. This task requires approximately 6 minutes to administer. Split-half reliability for Form V was found to be .89 (Benton, Varney, & Hamsher, 1978). Validity, as measured by correlations with WAIS-R visual-spatial subtests (Block Design $r = .68$ and Object Assembly $r = .69$), is higher than correlations with WAIS-R verbal subtests (i.e., Vocabulary $r = .28$; Trahan, 1998).

Spatial Visualization measures. Spatial visualization measures include multi-step spatial processing that may include other components (e.g., motor skills). The Hidden Figures Test (HFT; Ekstrom et al., 1976) is a test of recognizing embedded figures. It involves visually deconstructing components of complex visual stimuli. This paper-and-pencil test is separated into two parts, 2 pages per part, each part allowing for 12 minutes to complete. On the top of each page are five simple outline objects. Underneath these five objects are several complex patterns of lines. Within each complex pattern is one of the five objects from above. Every pattern has only one object. The participant tried to determine which object is embedded in each complex pattern. The score is the total correct minus a fraction of those marked incorrectly. Internal reliability on this measure is reported at .80 (Ekstrom et al.).

The Paper Folding Test (PFT; Ekstrom et al., 1976) task required individuals to imagine a paper has been folded in a specific manner and punctured with a pencil then unfolded. The participant chose which representation of the unfolded paper with holes would match the stimulus image. This task has two parts each allowing for 3 minutes to complete. Internal consistency for the PFT is .76 with validity estimate of .46, based on correlations with other spatial tasks (Visser, Ashton & Vernon, 2008).

Results

Preliminary Analyses

Normality: Measures of central tendency & skewness. Initial examination of data included a review of the Mean, Median, and Skewness for each dependent variable. MRT, CRT, JLO, PFT, and HFT all present with consistent measures of central tendency (i.e., Means and Medians were within 1-2 points; see Table 3). In contrast the WLT had a Mean of 107.97 and a Median of 45.00. Measurement of skewness for dependent variables were all between -1.0 and 1.0, except the WLT which had a measure of skewness of 1.55. Thus, tests of normality indicated normal distribution for all of the dependent variables except the WLT, which presents with a strongly skewed sample and bimodal distribution. To account for the non-normal distribution on the WLT, a log transformation was performed on those scores. After the transformation, the Mean was 1.76, Median was 1.65 and Skewness was .16. These transformed WLT scores were used in all subsequent analyses.

Group composition. To determine if groups differed by age, education, anxiety, depression, or estimated intelligence, separate independent samples t-tests were performed for each group (sex and condition). There were no significant differences between men and women on any of the above measures (p's >.05). There were also no significant differences between groups in the control condition and participants in the stereotype threat condition on measures of potential mediator/moderator variables (p's >.05; see table 4).

Potential covariate effects on dependent variables. Pearson correlation coefficients were calculated to determine if age, education, anxiety, depression, or estimated intelligence significantly correlated with any of the six dependent variables. Age significantly correlated with the JLO ($r = .31$, $p < .01$). Education significantly correlated with the PFT ($r = .23$, $p < .05$). Estimated intelligence was not significantly correlated with any measure (p's >.05). Depression, as measured by the CES-D, did not significantly correlate with any dependent variable (p's >.05). Anxiety, as measured by the BAI, was significantly correlated with the CRT ($r = .30$, $p < .01$). Therefore, the BAI was used as a covariate in subsequent CRT analyses, age was used as a covariate with the JLO, and education was used with the PFT.

Confirmation of dependent variable constructs. Given that each visual-spatial measure was chosen either because of previous documentation or theoretical fit into one of the three sub-types of visual-spatial skills, all six were correlated to determine the strength of association between measures. Two measures (those chosen to fit the spatial visualization sub-type) not only correlated highly with each other (PFT-HFT, $r = .50$, $p < .001$), but these two measures correlated significantly with all four of the other visual-spatial measures with the absolute value of the correlations ranging from .28 to .49 (p's <.01). Thus, these spatial visualization tasks correlated significantly with other visual-spatial tasks indicating they share some common constructs, but are within a range that indicates these also measure a unique construct. The fact that the measures correlated significantly with each other and correlated well with all the other visual-spatial tests argues that they indeed represent an integration of the visual-spatial skills, including skills assessed on the other measures. Other than correlations with the spatial visualization tests, the measures of mental rotations tests correlated highly within type (MRT-CRT, $r = .49$, $p < .001$), as was expected, and less well with measures from the spatial perception tasks (MRT-JLO, $r = .16$, $p = .17$; MRT-WLT, $r = -.28$, $p <.05$; CRT-JLO, $r = .26$, $p < .05$; CRT-WLT, $r = -.27$, $p <.05$) again, as expected. Lastly, the spatial perception tasks significantly correlated with each other (WLT-JLO, $r = -.31$, $p <.01$). Thus, the measures chosen for the visual-spatial constructs appear to measure hypothesized constructs adequately independently of one another.

Effects of menstruation and condition. Of the 47 women who participated in the study, 17 were successful in scheduling their appointment during menses. Although we did not have the means to measure the hormonal state of each female participant (which would have provided verification of hormonal status and quantification of hormonal levels for analysis), the number of women participating overall permitted an analysis of the effects of menstruation on the six visual-spatial tasks.

To determine the strength of effect of menstrual cycle and the interaction of menstrual cycle and stereotype threat condition on the dependent variables, hierarchical regression analyses were conducted. Separate equations were conducted for each dependent variable. For the equation with MRT as the outcome, menstruation was entered as step 1 and the interaction of stereotype threat condition and menstruation was entered in step 2. For the equations with HFT as the outcome, and the equations with WLT as the outcome, predictor variables were entered in the same manner as the MRT. For the equation with JLO as the outcome, age was entered into step 1, menstruation was entered into step 2 and the interaction of menstruation and stereotype threat condition was entered into step 3. For equations with the PFT and CRT as outcomes, predictor variables were entered in the same manner as for the JLO; thus, for JLO, PFT, and the CRT appropriate covariates (see preliminary analysis section) were entered in Step 1 followed by the standard predictors in Steps 2, and 3. Other than expected covariates, no other independent variables significantly predict visual-spatial performance (*p*'s > .10). Therefore, in an equation testing effects of menstrual status, menstruation does not appear to significantly impact performance on these measures.

Primary Analyses

To evaluate the explanatory power of each independent variable on the dependent variable, hierarchical regression analyses were performed. Hierarchical regression analysis was utilized because of its ability to not only identify main and interaction effects, but to evaluate the collective effects and effects of one variable after accounting for a potentially interrelated variable (Pedhazur, 1982). Variables were coded using standard coding procedures necessary to test for categorical variable effects in linear multiple regression equations.

Mental Rotations Test. A hierarchical regression analysis was conducted to determine how trait biology (sex), social knowledge (negative stereotype), and their interaction effect predicted three dimensional mental rotations performance. The outcome was MRT score and the predictor variables were sex, stereotype threat condition, and their interaction. The overall regression equation was significant $R^2 = .24$, F (3,75) = 7.90, p <.01. Sex was a significant unique predictor, accounting for 18% of the variance (ß = .50), with men obtaining higher scores overall on the MRT (t (78) = 4.65, p <.01), as was predicted. The mean score for men was M =22.21, SD= 8.07 and for women was M = 15.77, SD = 6.09. Also as was predicted, stereotype threat condition was not significant as a unique predictor (*t* (78) = 1.05, *p* = .30). The interaction of sex and stereotype threat condition uniquely predicted MRT scores, accounting for an additional 6% variance (*t* (78) = 2.45, *p* <.05; ß = .27). Although the interaction effect was predicted, the direction of effect was contrary to hypotheses: results indicate that men, not women, obtained lower scores when in the stereotype threat condition (see Figure 1). Planned contrasts for the interaction indicate that in the control condition men had significantly higher scores than women (*t* (34) = 4.66, *p* < .001). In the stereotype threat condition the scores between men and women were not significantly different (*t* (41) = 1.2, *p* = .24). Among women scores were not significantly different between the stereotype threat conditions and the control conditions (*t* (45) = -1.22, *p* = .23). Males scores in the stereotype threat condition were significantly lower than scores obtained by men in the control condition (*t* (30) = 2.01, *p* =.05). See table 5 for Beta weights and R^2 change for each predictor for $Y_{MRT} = A + B_{sex} + B_{condition} + B_{interaction}$.

Card Rotations Test. A hierarchical multiple regression analysis was conducted to determine how well biology (sex), social knowledge (negative stereotype), and their interaction effect predict a two-dimensional mental rotations task. The outcome was CRT score and the predictor variables were BAI, sex, stereotype threat condition, and the interaction of sex and stereotype threat. The overall regression equation was significant R^2 = .15, *F* (4,74) = 3.26, *p* <.05. As expected based on preliminary analyses the BAI was a significant predictor and accounted for 9% of the variance (*t* (78) = 3.16, *p* <.01; ß = .35). Sex was a significant unique predictor, accounting for an additional 5% of the variance (ß = .27), with men obtaining higher scores overall on the CRT (*t* (78) = 2.28, *p* <.05), as predicted. The mean score for men was *M* =105.09, *SD*= 32.13 and for women was *M* = 96.17, *SD* = 21.46.

Also as predicted stereotype threat condition was not a significant predictor (*t* (78) = .16, *p* = .88). The interaction of sex and stereotype threat condition was not a significant predictor of CRT scores (*t* (78) = .90, *p* = .38). In contrast to the hypothesis, the sex X stereotype threat interaction did not predict CRT scores. See table 6 for Beta weights and R^2 change for each predictor for $Y_{CRT} = A + B_{BAI} + B_{sex} + B_{condition} + B_{interaction}$.

Judgment of Line Orientation. A hierarchical multiple regression analysis was conducted similar to the MRT and CRT. The outcome was JLO score and the predictor variables were age, sex, stereotype threat condition, and the interaction of sex and stereotype threat. The overall regression equation was significant, $R^2 = .23$, F (4,74) = 5.58, p <.01. As expected based on preliminary analyses, age was a significant predictor and accounted for 10% of the variance (*t* (78) = 2.93, *p* <.01; ß = .30). Sex was a significant unique predictor, accounting for an additional 12 % of the variance, with men obtaining higher scores overall on the CRT (*t* (78) = 3.30, *p* <.01; ß = .36), as predicted. The mean score for men was *M* =27, *SD*= 2.17 and for women was *M* = 24.57, *SD* = 3.64. Also as predicted, stereotype threat condition was not a significant predictor (*t* (78) = -1.07, *p* = .29). The interaction of sex and stereotype threat condition was not a significant predictor (*t* (78) = -.10, *p* =.92). In contrast to an interaction hypothesis, the interaction of sex and stereotype threat did not predict the performance on the JLO. See table 7 for Beta weights and R^2 change for each predictor for $Y_{JLO} = A + B_{Age} + B_{sex} + B_{condition} + B_{interaction}$.

Water Level Test. A similar hierarchical multiple regression analysis was conducted. The outcome was WLT score and the predictor variables were sex entered in step 1, stereotype threat condition entered in step 2, and the interaction of sex and stereotype threat entered in step 3. The overall regression analysis was not significant, $R^2 = .06$, F (3,75) = 1.56, $p = .20$). No further analyses were warranted.

Hidden Figures Test. In this exploratory analysis, a hierarchical multiple regression analysis was conducted on this measure of spatial visualization. The outcome was HFT score and the predictor variables were sex, stereotype threat condition, and the interaction of sex and stereotype threat. The overall regression equation was not significant $R^2 = .06$, F (3,75) = 1.70, $p = .18$. Therefore, no further analyses were warranted.

Paper Folding Test. In this exploratory analysis, a multiple regression equation was conducted on this measure of spatial visualization. The outcome was the PFT score and the predictor variables were education, sex, stereotype threat condition, and their interaction. The overall regression equation was significant, $R^2 = .13$, F (4,74) = 2.64, $p < .05$. As predicted from the preliminary analyses, numbers of years in college was a significant predictor and accounts for 5% of the variance (t (78) = 2.09, $p < .05$ ß = .23). Sex was a significant predictor, accounted for an addition 5 % of the variance, with men obtaining higher scores overall (t (78) = 2.23, $p < .05$). The mean score for men was M =11.40, SD= 3.61 and for women was M = 9.76, SD = 3.56. Stereotype threat condition was not a significant predictor (t (78) = .87, p =.39). The interaction of sex and stereotype threat condition was also not a significant predictor (t (78) = .95, p = .35). See table 10 for Beta weights and R^2 change for each predictor for $Y_{PFT} = A + B_{sex} + B_{condition} + B_{interaction}$.

Post-Hoc Analyses on Female Participant Data

Subsequent analysis of the female participants was necessary to examine possible reasons why, in contrast to McGlone and Aronson (2006) and Martens et al. (2006), current female study participants were apparently unaffected by the stereotype threat condition/methodology. Similar to McGlone and Aronson (2006) and Martens et al. (2006), there was a significant interaction between sex and stereotype threat condition on the MRT in the current study. However, the stereotype threat condition appeared to affect men and not women, contrary to hypotheses, and was not apparent on any of the other five visual-spatial outcome measures.

Timing of the menses questionnaire in the procedure. To evaluate whether the timing of the menstrual questionnaire affected the outcomes, a small group of women (N = 8) was asked about menses following the visual-spatial measures (as opposed to prior to, as occurred in N = 39 participants). Independent samples t-tests for the 6 dependent variables comparing women who were given questions about menses prior to the "general instruction" form (questionnaire before group) and those women who were given the menses questionnaire following the entire study (questionnaire after group) were conducted. All of the *p* values exceeded 0.1, indicating that the timing of questions about menses did not affect results for this subsample.

Method for inducing stereotype threat. To examine how social comparison with same-sex test administrators may produce different results than social comparisons made with opposite-sex test administrators, a small group of women was tested (N = 12) by women rather than by men. No significant results were obtained for HFT, PFT, WLT, or the JLO. Results from the CRT and MRT were significant, $t\ (57) = 3.01$, $p < .001$ and $t\ (57) = 2.57$, $p < .01$, respectively. On the MRT women tested by men ($M = 15.77$, $SD = 6.09$) scored higher than women tested by women ($M = 11.00$, $SD = 3.87$; see Figure 3). The effect size was large ($d = .93$). The same was true for the CRT with women tested by men ($M = 96.17$, $SD = 21.46$)

having higher scores than women tested by women (*M* = 61.33, *SD* = 37.60; see Figure 4), with a large effect size (*d* = 1.14).

Discussion

The purpose of this study was to integrate social psychological and biopsychological theories in evaluating the potential variables associated with well-documented differences in visual-spatial performance between men and women. Divergent lines of research have separately evaluated the cause of the visual-spatial sex differences. Utilizing a biopsychosocial framework, sources of variance were evaluated based on biological traits (men/women) and social constructs (negative stereotypes) after accounting for biological state (menses) and psychological state (mood/anxiety). Regression analyses provided data on the relative explanatory power of specific variables, contributing value information how both social and inborn factors relate to visual-spatial skill performance.

Hypotheses were based on previous literature, meta-analyses on sex differences, and social psychology studies that utilized one measure of visual-spatial processing, the Vandenberg and Kuse Mental Rotations Test (MRT). Large sex differences have consistently been identified with the MRT, which has also been sensitive to stereotype threat effects in previous studies (McGlone & Aronson, 2006). However, the MRT represents only one type of visual-spatial skill, mental rotation. Therefore, measures from other types of visual-spatial skills and another measure of mental rotation were utilized to clarify reported sex differences on the broad category of tests known as visual-spatial tests.

The three visual-spatial constructs as defined by Linn and Petersen (1985) were supported through correlational analyses. Specifically, the measures of mental rotations and spatial perception both correlated better within visual-spatial skill subtype than with other measures of presumably distinct subtypes. In addition, the spatial visualization subtests correlated significantly with all other visual-spatial measures, but within a range that indicates that these measures do indeed represent their own distinct subtype. Given the fact that these spatial visualization tasks (PFT and HFT) measure a combination of visual-spatial skills and they generally lack consistency in detecting group differences, the place of these measures in future research is limited.

In support of a biologically based difference between men and women as a factor in visual-spatial performance, men performed better than women on four of six visual-spatial measures. Only the HFT and WLT performance were not related to sex of the participant, which was unexpected but not entirely without precedent. In the Voyer et al. (1995) meta-analysis the WLT was identified as a measure of Spatial Perception in 30 studies in their review, with as many as eight of those failing to identify significant sex differences. Given the heterogeneity of the WLT tests due to each laboratory designing its own stimuli based on descriptions, models, and methods in previously published studies, it is difficult to know if this inconsistency is a result of stimulus design and implementation or the nature of the construct studied. Furthermore, it argues for standardization of the measure if it continues to be used in visual-spatial research. The null result for the HFT, a spatial visualization task, is not unexpected given the highly inconsistent results from measures in this general sub-category of visual-spatial tests.

Although the outcome measures correlated better within visual-spatial sub-type than between visual-spatial sub-types, significant variability existed within each sub-type with regards to the amount of variance accounted for by sex of the participant. That is, Mental Rotations Tasks (MRT and CRT) combined had the most variance accounted for by the sex of the participant and within this domain the MRT had the strongest effect with 18% of the variance accounted for by sex of the participant contrasted to the 5% accounted for by sex on the CRT after accounting for the effects of state anxiety. The MRT has historically had the highest effect size for sex differences (Voyer et al., 1995), providing a clear advantage over other visual-spatial measures with good reliability and validity and a strong literature documenting its development, utilization, and consistency in distinguishing groups. The MRT clearly is unique, demonstrating the largest sex differences in this study and demonstrated to respond to explicit instructions of a negative stereotype. Therefore, the MRT should continue to be included in future visual-spatial and stereotype threat studies.

Within the Spatial Perception sub-type of visual-spatial skills, the JLO was predicted by sex with 12% of the variance accounted for by this predictor after accounting for age. Although the WLT, the other Spatial Perception task, was significantly correlated with the JLO demonstrating shared constructs, the WLT was not significantly predicted by any variables of interest. The WLT appeared to be a task on which participants appeared to either understand or not understand the concept. Scores were bimodally distributed, with the largest percentage of respondents performing as would be expected of individuals who clearly understood the Piagetian concept and obtaining scores skewed towards zero (Mean of 107.97 and a Median of 45.00; see Figure 2). Because of this distribution, scores on the WLT were logarithmically transformed. Lastly, the Spatial Visualization tasks inconsistently identified sex differences. The HFT was not significantly predicted by sex of the participant, whereas 5% of the PFT was accounted for by sex of the participant after accounting for education.

Men were not expected to be affected by the stereotype threat methods across measures because it was expected that the current sample would perform consistently with other studies, whose hypotheses were based on the sample having had some previous experience with the respective stereotype under investigation. However, in contrast to the hypothesized direction of the interaction effect for the mental rotations tasks, men, not women, in the stereotype threat condition obtained the lower scores. This effect was only noted for the one test with the largest sex effect, the MRT. This result is not consistent with prior research. However, previous research has relied on the assumption that negative stereotypes regarding performance on visual-spatial skills affected women because presumably the women would have some prior experience with the negative stereotype and would "know" that their performance was poorer compared to men. It is possible that the men in this study were naïve to the stereotype regarding a male advantage on visual-spatial measures; therefore, they responded to explicit instructions by performing worse when the negative stereotype was explicitly stated.

Female Participants' Results

Given the failure to detect or induce a stereotype threat effect in women, further exploration was warranted to help direct future research. With the purpose of this study designed to examine the multi-dimensional sources of the visual-spatial sex effect, two factors considered primary are reviewed: state biology (menses) and social constructs (environmental variations).

Across all visual-spatial tasks, conditions, and their interactions, there was no demonstrable effect from the female participant biological state (menstrual status). This contrasts a growing literature in which women performed better on visual spatial tasks during menstruation theoretically due to lower levels of female hormone during that phase. This inconsistency may reflect the differences in measurement methods, a hormonal effect detectable only through direct quantification and not by indirect biological markers such as menses, or simply that the effect is not detectable across these specific visual-spatial measures. While measurement using self-report of menstruation as a biological marker for hormone changes does not provide a quantification of the hormones, the fact that this biological marker does not predict visual-spatial performance provides some evidence that argue against the socially held negative stereotypes regarding menstruation and cognition. If that is indeed the case, then these results may imply that the hormone literature using quantifiable methods may not be widely applicable to the general testing and/or clinical situation because quantification of hormones is not a standard of clinical practice, and the biological markers are too global. Further research exploring these visual-spatial tasks with quantifiable hormone levels, self-report of the woman's experience, comparability with easily observable biological markers, such as menses, and effect size estimates would be instructive. In addition, these findings should be replicated with larger, homogenous samples.

The design of any laboratory experiment establishes certain environmental variations. Specifically, two social demands were considered relevant in this study and may provide some explanatory power for the failure to induce/detect a stereotype threat effect in women. These social demands were: 1) asking about menses and, 2) using same-sex or opposite-sex test administrators. As noted above, menstruation does not appear to specifically influence the current outcomes. However, asking about menstruation may have inadvertently primed the female participants to be more focused on the negative stereotypes of women during menses rather than being a woman in general. Separately, the explicit instruction methodology for inducing stereotype threat in women must be explored as an alternative explanation for the null results. In isolating the various methodological components that contribute to the stereotype threat effect, this study employed one-to-one testing situations with opposite sex administrators. Thus, the effect of explicit instruction was isolated and found not to affect the performance by women. Therefore, the factors necessary to induce the stereotype threat effect in women may involve other components, such as the sex of the administrator, same-sex group testing situations, and/or mixed-sex group testing situations.

Results from the timing of the questionnaire that asked about menses (before or after the visual-spatial measures) indicate that timing of the questionnaire is not a likely reason for these results. In contrast, the second environmental variation, opposite-sex or same-sex test administrators may have had some effect. Results from this follow-up data indicate that women who were tested by women showed a reduction in test scores overall, as compared to women tested by men. In previous stereotype threat studies using visual-spatial measures, participants have typically been tested in mixed-sex groups led by men. In those studies that had only male leaders, specifying the necessary and sufficient factors to induce the effect is impossible – in other words, were the women in the stereotype threat condition affected by the male administrator, male participant presence, female participant presence, or the knowledge of the negative stereotype threat. In the one-to-one testing situation, these factors can be manipulated/controlled individually. The fact that women tested by women responded in a manner consistent with the negative stereotype threat may stem from a social competition that occurs when a person of the same sex is present. It is unknown whether a similar effect may exist for male participants when tested in a one-to-one situation, an important question in further research.

Social psychologists may argue that female participants were not naïve to the negative stereotype regarding women and visual-spatial skills, therefore women in the control condition were similarly affected by the negative stereotype threat effect as those who were explicitly reminded about the negative stereotype. That is, there may be a floor effect in this stereotype and inducing additional threat beyond that already inherent across all participants had little effect. Though we did not systematically test participants' perceptions of the study's true hypotheses, informal questioning indicated that none of the women in the control condition identified sex comparison or the fact that women have lower scores as a part of the hypothesis. Thus, there was no indication that women in the control condition were aware of the negative stereotype and were affected by it without the explicit instructions. In contrast, social psychologists may separately argue that in the presence of men – in this case, the male test administrator –, women are less concerned if they under perform as other members of their group (women) would not know they have "supported" the negative stereotype, whereas in the presence of other women the threat of supporting the negative stereotype is stronger and induces more anxiety. Indeed, this may be true for men as well and should be explored.

Furthermore, social psychologists have identified the role of identification with the domain assessed (i.e., visual-spatial skills) as a mediating variable. It may be that undergraduate women majoring in psychology place a low priority on development or status of their visual-spatial processing; however, this does not explain the differences observed with same-sex test administrators or why male participants would place greater value on their visual-spatial performance. Thus, there may be two distinct value systems mediating the outcomes: 1) personal value on performance within a domain and, 2) value of performance because of sensitivity to group stereotypes. That is, if an individual has a strong identification with the domain assessed the introduction of a negative stereotype may impact their performance regardless of concern about the group stereotype because of the personal value they place on their performance in that domain. On the other hand, if the individual is particularly sensitive to any stereotype about their group, they may also be affected by a negative stereotype threat because of the fear of supporting a negative view of their group, regardless of their personal value on the skills assessed. With regards to the male participants who were primarily recruited from psychology and other social science and liberal arts courses, the value placed on their performance may not be due to the actual domain assessed but rather that predominant female cohorts in these majors may selectively sensitize them to male and female stereotypes. In addition, the women who have more common experience with the female cohort may not be as affected by the male administrator, but more so by the female administrator because of the fear of supporting the negative stereotype.

Evolutionary theorists may argue that male participants would benefit from demonstrating their skills in the presence of women, known as the “show-off” hypothesis (Hawkes, 1991). Therefore, stereotype threat may elicit greater anxiety in these opposite-sex testing situations when they are instructed that there is a real likelihood they will not perform well. Whereas women would gain little benefit from “showing up” the male administrator, but may experience competition from a female participant and want to demonstrate superiority, thus, inducing anxiety and reduced scores. An evolutionary theoretical perspective may also provide some explanation for the significant sex-related differences observed on the MRT in contrast to the smaller differences observed with the other three visual-spatial measures identifying a group difference. Specifically, the need to rapidly decipher visual-cues in our three-dimensional world to determine the presence of potential prey and/or predictor was necessary for survival (Silverman, Choi, & Peters, 2007). In addition, two-dimensional processing, although present in extreme early art representations pre-dating written language, is less related to survival and a less practiced skill; hence, no sex differences on these tasks.

Feminist psychologists may argue that women in higher education are experienced in personally deflecting/diminishing negative stereotypes and do not readily conform or are easily influenced by explicit instructions about a negative stereotype. Therefore, they would work hard in the presence of men to demonstrate they are skilled at the current task, whereas in the presence of other women, they may have less to prove and therefore do not exert the same effort. In addition, the women in this sample were recruited from a commuter campus, with many of the students balancing other life/work demands, thus, developing confidence in their management of multiple responsibilities and resilience to negative stereotypes.

While data gathered in the current study can not identify the exact mechanism of observed effects, it is clear that follow-up studies on stereotype threat should include attention toward the sex of the administrator, one-to-one versus group testing, group composition, awareness of the negative stereotype, personal values on performance of domains assessed, and sensitivity towards any negative stereotype for their group. With current cultural climate placing greater emphasis on reducing sex-differences in areas like math and visual-spatial skills, a shift towards smaller sex-differences may become more common with time. Although the research on this effect is rather current (within the current decade), the historical trend and parental influences may be just starting to shift the sex differences data.

While this study was conducted in a laboratory setting, clinical situations are similarly affected. In one-to-one neuropsychological testing situations effort is exerted to standardize administrations, maintain rapport and maintain motivation. Internalized negative stereotypes that become personal expectations based on an individual's identified group (i.e., stereotype --

a woman isn't good at numbers; individual response -- therefore I can't be good at numbers) can have an impact on performance and result in the self-fulfilling prophecy phenomenon. Furthermore, as the current data suggest, the sex of the test administrator in the one-to-one testing situation is a variable that may be having a greater impact on results and requires further clinical investigation. Therapeutically, belief in a stereotype may also impact progress, (i.e., "criminals never change") resulting in reduced effort to change.

The laboratory setting also extends to small educational settings with same-sex or mixed-sex teachers and classmates. In these educational settings, belief in stereotypes can also lead to anxious responding and/or reduced efforts. It has been documented that because of negative stereotypes a person may purposely reduce their identification with the domain (i.e., academics) so that failure can knowingly be attributed to reduced effort and not personal deficiency (Aronson et al., 1999).

Exploring the "sex effect" for visual-spatial skills is ripe for misinterpretation and can lead to unintended propagation of a negative stereotype. Although the outcomes of this study identify various contributing factors to the effect (biological trait, social knowledge, and social experience) the contributions from biological trait tends to attract negative attention. Thus, it is important to put the sex differences into a broader context. It is not accurate to say that all women performed worse than all men on all visual-spatial measures. Indeed, two measures did not find a significant sex difference. The fact remains that individual differences are far greater than the group differences, with some men obtaining lower scores than some women and vice versa. Thus, on the individual level no assumptions can be made about the expected performance by a man or a woman. However, at the group level, knowledge of the differences can encourage educators to include multi-modal teaching strategies so that visual information (graphs, geometry, art, etc.) are supplemented with verbal and tactile learning opportunities, thereby, allowing men, women, boys, and girls to experience supportive learning opportunities.

Limitations exist in all research. In the current study, recruitment of male participants was difficult. By focusing early recruitment on psychology courses, which contain larger numbers of women compared to men, this limitation was expected and managed through the use of direct compensation for non-psychology students. However, the lower rates of recruitment and eventual sample reduced power to detect small effects and may explain some lack of findings. While other studies in stereotype threat effect find various sizes of the effect, the fact that this study controlled the social experiences, tested only the social knowledge component of stereotype threat, and tested the effects on a lesser known negative stereotype, it is likely that the strength of this effect is smaller than the effect for other better known negative stereotypes when using both social experience and social knowledge to elicit the effect. It should be noted, however, that in the current study the *p* values for the non-significant primary analyses did not approach significance. Furthermore, if more men had been recruited it would have permitted an exploration of same-sex (male administrator – male participant) effects. In addition, the effort to have testing occur during menses was challenging and was not accomplished for a sizeable number of women. Rather than controlling the state biology through methodology, the study statistically evaluated the potential effects from this source and found no effects. As the study was not designed to measure state biology, but rather control for it, hormone status was not quantifiably measured.

In conclusion, the results from the current study highlight the relatively young state of stereotype threat theory with much of the effect still to be explored. Where the current methodology was expected to produce effects in women, it conversely only affected men. In addition, changing the sex of the administrator significantly impacted the performance by women; such an effect remains to be tested in men. Individual test administration may have an effect as well. The impact from deviations from previous stereotype threat protocols necessary to test a range of visual-spatial skills highlights the need to understand the necessary and sufficient factors to produce the effect. Further investigations drawn from the current findings indicate that the effects of negative stereotypes may be dependent on the population, type of task, sex of the "authority" figure in relation to the subject, sensitivity to stereotypes, identification with domain assessed, and prior awareness/experience with a negative stereotype.

References

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35,* 29-46.

Beck, A.T., & Steer, R. A. (1993). *Beck Anxiety Inventory Manual.* San Antonio, TX: Psychological Corporation.

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin, 32,* 1059-1071.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136, 2,* 256-276.

Benton, A. L., Hamsher, K. deS., Varney, N. R., & Spreen, O. (1983). *Contributions to neuropsychological assessment.* New York: Oxford University Press.

Benton, A. L., Varney, N. R., & Hamsher, K. deS. (1978). Visuospatial judgment. A clinical test. *Archives of Neurology, 35,* 364-367.

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

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: A revision of the National Adult Reading Test. *The Clinical Neuropsychologist, 3,* 129-136.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge:

The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology, 40*, 247-255.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology, 39,* 626-633.

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

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Manual for kit of factor referenced cognitive tests*. Princeton, New Jersey: Educational Testing Service.

Fancher, J. B., & Soliday, E. (2008). *Visual-spatial abilities across the menstrual cycle: A clinical epidemiological review.* Unpublished manuscript.

Halpern, D. F., & Tan, U. (2001). Stereotypes and steroids: Using a psychobiosocial model to understand cognitive sex differences. *Brain and Cognition, 45,* 392-414.

Hampson, E., 1990. Estrogen-related variations in human spatial and articulatory-motor skills. Psychoneuroendocrinology, 15, 97-111.

Hawkes, K. (1991). Showing off: Tests of an hypothesis about men's foraging goals. *Ethology & Sociobiology, 12*, 29-54.

Hothersall, D. (2004). *History of Psychology* (4th ed.) New York: McGraw-Hill.

Inzlicht, M., Aronson, J. Good, C., & McKay, L. (2006). A particular resiliency to threatening environments. *Journal of Experimental Social Psychology, 42,* 323-336.

Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95,* 796-805.

Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479-1498.

Martens, A., Johns, M., Greenverg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. *Journal of Experimental Social Psychology, 42,* 236-243.

Marx, D. M. & Stapel, D. A. (2006). It depends on your perspective: The role of self-relevance in stereotype-based underperformance. *Journal of Experimental Social Psychology, 42,* 768-775.

Marx, D. M., & Stapel, D. A. (2006). It's all in the timing: Measuring emotional reactions to stereotype before and after taking a test. *European Journal of Social Psychology, 36,* 687-698.

Marx, D. M., & Stapel, D. A. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology, 91,* 243-254.

McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology, 27,* 486-493.

McGlone, M. S., & Aronson, J. (2007). Forewarning and forearming stereotype-threatened students. *Communication Education, 56,* 119-133.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29,* 782-289.

Osborne, J. W. (2007). Linking stereotype threat and anxiety. *Educational Psychology, 27,* 135-154.

Pedhazur, E. J. (1982). *Multiple Regression in Behavioral Research 2nd* Ed. Fort Worth, TX: Holt, Rinehart, & Winston.

Piaget, J., & Inhelder, B. (1956). *The Child's Conception of Space.* London: Routledge and Kegan Paul.

Radloff, L.S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385-401.

Ryan, J. J., & Paolo, A. M. (1992). A screening procedure for estimating premorbid intelligence in the elderly. *The Clinical Neuropsychologist, 6,* 53-62.

Sawyer, T. P. & Hollis-Sawyer, J. A. (2005). Predicting stereotype threat, test anxiety, and cognitive ability test performance: An examination of three models. *International Journal of Testing, 5,* 225-246.

Schimel, J., Arndt, J., Banko, K. M., & Cook, A. (2004). Not all self-affirmations were created equal: The cognitive and social benefits of affirming the intrinsic (vs. extrinsic) self. *Social Cognition, 22,* 75-99.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology, 38,* 194-201.

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

Seibt, B., & Fö rster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology, 87,* 38-56.

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of Experimental Social Psychology, 39,* 68-74.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10,* 81-84.

Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Archives of Sexual Behavior, 36,* 261-268.

Silverman, I. & Eals, M. (1994). The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology & Sociobiology, 15,* 95-105.

Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35,* 4-28.

Steele, J. R., & Ambady, N. (2006). "Math is hard!" The effect of gender priming on women's attitudes. *Journal of Experimental Social Psychology, 42,* 428-436.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and intellectual performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-811.

Strauss, E., Sherman, E., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms and commentary. Third Edition.* Oxford University Press.

Summers, L. H. (2005). Remarks at NBER conference on diversifying the science & engineering workforce. Retrieved April 10, 2007, from Harvard University, Office of the President Web site: http://www.president.harvard.edu/speeches/ 2005/nber.html.

Thomas, H., & Lohaus, A. (1993). Modeling growth and individual differences in spatial tasks. *Monographs of the Society for Research in Child Development, 58,* i-159.

Thomas, H., Lohaus, A., & Kessler, T. (1999). Stability and change in longitudinal water-level task performance. *Developmental Psychology, 35,* 1024-1037.

Trahan, D. E. (1998). Judgment of line orientation in patients with unilateral cerebrovascular lesions. *Assessment, 5,* 227-235.

Urbaniak, G. C. & Plous, S. (2008). Research randomizer [Computer software]. Retrieved from http://www.randomizer.org

Uttl, B. (2002). North American Adult Reading Test: Age norms, reliability, and validity. *Journal of Clinical and Experimental Neuropsychology, 24*, 1123-1137.

Vandenburg S. G. & Kuse, A. R. (1978). Mental rotation: A group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47,* 599-604.

Vasta, R., Rosenberg, D., Knott, J. A., & Gaze, C. E. (1997). Experience and the Water Level Task revisited: Does expertise exact a price? *Psychological Science, 8*, 336-339.

Visser, B.A., Ashton, M.C., & Vernon, P.A. (2008). What makes you think you're so smart? *Journal of Individual Differences, 29,* 35-44.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250-270.

Wilson, J. R., DeFries, J. C. McClearn, G. E., Vandenberg, S. G., Johnson, R. C., & Rashad, M. N. (1975). Cognitive abilities: Use of family data as a control to assess sex and age differences in two ethnic groups. *International Journal of Aging and Human Development, 6*, 261-276.

Table 1

Subtypes of Visual-Spatial Skills and Representative Measures Employed in the Current Study

Subtype	Measures
Mental Rotations	Mental Rotations Test (MRT)
	Card Rotations Test (CRT)
Spatial Perception	Water Level Test (WLT)
	Judgment of Line Orientation (JLO)
Spatial Visualization	Hidden Figures Test (HFT)
	Paper Folding Test (PFT)

Table 2

General Instructions Provided Before Task Administration

Control Condition

You are about to complete a series of tasks designed to better help us understand the components of different visual-spatial skills. These tasks are designed to be challenging. Please do your best as this is a genuine measure of your visual-spatial intelligence and will be used for future psychological research. Before you begin please complete the following demographic information and this short written task. On the paper below list 6 reasons why it is different to live in the Pacific Northwest than in the Southern United States. And in a few short sentences describe what it is like to live here.

Stereotype Threat Condition*

You are about to complete a series of visual-spatial tasks designed to better help us understand why men (women) perform better than women (men) on these measures. These tasks are designed to be challenging. Please do your best as this is a genuine measure of your visual-spatial intelligence and will be used for future psychological research. Before you begin please complete the following demographic information and this short written task. On the paper below list 6 reasons why men and women are different. And in a few short sentences describe what it is like to be a woman (man).

* If it is a male participant, wording indicates women perform better. The converse is true for female participants.

Table 3

Means, Medians, & Skewness for Each of the Dependent Variables (N=79)

	Mean (SD)	Median	Skewness
MRT	18.38 (7.61)	16.75	.582
CRT	99.78 (26.48)	101.00	-.46
JLO	25.56 (3.33)	26.00	-.98
WLT	107.98 (122.45)	45.00	1.55
HFT	12.51 (7.19)	11.75	.37
PFT	10.42 (3.65)	9.75	-.02

Table 4

Demographic Variables Means and Standard Deviations by Group (N=79)

	Age	Education	BAI	CES-D	eFSIQ
Women	25.41(8.67)	14.68(1.12)	9.83(8.71)	11.70(8.52)	108.02(7.02)
Men	25.78(7.78)	14.59(1.13)	7.00(4.51)	7.00(4.51)	106.18(18)
ST	25.79(7.66)	14.58(1.14)	8.65(6.83)	10.65(7.33)	106.07(17.78)
Control	25.40(9.05)	14.72(1.11)	8.72(8.13)	10.75(7.64)	108.71(6.92)

* $p < .05$

Table 5

Summary of Hierarchical Regression Analysis for Variables Predicting 3D Mental Rotations Performance (MRT; N = 79)

Variable		*B*	*SE B*	ß	R^2
Step	1				.18*
	Sex of Participant	3.22	.80	.42*	
Step	2				.18*
	Sex of Participant	3.16	.81	.41*	
	Stereotype Threat Condition	.326	.53	.06	
Step	3				.24*
	Sex of Participant	3.88	.835	.50*	
	Stereotype Threat Condition	.55	.52	.11	
	Interaction	1.28	.52	.27*	

* $p < .01$

Table 6

Summary of Hierarchical Regression Analysis for Variables Predicting 2D Mental Rotations Performance (CRT; N = 79)

Variable		*B*	*SE B*	β	R^2
Step	1				.09**
	BAI	1.07	.39	.30**	
Step	2				.14**
	BAI	1.22	.39	.34**	
	Sex of Participant	6.19	2.90	.23*	
Step	3				.14**
	BAI	1.22	.39	.34**	
	Sex of Participant	6.19	2.95	.23*	
	Stereotype Threat Condition	1.65E-03	1.91	.00	
Step	4				.15*
	BAI	1.24	.39	.35**	
	Sex of Participant	7.19	3.15	.27*	
	Stereotype Threat Condition	.30	1.93	.02	
	Interaction	1.73	1.93	.11	

* p <.05

** p <.01

Table 7

Summary of Hierarchical Regression Analysis for Variables Predicting 2D Line Orientation (JLO; N = 79)

Variable		*B*	*SE B*	ß	R^2
Step	1				.10*
	Age	.13	.04	.31*	
Step	2				.22*
	Age	.12	.04	.30*	
	Sex of Participant	1.19	.34	.35*	
Step	3				.23*
	Age	.12	.04	.30*	
	Sex of Participant	1.24	.34	.37*	
	Stereotype Threat Condition	-.24	.23	-.11	
Step	4				.23*
	Age	.12	.04	.30*	
	Sex of Participant	1.22	.37	.36*	
	Stereotype Threat Condition	-.25	.23	-.11	
	Interaction	-.2.30E-02	.23	-.01	

* *p* <.01

Table 8

Summary of Hierarchical Regression Analysis for Variables Predicting Line Representation of Water Levels (WLT; N = 79)

Variable		*B*	*SE B*	ß	R^2
Step	1				.05
	Sex of Participant	-.11	.06	-.21	
Step	2				.06
	Sex of Participant	-.12	.06	-.23	
	Stereotype Threat Condition	3.97E-02	.04	.12	
Step	3				.06
	Sex of Participant	-.12	.06	-.24	
	Stereotype Threat Condition	3.77E-02	.04	.12	
	Interaction	-1.14E-02	.04	-.04	

Table 9

Summary of Hierarchical Regression Analysis for Variables Predicting Dis-embedding Figures from Within Complex Patterns (HFT; N = 79)

Variable		*B*	*SE B*	ß	R^2
Step	1				.06*
	Sex of Participant	1.77	.80	.24*	
Step	2				.06
	Sex of Participant	1.78	.82	.24*	
	Stereotype Threat Condition	-4.36E-02	-.54	-.01	
Step	3				.06
	Sex of Participant	1.59	.88	.22	
	Stereotype Threat Condition	-9.98E-02	.55	-.02	
	Interaction	-.32	.55	-.07	

* $p < .05$

Table 10

Summary of Hierarchical Regression Analysis for Variables Predicting Spatial Visualization of Paper Folding and Hole Punches (PFT; N = 79)

Variable		*B*	*SE B*	ß	R^2
Step	1				05*
	Education	.79	.37	.23*	
Step	2				.11*
	Education	.82	.37	.24*	
	Sex of Participant	.89	.42	.23*	
Step	3				.11*
	Education	.80	.37	.24*	
	Sex of Participant	.85	.42	.22*	
	Stereotype Threat Condition	.20	.27	.08	
Step	4				.13*
	Education	.77	.37	.23*	
	Sex of Participant	1.00	.45	.26*	
	Stereotype Threat Condition	.25	.28	.10	
	Interaction	.27	.28	.11	

* p <.05

Figure 1. Interaction Effect between Sex and Stereotype Threat Condition on the MRT.

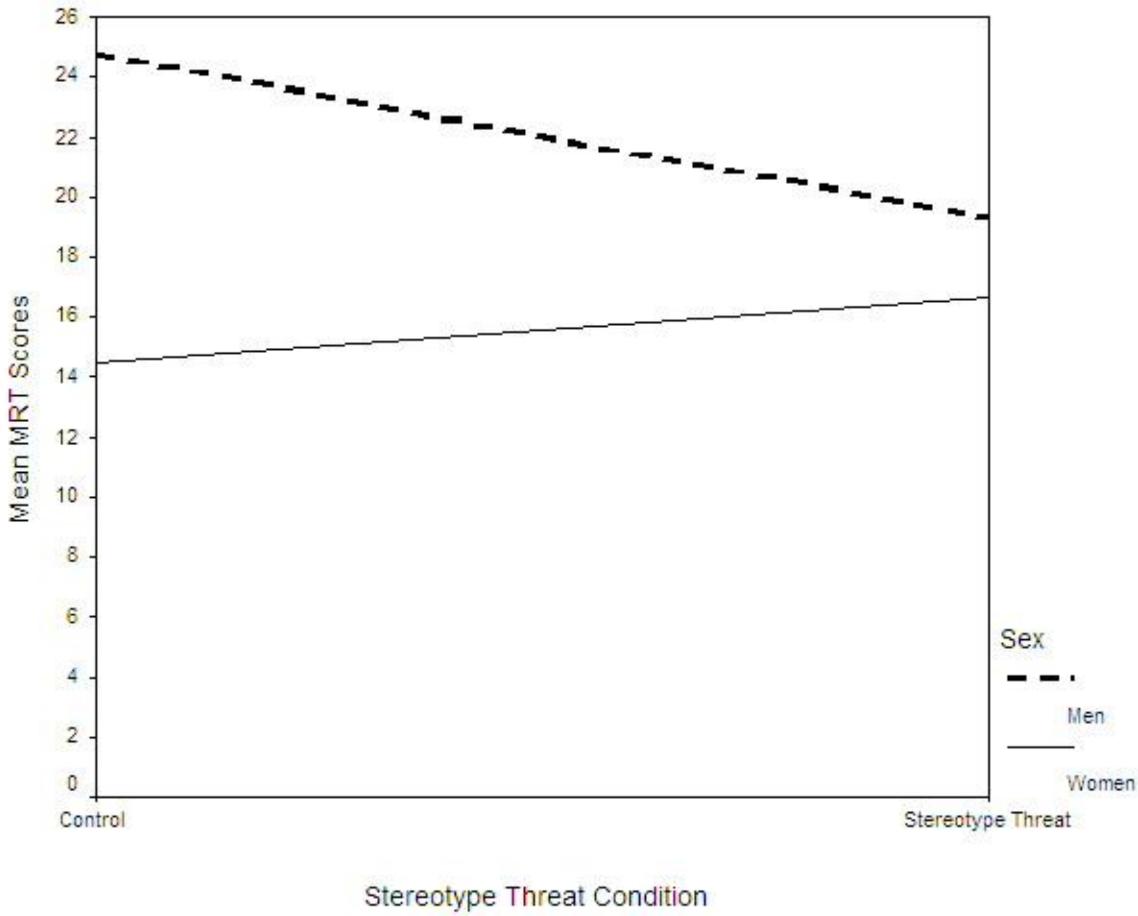

Figure 2. Frequency of WLT scores demonstrating a skewed distribution

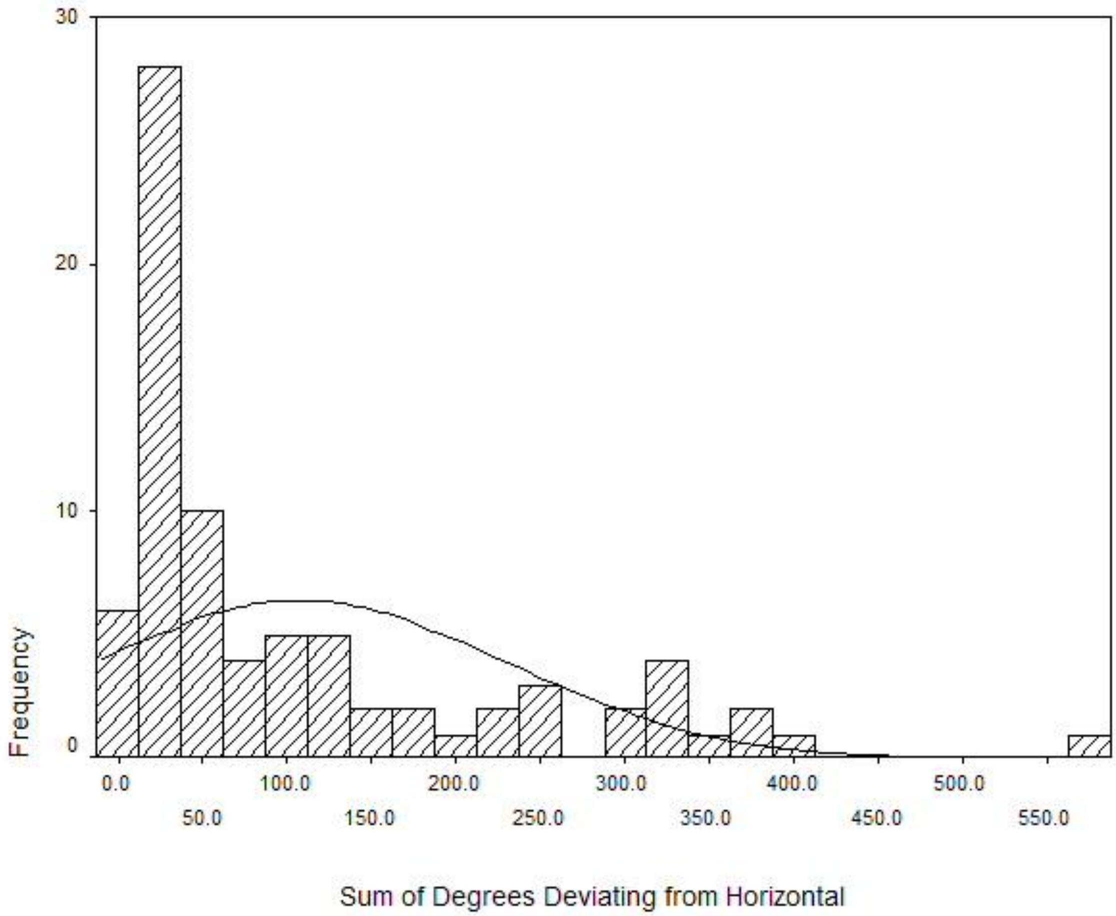

Figure 3. Results for the MRT of Women Participants Tested either by Men or Women.

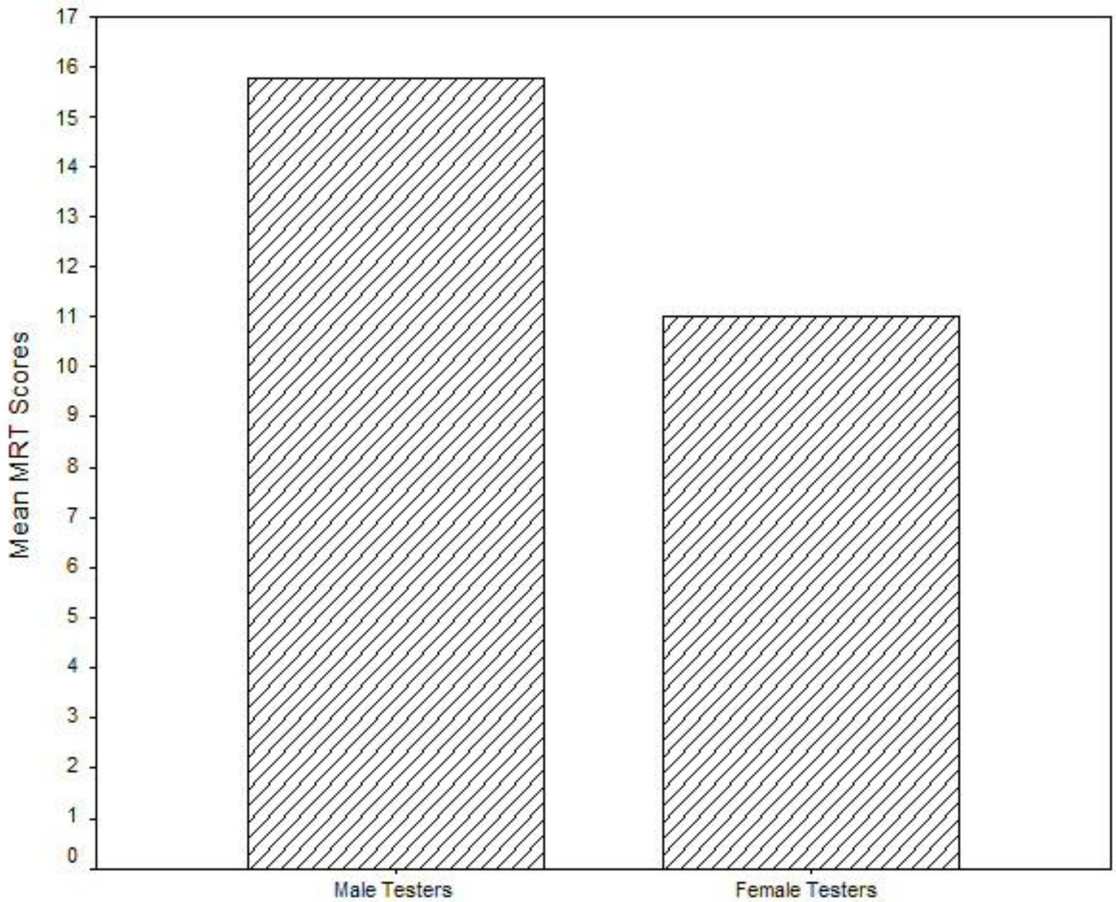

Figure 4. Results for the CRT of Women Participants Tested either by Men or Women.