# AN EXAMINATION OF STEREOTYPE THREAT EFFECTS ON KNOWLEDGE ACQUISITION IN AN EXPLORATORY LEARNING PARADIGM

By

James Grand

## A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

## DOCTOR OF PHILSOPHY

Psychology

### ABSTRACT

## AN EXAMINATION OF STEREOTYPE THREAT EFFECTS ON KNOWLEDGE ACQUISITION IN AN EXPLORATORY LEARNING PARADIGM

By

## James Grand

Stereotype threat describes the situation where an individual is faced with the risk of upholding a negative stereotype about a subgroup to which that person belongs based on his/her actions (Steele & Aronson, 1995). Empirical investigations of stereotype threat effects across a variety of individuals, subgroups, and contexts have identified a number of undesirable consequences related to performance on domain-relevant tasks (e.g., Nguyen & Ryan, 2008; Steele, 1997; Steele & Aronson, 1998; Steele, Spencer & Aronson, 2002). Efforts to identify the psychological mechanisms and processes most directly affected by stereotype threat have indicated that one of its most detrimental influences is exerted on individuals' working memory capacity. More specifically, the added cognitive and emotional-regulatory strain introduced by the presence of stereotype threat uses up a portion of one's limited working memory capacity, thus "hijacking" cognitive resources that could otherwise have been put towards completing task-relevant activities/performance (Schmader, Johns, & Forbes, 2008).

Given the importance of working memory to the development of new knowledge, skills, and abilities (cf., Feldman Barrett, Tugade, & Engle, 2004), the primary goal of the present investigation was to extend and build upon recent research examining the acquisition of taskrelevant knowledge by individuals facing conditions of stereotype threat during their learning activities (Rydell, Rydell, & Boucher, 2010; Rydell, Shiffrin, Boucher, Van Loo, & Rydell, 2010). Guided by an empirically grounded taxonomy of critical learning outcomes (Kraiger, Ford, & Salas, 1993), the knowledge organization and development of task strategies were examined for 145 female learners assigned into either stereotype threat or control conditions. Individuals were tasked with learning to operate a low-fidelity computer-based radar tracking simulation over the course of three experimental sessions held on consecutive days. Based on principles of active learning (Bell & Kozlowski, 2008), the presentation of content/materials followed an exploratory learning paradigm which facilitates task comprehension through the use/improvement of learners' inferential reasoning capabilities (e.g., McDaniel & Schlager, 1990).

Key findings of this study indicated that, unlike females who learned the task under control conditions, females facing stereotype threat experienced the greatest difficulty acquiring effective heuristics critical to improving task performance. Examination of participants' knowledge structures revealed that although female learners under stereotype threat were capable of deducing advanced relations amongst relevant task concepts over time, they appeared to do so in a manner that was far less efficient and, consequently, less conducive to performance when required to apply their knowledge in more demanding task conditions. Further analyses indicated that females under conditions of stereotype threat were not only less accurate at applying their learned knowledge to task-critical decisions, but the manner in which they had learned to interpret information presented to them in the task was generally also less optimal. Lastly, the observed pattern of results revealed that the above effects did not manifest immediately during initial onset of learning activities and required time for meaningful differences to emerge, suggesting that longitudinal examinations of stereotype threat effects are an important direction for future research. Copyright by James Grand 2012

## ACKNOWLEDGEMENTS

I would like to thank my parents (Barry and Kathy), brother (William), and sister (Lacy) for the support and encouragement they have provided, and continue to provide, in everything I do; I could not have asked for a more loving or caring family. I would also like to thank Jennifer Wessel for her friendship throughout our graduate school career, as well as the personal and professional relationship that has emerged from that; I am forever grateful for the happiness and fun you bring to my life. Finally, I would like to thank the members of my dissertation committee—Ann Marie Ryan, Steve Kozlowski, Neal Schmitt, Tim Pleskac, and Georgia Chao—whose investment in my education and the challenges they have pushed me to take on have left a lasting impact which will not soon be forgotten.

# **TABLE OF CONTENTS**

LIST OF TABLES	viii
LIST OF FIGURES	X
INTRODUCTION	1
An Overview of Stereotype Threat: Theory, Consequences, and Criticisms	5
Nature of stereotype threat	6
Outcomes of stereotype threat.	
Criticisms of stereotype threat.	
Stereotype Threat at Learning: Rationale, Applications, and Implications	
Conceptual background for stereotype threat at learning.	
Examinations of stereotype threat at learning	
Research/practical implications of stereotype threat at learning.	
Research Hypotheses	
Stereotype threat and knowledge organization.	
Stereotype threat and cognitive strategy	
Stereotype threat and performance.	55
METHOD	
Participants	57
Experimental Task	58
Procedure	
Online signup.	
Experimental sessions.	
Task introduction and familiarization trial.	65
Practice trials	66
Performance trial	67
Exploratory learning recommendations.	69
Experimental manipulation	71
Measures	
Demographics.	
Cognitive ability	
Math domain identification.	74
Working memory.	74
Metacognitive activity.	
Manipulation checks.	77
Declarative knowledge	
Knowledge structure assessment.	
Strategic learning behaviors	
Decision-making strategy.	
Task performance.	87

RESULTS	
Descriptive Statistics and Data Cleaning	
Manipulation Check	
Knowledge Structure Analyses	
Knowledge structure similarity	
Knowledge structure correlation	
Knowledge structure coherence	
Number of knowledge structure links	
Knowledge structure clustering	
Cognitive Strategy Analyses	
Strategic learning behaviors	
Knowledge acquisition behaviors	
Task practice behaviors	
Self-regulation	
Decision-making strategy	
Task Performance	
DISCUSSION	
Summary of Key Findings	
Stereotype Threat Effects on Knowledge Organization	
Stereotype Threat Effects on Cognitive Strategy Acquisition	
Stereotype Threat Effects on Task Performance	
Implications and Directions for Future Research	
Study Limitations and Generalizability	193
Conclusion	
FOOTNOTES	
APPENDICES	
APPENDIX A	
APPENDIX B	
APPENDIX C	
APPENDIX D	
APPENDIX E	
APPENDIX F	
APPENDIX G	
APPENDIX H	
APPENDIX I	
APPENDIX J	
APPENDIX K	
REFERENCES	

# LIST OF TABLES

Table 1. Summary of Rydell and Colleagues' Multi-study Experiments Examining Stereotype      Threat Effects at Learning      3	1
Table 2. Total Sample Size and Attrition Rates across Days by Sex and   Experimental Condition	8
Table 3. Subdecision Outcomes and Relevant Identifying Information Cues/Values      in TANDEM      6	51
Table 4. Rules of Engagement for Determining Final Engagement Decisions 6	52
Table 5. Summary of Experimental Session Sequence and Timings 6	4
Table 6. Distribution of Target Characteristics Across all Scenarios for Practice Trial Targets $(n = 63)$ and Performance Trial Targets $(n = 126)$	58
Table 7. TANDEM Knowledge Concepts with Descriptions	'9
Table 8. Relative Probabilities Shared between Subdecision Outcomes and Final      Engagement Decision Outcomes      8	32
Table 9. Means, Standard Deviations and Interrcorrelations for Study Variables 9	0
Table 10. MRCM Parameter Estimates for Female's Knowledge Structure Similarity with Males and the Top 15 Performers (Hypotheses 1 & 6)	1
Table 11. MRCM Parameter Estimates for Female's Knowledge Structure Correlationwith Males and the Top 15 Performers (Hypotheses 2 & 7)11	3
Table 12. MRCM Parameter Estimates for Female's Knowledge Structure Coherence (Hypotheses 3 & 8)	4
Table 13. MRCM Parameter Estimates for Number of Links in Female's Knowledge      Structures (Hypotheses 4 & 9)	5
Table 14. MRCM Parameter Estimates for Graph Theoretic Metrics (Exploratory Analyses) 12	8
Table 15. MRCM Parameter Estimates for Time Spent on Task Manual Sections   (Hypothesis 11)	3
Table 16. MRCM Parameter Estimates for Task Practice Behaviors (Hypothesis 11)    13	8

Table 17.	MRCM Parameter Estimates for Female's Metacognitive Activity (Hypothesis 11)	142
Table 18.	MRCM Parameter Estimates for Performance Outcomes Measured during Learning/Practice Trials (Hypothesis 13 & 14)	162
Table 19.	MRCM Parameter Estimates for Performance Outcomes Measured during Performance Trials (Hypothesis 13 & 14)	164
Table 20.	MRCM Parameter Estimates for Performance on the Declarative Knowledge Assessments	167
Table 21.	Hypothesis Summary	169

# LIST OF FIGURES

Figure 1. Conceptualization of stereotype threat as a cognitive imbalance triggered by person and/or situation factors (Schmader, Johns, & Forbes, 2008)	7
Figure 2. Integrated process model of stereotype threat effects on performance (Schmader, Johns, & Forbes, 2008)	. 13
Figure 3. Classification system for learning outcomes (adapted from Kraiger, Ford, & Salas, 1993)	. 24
Figure 4. TANDEM graphical user interface	. 60
Figure 5. Sequencing of exploratory learning recommendations and manipulation instructions during daily practice rounds	. 70
Figure 6. Cognitive strategy heuristic for TANDEM performance	. 86
Figure 7. Knowledge structures for female participants in the stereotype threat and control conditions averaged across days	117
Figure 8. Average knowledge structures for female participants in the stereotype threat and control conditions at end of Day 1	118
Figure 9. Average knowledge structures for female participants in the stereotype threat and control conditions at end of Day 2	119
Figure 10. Average knowledge structures for female participants in the stereotype threat and control conditions at end of Day 3	120
Figure 11. Cumulative average time spent viewing manual pages during learning trials	131
Figure 12. Average time spent viewing manual pages during learning trials	132
Figure 13. Female's average task practice behaviors across learning trials	140
Figure 14. Observed and optimal decision weights for Type cue (Surface) on decision to Warn rather than Clear targets for stereotype threat females and control females at each day	151
Figure 15. Observed and optimal decision weights for Type cue (Surface) on decision to Mark rather than Clear targets for stereotype threat females and control females at each day	152

Figure 16.	Observed and optimal decision weights for Type cue (Sub) on decision to Warn rather than Clear targets for stereotype threat females and control females at each day	153
Figure 17.	Observed and optimal decision weights for Type cue (Sub) on decision to Mark rather than Clear targets for stereotype threat females and control females at each day	154
Figure 18.	Observed and optimal decision weights for Class cue (Military) on decision to Mark rather than Clear targets for stereotype threat females and control females at each day	155
Figure 19.	Observed and optimal decision weights for Class cue (Military) on decision to Mark rather than Clear targets for stereotype threat females and control females at each day	156
Figure 20.	Observed and optimal decision weights for Intent cue (Hostile) on decision to Warn rather than Clear targets for stereotype threat females and control females at each day	157
Figure 21.	Observed and optimal decision weights for Intent cue (Hostile) on decision to Mark rather than Clear targets for stereotype threat females and control females at each day	158

#### **INTRODUCTION**

White men can't jump, women can't drive, and three men can't take care of a baby. Aside from their familiarity as popular entertainment punch lines, each of these events also share a subtler and potentially more surprising feature—under the right circumstances, empirical research suggests there is some truth to their claims (Bosson, Haymovitz, & Pinel, 2004; Stone, Lynch, Sjomerling, & Darley, 1999; Yeung & von Hippel, 2008). The specific circumstances in question here refer to instances of *stereotype threat*, a predicament in which an individual is faced with "the risk of confirming, as self-characteristic, a negative stereotype about one's group" based on his/her actions (p. 797, Steele & Aronson, 1995). More specifically, stereotype threat theory posits that the presence of a culturally-shared stereotype which implicates a subgroup is less capable at specific domain tasks or possesses deficient knowledge, skills, or abilities in a domain can lead to a variety of undesirable consequences for individuals identified with the disadvantaged subgroup on domain-relevant tasks (Steele, 1997; Steele & Aronson, 1998; Steele, Spencer & Aronson, 2002).

Although the most widely documented of these undesirable consequences is reduced performance on intellective ability/knowledge tests (e.g., Cole, Matheson, & Anisman, 2007; Good, Aronson, & Harder, 2008; Keller, 2007; Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995; Walton & Cohen, 2003; see Nguyen & Ryan, 2008, for a meta-analysis), stereotype threat has been linked to a variety of other negative outcomes as well. Poorer functioning at physical and social activities (Stone & McWhinnie, 2008; Kray, Galinksy, & Thompson, 2002, respectively), a higher prevalence of internal versus external attributions to failure (Koch, Müller, & Sieverding, 2008), greater engagement in self-handicapping behaviors (erecting barriers to performance that provide a "fallback excuse" for potential failures, Keller,

2002; Steele & Aronson, 1995; Stone, 2002), adoption of performance-avoidance goals (Brodish & Devine, 2009; Smith, 2004; 2006; Smith, Sansone, & White, 2007), discounting the validity, importance, or appropriateness of a task (Keller, 2002; Lesko & Corpus, 2006), and attempts to distance oneself from the stereotyped group (Pronin, Steele, & Ross, 2004; Steele & Aronson, 1995) or "disengage" from the task domain (Crocker, Major, & Steele, 1998; Major, Spencer, Schmader, Wolfe, & Crocker, 1998) have all been linked to stereotype threat. While stereotype threat is most commonly invoked in explanations for the underachievement of minorities (i.e., females and non-Whites) in specific domains, the effect has been shown to generalize to majority subgroups as well. For example, when faced with stereotypes about Asian students' superiority at mathematics and intelligence testing, White males have been shown to perform significantly worse on mathematics tests and to disengage more strongly from the task domain by diminishing the importance/self-relevance of their intellect than in situations where no such stereotype is mentioned (Aronson, Lustina, Good, Keough, Steele, & Brown, 1999; von Hippel, von Hippel, Conway, Preacher, Schooler, & Radvansky, 2005).

The demonstrable effects of stereotype threat thus span a variety of outcomes and are inclusive of virtually all subgroup categories. However, the examination and application of stereotype threat as a phenomenon of interest has primarily been restricted to instances in which achievement or evaluation are the central criteria. Stated differently, the development of stereotype threat theory and investigations of its influence have primarily been of interest to researchers and practitioners *at performance*, defined here as any point in time where an individual is asked to demonstrate some domain knowledge, skill, or ability for the purposes of explicitly diagnosing or measuring that individual's domain competence. Although a subtle distinction (and one that has been inherent in treatments of the concept since its inception, Steele

& Aronson, 1995), this limited scope takes for granted a fundamental tenet and extrapolation of the theory: while the negative stereotypes which lend strength to a situational threat are domain-/ability-specific, the influence of those stereotypes are not necessarily restricted to the manner by which the domain is encountered or the capability expressed. For example, regardless of whether women are asked to complete a difficult test of mathematical ability (an explicitly evaluative context, Spencer et al., 1999), teach young students whose mathematical ability is later assessed (context in which performance of the female instructor is not directly evaluated, Beilock, Gunderson, Ramirez, & Levine, 2010), or learn about novel mathematical operations (context in which performance is not the immediate focus of attention, Rydell, Rydell, & Boucher, 2010), the stereotype "women are less proficient at mathematics" is equally relevant to women participating in these activities.

Of particular interest is the latter of these examples, which implies that stereotype threat could potentially impair the knowledge acquisition process of affected individuals. Though variability in the definition of "intelligence" and related theories about the extent to which individual differences contribute to intellective performance abound (e.g., Sternberg, Conway, Ketron, & Bernstein, 1981; Sternberg & Grigorenko, 2004; Wagner & Sternberg, 1984), there is virtually no disagreement that learning, training, and the accrual of performance-relevant knowledge and experience is critical to successful performance achievement (cf., Baldwin, Ford, & Blume, 2009; Goldstein & Ford, 2002; Kraiger, Ford, & Salas, 1993). To the extent that domain-relevant negative stereotypes impact the acquisition of knowledge, skills, or abilities needed by individuals to effectively perform in a given domain, the subsequent domain achievement of threatened individuals would also be expected to suffer. As will be elaborated further in the sections to follow, this possibility has significant practical and research

implications—not least of which is that the current state of the literature is unable to adequately answer whether examinations of stereotype threat at performance are instances of an insidious situational pressure producing differences between individuals who possess equally sturdy intellective foundations or one that capitalizes on preexisting instabilities.

This simple yet integral concept serves as the primary impetus of the present research effort. The goal of this study is to extend the conceptualization and associated consequences of stereotype threat theory by examining its effects at learning, defined here as any time during which individuals engage in non-evaluative experiences and activities designed to contribute to the development of one's competencies/capabilities through the acquisition and retention of domain-/task-relevant knowledge. The conceptual and methodological rationale for this research begins with an overview of stereotype threat theory's conceptualization, consequences (both proximal and distal), and criticisms. Attention is next directed towards a discussion of learning in the context of stereotype threat theory. Using Kraiger et al.'s (1993) learning outcomes classification scheme as an organizing framework, research from the literature on learning/knowledge acquisition is summarized to characterize the manner by which stereotype threat effects are likely to impact these processes as well as delineate the specific conceptualization of learning pursued in the present study. This section also includes a detailed examination of the first published studies investigating the role of stereotype threat during learning (Rydell, Rydell, & Boucher, 2010; Rydell, Shiffrin, Boucher, Van Loo, & Rydell, 2010) in order to provide some context regarding the contributions which the present research stands to add. Lastly, the formal research hypotheses and their accompanying rationale are advanced which lay out the intended direction of the present study.

## An Overview of Stereotype Threat: Theory, Consequences, and Criticisms

No doubt owing to its provocative conclusions and relatively intuitive logic, stereotype threat theory's account for why some groups of individuals have tended to underperform in specific domains has stimulated interest in both media (e.g., Chandler, 1999; Cloud, 2009; Rivers, 2007) and scholarly outlets. In the 15+ years since Steele and Aronson (1995) published their seminal article on the topic, over 300 empirical studies have been published which attempt to examine the achievement deficiencies elicited by stereotype threat effects. As noted previously, the most common investigations of stereotype threat at performance have been directed towards explaining group performance discrepancies in cognitive ability testing. Examples of stereotype threat's effects have been documented across a variety of testing domains and subgroups, including females and mathematical ability testing (Good et al, 2008; Spencer et al., 1999; Walsh, Hickey, & Duffy, 1999), Black students on general cognitive ability exams (Brown & Day, 2006; McKay, Doverspike, Bowen-Hilton & Martin, 2002, 2003; Steele & Aronson, 1995), and low socioeconomic status individuals on verbal ability tests (Croizet & Claire, 1998; Harrison, Stevens, Monty, & Coakley, 2006), among many others.

Nevertheless, the theoretical basis for the process by which stereotype threat is experienced by individuals and ultimately exerts its influence on their performance/achievement is believed to be the same regardless of its application. In their earliest formulation, Steele and Aronson (1995) proposed that stereotype threat operates by activating a number of proximal affective, behavioral, and cognitive mechanisms detrimental to performance, including "distraction, narrowed attention, anxiety, self-consciousness, withdrawal of effort, [or] overeffort" (p. 809). Although these intervening processes were believed to vary in importance and salience based on the conditions of the performance situation, the authors' primary conclusions

were that stereotype threat leads to both cognitive processing inefficiencies (i.e., threatened individuals spend more time doing fewer things less accurately) and lowered performance expectations/motivation on the part of threatened individuals. Since that time, extensive efforts have been invested in specifically isolating and examining these fundamental operations of stereotype threat and their impact on performance outcomes. Arguably the most complete theoretical treatment and review of the stereotype threat literature to date was presented by Schmader, Johns, and Forbes (2008). Based on their review of the relevant literature, these authors proposed an integrated conceptual representation of the intrapersonal dynamics believed to underlie experiences of stereotype threat as well as a model describing the processes through which stereotype threat influences psychological functioning and performance. Given their richness and ambitious incorporation of the large majority of the stereotype threat literature, these models will be used as the primary conceptualization of stereotype threat in the present study and are described in greater detail below.

**Nature of stereotype threat.** Schmader et al. (2008) state that stereotype threat stems from the activation of three intrapersonal constructs: an individual's concept of his/her group membership, concept of the ability domain in question, and his/her self-concept (Figure 1). However, it is not the mere engagement of these concepts that encapsulates stereotype threat, but rather the *propositional relations* that exist and are altered among them. Semantically, propositional relations describe the evaluations and beliefs that individuals explicitly form in their attempts to validate automatic and associative appraisals of a situation (e.g., a negative reaction to a score one receives on a math test translates into the propositional relation "I am bad at math") (cf., Gawronski & Bodenhausen, 2006). For any given context, a positive propositional relation implies that two concepts coincide with one another (e.g., My group has this ability; I



*Figure 1*. Conceptualization of stereotype threat as a cognitive imbalance triggered by person and/or situation factors (Schmader, Johns, & Forbes, 2008)

am like my group; I have this ability) whereas a negative relation implies that two concepts oppose one another (e.g., My group does not have this ability; I am not like my group; I do not have this ability). On the basis of this relational framework and other similar research (Heider, 1958; Nosek, Banaji, & Greenwald, 2002), Schmader et al. (2008) posit that stereotype threat manifests from a situationally induced imbalance in the implied propositional relations among an individual's concept of group, ability, and self that the individual is driven, yet struggles, to resolve. More specifically, stereotype threat is experienced when a negative propositional relation between one's group membership and the ability domain is engendered which is seemingly irreconcilable with positive propositional relations between the self-and-group and the self-and-domain (e.g., My group does not have this ability, I am like my group, but I have this ability).

As exemplified in Figure 1, the cognitive imbalances that elicit stereotype threat arise from the simultaneous activation of situational primes across the three relational links between group, ability, and self, each of which may be further influenced by certain individual difference characteristics. In the first link between one's group and ability, external environmental cues signal that one's group is considered deficient in the ability domain and thus infer a negative propositional relation exists between those concepts. In studies of stereotype threat, these cues are most commonly introduced through the manipulation of negative stereotypes relevant to the situation and actors. These experimental manipulations—often the hallmark and defining criticism of stereotype threat research (e.g., Cullen, Hardison & Sackett, 2004; Cullen, Waters, & Sackett, 2006)—have been presented in many shapes and forms, including altering individual's perceptions of the diagnosticity of a test (Kray et al., 2001; Steele & Aronson, 1995), the salience/explicitness of a negative performance stereotype (Grand, Ryan, Schmitt, & Hmurovic, 2011; Spencer et al., 1999), or the manner by which a domain task is described (Frantz, Cuddy, Burnett, Ray, & Hart, 2004; Stone et al., 1999). Additionally, individual differences such as stigma consciousness (Brown & Lee, 2005; Brown & Pinel, 2003), group-based rejection sensitivity (Mendoza-Denton, Purdie, Downey, & Davis, 2002), and stereotype knowledge/belief (Keifer & Sekagueptewa, 2007; Schmader, Johns, & Barguissau, 2004) can facilitate the ease with which negative group-ability domain stereotypes are adopted and, by extension, the corresponding negative propositional relation activated.

The second link contributing to the experience of stereotype threat is proposed to exist between a person's self-concept and his/her membership in the stereotyped group. In this case, situational cues promote recognition of a "collective self" as a representative indicator of one's self-concept, thereby encouraging a positive propositional relation between the self-and-group

that deemphasizes an individual's unique strengths, weaknesses, and characteristics (e.g., Marx, Stapel, & Muller, 2005; Shih, Pittinsky, & Ambady, 1999). Experimental primes of the selfgroup link have ranged from pre-performance questionnaires soliciting group identity-relevant information (Ambady, Shih, Kim, & Pittinsky, 2001; McGlone & Aronson, 2006; Shih et al., 1999; Shih, Pittinsky, & Trahan, 2006; Yopyk & Prentice, 2005), to having individuals interact with out-group members prior to performance (Marx & Goff, 2005; Stone & McWhinnie, 2008), to simply asking individuals to provide their gender/race on pre-test demographics (Steele & Aronson, 1995). The results from such experimental studies have generally found performance decrements for the stereotyped group when self-group membership is made salient before rather than after performance (though this point is not without contention, see Stricker & Ward, 2004, Danaher & Crandall, 2008, and Stricker & Ward, 2008).

Additionally, although certain minority groups (females, African Americans, etc.) are most often the focus of stereotype threat researchers, neither the ease of visibility nor the proportional demographic status of one's group is a prerequisite for experiencing stereotype threat. Individuals from less readily detectable group categories—such as those based on socioeconomic status (Croizet & Claire, 1998; Harrison et al., 2006) or mental illness (Quinn, Kahng, & Crocker, 2004)—and even from groups typically considered culturally dominant or in the majority—such as men (Koenig & Eagly, 2005) or Whites (Aronson et al., 1999; Stone, 2002)—are susceptible to stereotype threat under certain circumstances. However, it is certainly the case that members from minority or low-status groups face a far higher prevalence of negative stereotypes, and thus run the risk of more regular exposure to conditions that are favorable to stereotype threat (Gonzalez, Blanton, & Williams, 2002; Shih et al., 1999). Furthermore, this linkage suggests that individuals more likely to exhibit a positive self-group

concept even when situational primes are ambiguous may be more susceptible to threat, a finding supported by research on the effects of group identification in threatening performance situations (Marx et al., 2005; Ployhart, Ziegert, & McFarland, 2003; Schmader, 2002).

The final link in the stereotype threat imbalance is the positive propositional relation between self-and-domain such that an individual's self-concept is associated with doing well in that context due to expectations of success or a high motivation to achieve (Schmader et al., 2008). Personal stake or investment in a domain/outcome has been advanced as a critical precondition for the elicitation of stereotype threat (Steele, 1997; Steele & Aronson, 1995; Steele & Davies, 2003); generally, the more that individuals care about a given domain or doing well in it, the more susceptible they are to threat (Aronson et al. 1999; Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Hess, Auman, Colcombe, & Rahhal, 2003; Keller, 2007; Levy, 1996; Levens, Désert, Croizet, & Darcis, 2000; Spencer et al., 1999; Stone et al., 1999; Wout, Danso, Jackson, & Spencer, 2008; but see Nguyen & Ryan, 2008). This positive self-ability relation has been experimentally elicited primarily by either indicating to participants that an experimental task is challenging but within the scope of their abilities or by simply selecting participants with documented success in the ability or domain (e.g., Aronson et al., 1999; Brown & Pinel, 2003; Josephs, Newman, Brown, & Beer, 2003; Schmader & Johns, 2003; Spencer et al., 1999; Steele & Aronson, 1995). For many, this link represents perhaps the most unfortunate dimension of stereotype threat as it implies that individuals who are the most highly motivated and driven to succeed are those at greatest risk of falling prey to the Sisyphean struggles engendered by the phenomenon (cf., Steele, 1997).

In sum, stereotype threat encapsulates a specific concoction and relational network of situational and intrapersonal characteristics. Although alternative conceptualizations of the

phenomenon exist, the broader appeal and utility of Schmader et al.'s (2008) model is its emphasis on the nature of stereotype threat as an *emergent* situational phenomenon requiring *multiple* conditions be met in order for the phenomenon to occur. Though further research is needed to support the claim, a primary implication of this framework is that stereotype threat cannot (or is highly unlikely to) occur unless all portions of this cognitive imbalance are brought to bear in a situation. Additionally, this model suggests that it may not be possible (or, at minimum, plausible) to vary the "amount" of stereotype threat in a single performance instance in the same manner one might vary other characteristics such as time pressure or cognitive load; instead, the manifestation of threat is influenced by targeting the three core concepts of group, self, and ability and the formation of the conflicting propositional relations among them<sup>1</sup>. Thus, it is perhaps most conceptually accurate to treat stereotype threat (regardless of its application) as the confluence of a defined set of situational and individual characteristics that-when mixed or experienced together in a prescribed fashion—can lead to cognitively and affectively unsettling states/processes that are not conducive to productive psychological functioning. Precisely what those unsettling states/processes are is the topic of the following section.

**Outcomes of stereotype threat.** To this point, the discussion has centered on the manner by which stereotype threat manifests as a cognitive imbalance. However, as with nearly all models of cognitive disruption or inconsistency (e.g., Baumeister & Vohs, 2004; Higgins, 1987; Festinger, 1957) the critical assumption of stereotype threat theory is that the discrepancy it produces prompts a state of unresolved tension within the stereotyped individual that he/she is motivated to dispel. In response, a variety of psychological resources and processes are engaged to aid the resolution effort. In and of itself, such a response is not inherently negative and serves an important homeostatic function for the individual (cf., Pressing, 1999); but, when one

considers that that "brainpower" could otherwise be directed towards the demands imposed by the task domain and, more importantly, is disproportionately experienced by only certain group members, this imbalance and the subsequent chain of events it sets off is substantially more problematic.

Based on their review of the literature, Schmader et al. (2008) developed an integrated model of the process by which stereotype threat is believed to influence performance achievement (Figure 2). The authors devote a great deal of effort to precisely explicating and defending the empirical rationale/support for their framework, much of which is well beyond the scope of the present discussion. Consequently, only the major relations and their relevance to the present study will be highlighted. To begin, Schmader et al. (2008) suggest that three interconnected responses are aroused in reaction to the cognitive imbalance and subsequent ruminations/appraisals generated by stereotype threat: a heightened state of physiological stress/anxiety (e.g., Murphy, Steele, & Gross, 2007; Blascovich, Spencer, Quinn, & Steele, 2001; Croizet, Després, Gauzins, Huguet, Leyens, & Méot, 2004); hyper-vigilant monitoring of both perceived personal performance and feedback/cues that indicate one is threatened or is being influenced by the negative stereotype (e.g., Beilock, Rydell, & McConnell, 2007; Ben-Zeev, Fein, & Inzlicht, 2005; Forbes, Seibt & Förster, 2004; Schmader, & Allen, 2008; Johns, Inzlicht, & Schmader, 2007); and thought suppression processes directed towards regulating negative cognitions and affect (e.g., Johns et al., 2007; von Hippel et al., 2005; Wraga, Helt, Jacobs, & Sullivan, 2007). In turn, each of these factors is believed to consume resources from an individual's working memory (cf., Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Matheson & Cole, 2004; Muraven & Baumeister, 2000; Smith & Henry, 1996; Wenzlaff & Wegner, 2000). Simply described, working memory can be conceptualized as the limited capacity "cognitive



*Figure 2*. Integrated process model of stereotype threat effects on performance (Schmader, Johns, & Forbes, 2008)

workspace" that individuals employ to coordinate the storage and controlled attention of immediately relevant thoughts, operations, and information (Baddeley, 1986, 1997; Baddely & Hitch, 1974; Engle, 2002; Kane, Conway, Hambrick, & Engle, 2007). Consequently, the siphoning and disruption of working memory resources by the experience of stereotype threat reduces the efficiency and capability with which one can effectively manage and complete cognitively demanding tasks (Beilock et al., 2006; Beilock et al., 2007; Schmader, 2010; Schmader, Forbes, Zhang, & Berry Mendes, 2009; Schmader & Johns, 2003), thereby inhibiting an individual from reaching their performance potential.

As is apparent from Figure 2, working memory is posited to be the most proximal mechanism through which experiences of stereotype threat impede functioning on cognitivebased tasks and thus warrants closer examination. In their seminal manuscript on the topic of working memory, Baddeley and Hitch (1974) summarized preliminary evidence for the existence of a cognitive processing system that operated similarly to, yet distinctive from, previous conceptualizations of short-term memory. Early treatments of these authors' working memory hypothesis proposed a tripartite system composed of a superordinate *central executive* responsible for controlled processing and attention among two "slave systems," the *phonological loop* and the *visuospatial sketchpad*, which coordinated the temporary storage (~1-2 seconds) and manipulation/rehearsal of auditory/speech-based and visuospatial information, respectively (Baddeley, 1986, 1992). Baddeley (2000) later amended this framework by incorporating a fourth system termed the *episodic buffer* that assumes some of the controlled processing functions from the executive control. Specifically, the episodic buffer is characterized as an interface for integrating information from the phonological loop and visuospatial sketchpad with information from long-term memory to form brief "episodic memories" over short periods of time, leaving the executive control primarily responsible for directing attentional efforts (Baddeley, 2001).

Although Baddeley's model (1986, 2000) is widely cited as the preeminent conceptualization of working memory, research spearheaded by Engle, Kane and colleagues (e.g., Engle, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane & Engle, 2003; Kane et al., 2007; Kane, Hambrick, Tuholski, Wilhem, Payne, & Engle, 2004) has adopted a slightly altered perspective on the functioning of this cognitive system that integrates well with stereotype threat theory. The primary point of emphasis these researchers advance attempts to more discretely distinguish "storage" functions of working memory from its role as a domain-general executive attention process. As Engle (2002) relates:

The term capacity, as used in discussions of short-term memory (STM), often conjures up images of a limited number of items or chunks that can be stored (e.g.,  $7 \pm 2$ ). However, my sense is that [working memory] WM capacity is not about individual differences in how many items can be stored per se but about differences in the ability to control

attention to maintain information in an active, quickly retrievable state. Thus, WM capacity is just as important in retention of a single representation, such as the representation of a goal or of the status of a changing variable, as it is in determining how many representations can be maintained. WM capacity is not directly about memory—it is about using attention to maintain or suppress information. WM capacity is about memory only indirectly. Greater WM capacity does mean that more items can be maintained as active, but this is a result of greater ability to control attention, not a larger memory store. Thus, greater WM capacity also means greater ability to use attention to avoid distraction. (p. 20)

Additionally, this interpretation further implies that short-term memory is essentially a subset of working memory, with short-term memory performing functions akin to the phonological loop and visuospatial sketchpad and working memory coordinating controlled attention (Engle et al., 1999). Although still largely consistent with Baddeley's framework (1986, 2000), empirical investigations based on this perspective seem to support the notion that short-term memory processes are domain-specific (that is, the chunking, rehearsing, coding, storing, etc. of information is specific to a particular domain) whereas the executive control of working memory is agnostic and functions more consistently across multiple domains in maintaining particular memory representations in active and easily accessible states (Engle & Kane, 2004; Kane & Engle, 2003; Kane et al., 2004).

Given the above depiction, it is possible to more precisely explicate the manner by which stereotype threat exerts its influence on meaningful outcomes of interest by considering its direct impact on working memory<sup>2</sup>. Working memory capacity has been implicated in a large variety of intellective activities, including reading comprehension (Daneman & Merikle, 1996), problem-solving based on listening comprehension (Adams & Hitch, 1997; Carpenter, Just, & Shell, 1990), advanced reasoning (Kyllonen & Christal, 1990), strategy adaptation (Schunn & Reder, 2001), multitasking (König, Bühner, & Mürling, 2005), Stroop color naming (Kane & Engle, 2003), and visuospatial reasoning (Kane et al., 2004), among others (see Feldman Barrett,

Tugade, & Engle, 2004, for further review). Additionally, many researchers consider working memory the primary processing component underlying general fluid intelligence (Cattell, 1943), or the ability to reason logically, solve novel problems, and adapt to new circumstances (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle et al., 1999; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Kyllonen, 1996; Kyllonen & Christal, 1990).

Common to all such heavily working memory-centric and fluid intelligence tasks is that they necessitate focused direction, effortful processing, and dynamic self-regulation of one's controlled attention in order to successfully complete—characteristics which leave them particularly susceptible to the added demands of stereotype threat. As described previously, the cognitive imbalance imparted through the presence of a negative domain-relevant stereotype elicits increased stress from threatened individuals and spurs them to adopt (consciously or not) added monitoring processes to gauge the extent to which their performance or actions are confirming the stereotype (e.g., Grand et al., 2011; Ployhart et al., 2003; Rydell, Rydell, & Boucher, 2010). These responses are believed to engender further internal appraisal processes through which the individual attempts to reconcile the contradictory state and that tend to elicit increased focus/attention towards negative thoughts and emotions (e.g., Forbes et al., 2007), leading to subsequent cognitive effort to subdue those responses. All the while, these added, task-irrelevant attentional demands are indiscriminately filtered through the working memory system.

Working memory capacity facilitates functioning on intellective activities by enabling individuals to *both* maintain one's attention on task-relevant facets of a situation and suppress interference from non-relevant components (Rosen & Engle, 1998). However, working memory has limits on its capacity (e.g., Engle, 2002; Kane et al., 2004), and such controlling and

suppressing functions are cognitively demanding. Thus to the extent that stereotype threat engenders more task-irrelevant foci to suppress, less of one's limited working memory capacity can be allocated to rehearsing, integrating, and manipulating information relevant to achieving goal-directed objectives. In short, the primary harm introduced by stereotype threat is the contribution of irrelevant affective and cognitive stimuli that unnecessarily hijack an individual's working memory resources, thereby leaving fewer cognitive resources to devote to task demands (Schmader et al., 2003; Schmader et al., 2008).

This proposition is also generally supported by the observed pattern of results concerning stereotype threat effects. For example, there is evidence to support the claim that domain activities must be both difficult and relatively complex for stereotype threat effects to manifest (e.g., Nguyen & Ryan, 2008; Quinn & Spencer, 2001; Steele & Aronson, 1995). Undertaking tasks that are simple and/or well-rehearsed seldom leads to underachievement by threatened individuals (e.g., Beilock et al., 2007; O'Brien & Crandall, 2003). From an attentional resource allocation perspective, participating in difficult and complex tasks presumably pushes the bounds of one's working memory to its limits. When conditions conducive to stereotype threat are then made salient in such situations, the added cognitive demands simply overwhelm the capacity of stereotype threatened individuals and prevent them from focusing on required task demands. Note that this rationale does not imply that the underachievement of stereotype threatened individuals necessarily results from reduced effort or motivation to succeed on their part. In fact, and as would be predicted based on the working memory models detailed above, targeted individuals have been shown to exert equal amounts of (or, in many cases, more) effort and persist in those efforts longer than non-threatened individuals (Forbes et al., 2007; Jamieson & Harkins, 2007; Kray, Thompson, & Galinsky, 2001; O'Brien & Crandall, 2003; Rydell, Shiffrin,

et al., 2010). However, this effort is inefficiently oriented towards managing task-irrelevant demands that are not present for non-threatened individuals, leading to quicker mental fatigue and exhausting more time and mental energy on things that do not contribute to task accomplishment—essentially forcing these individuals to do more while achieving less (cf., Grier et al., 2003; Steele & Aronson, 1995).

**Criticisms of stereotype threat.** It is pertinent to briefly address two broad concerns commonly levied against stereotype threat theory that also bear on the proposed research. First, some researchers have advocated that stereotype threat is more appropriately considered through the lens of motivational states and goal-orientation primes rather than the working memory cognitive models described herein (e.g., Marx & Stapel, 2006a, 2006b; Wheeler & Petty, 2001). For example, Grimm, Markman, Maddox, and Baldwin (2009) suggest that stereotype threat can be interpreted as a mismatch in regulatory focus driven by differences in the reward structure of the task environment and the prevention/failure-avoidance states that are primed by the introduction of negative stereotypes (i.e., tasks in which "doing something better" implies greater success do not reward avoiding failure, Canidu et al., 2005; Seibt & Förster, 2004). In their study, Grimm et al. (2009) present convincing evidence that one can eliminate the performance discrepancies caused by stereotype threat for women on mathematics tests by simply changing the performance criteria of the task and instructing participants that their goal on a performance task is to avoid losing a certain number of points (rather than gaining a certain number of points)—a reward structure presumably better aligned with the prevention/avoidance state held by negatively stereotyped women which should thus facilitate higher achievement motivation.

However, the conceptual rationale underlying such effects does not necessarily contradict, supersede, or invalidate the propositions of the cognitive imbalance or working memory models.

Instead, they largely complement one another. Grimm et al.'s (2009) findings are consistent with the notion that in the face of intrapersonal conflict about their and their' group's abilities, stereotype threatened individuals engage in more effortful situational monitoring that taxes working memory resources (e.g., Beilock & Carr, 2005; Beilock et al., 2006; Beilock et al., 2007; Schmader et al., 2008) and that alterations to the structure of the environment can make this effortful process more or less beneficial to performance. For example, Forbes et al. (2007) and Jamieson and Harkins (2007) report that threatened individuals are typically more motivated to correct errors in performance contexts, a strategy that facilitates equal or better achievement compared to non-threatened individuals in situations where it is functional. However, when the performance context is less conducive to these strategies as a result of greater working memory demands (i.e., more difficult task, stricter time limit, less space for error correction), threatened individuals operating under such heightened motivational states tend to underperform (cf., Harkins, 2006; Nguyen & Ryan, 2008). Such findings are supportive of the notion that threatened versus non-threatened individuals demonstrate differences in the processes/manner by which their performance outcomes are achieved (e.g., Beilock & Carr, 2005; Rydell, Shiffrin, et al., 2010). In short, the motivational approaches to stereotype threat offer valuable insights into the mechanisms of the phenomenon and are likely to supplement its theoretical underpinnings, but they do not invalidate cognitive/working memory accounts.

The final battery of criticisms typically voiced against stereotype threat theory bear less on the theoretical end of the quotient and more heavily on its application. For more than a decade, organizational researchers have noted a variety of difficulties with detecting the effects of stereotype threat in "real-world" assessment and performance situations, implying that its feasibility as a phenomenon of import may thus be limited (Cullen et al., 2004; Cullen et al.,

2006; Good, Aronson, & Inzlicht, 2003; Sackett, Hardison, & Cullen, 2004; Sackett & Ryan, 2012; Sackett, Schmitt, Ellingson & Kabin, 2001; Schmidt, 2002; Stricker & Ward, 2004). In large part, these concerns can be summarized as follows:

- The manner by which stereotype threat is introduced into a situation is too unrealistic and/or unethical and therefore the likelihood it would be elicited in an evaluative performance situation is virtually nonexistent.
- 2. Analyses of large datasets in which between-group performance differences attributable to stereotype threat might be expected have not supported the theory's predictions.
- 3. For those who acknowledge that stereotype threat may be a legitimate concern in high stakes performance situations, the stereotype threat removal strategies offered by researchers as solutions to the issue (e.g., minimizing the diagnosticity/evaluative purposes of a test, priming alternative identities, etc.) are too impractical to be implemented.

While deconstructing and addressing each of these concerns is well beyond the means of this paper (cf., Steele & Davies, 2003; Stricker & Ward, 2008), they do serve to highlight certain key points relevant to the present study. First, as noted previously, stereotype threat is the characterization of an emergent cognitive imbalance predicated on the activation of specific relations among a small set of situational and intrapersonal characteristics. There can be little argument that these restrictions dictate strict boundary conditions for when one might expect to observe stereotype threat effects and when it would be appropriate to conclude that stereotype threat is operating. However, both the priming of negative stereotype cues (e.g., Nguyen, O'Neal, & Ryan, 2003, etc.) and attempts to minimize/reverse stereotype threat effects (e.g., Grimm et al.,

2009; Stricker & Ward, 2004) may occur through many subtly different paths, not all of which have been fully tested or documented. Furthermore, even if the base rate of stereotype threat occurrences are minimal in evaluative work or academic performance contexts, there are still other applications where stereotype threat effects may more readily manifest that could potentially influence important outcomes that have not yet been adequately examined—such as training or learning contexts.

Second, all performance is not created equal. Although it is common to presume that similar between-group performance scores indicate that group members operate in functionally equivalent manners, the *process* by which performance outcomes are generated is a crucial consideration. For example, stereotype threat researchers have not typically drawn distinctions between maximal (i.e., assessments aimed at evaluating individual's highest predicted achievement) and typical (i.e., assessments aimed at evaluating individual's day-to-day/sustainable achievement) performance contexts when evaluating threat effects. However, to the extent that *how* one performs a task is as important as *what* one is capable of performing, stereotype threat may influence behavior/achievement in manners not easily observable in contexts traditionally examined in stereotype threat studies (e.g., testing/assessment, selection).

In sum, the characterization of stereotype threat as a cognitive, behavioral, and affective experience stemming from an emergent and dynamic process sensitive to the situational characteristics in which an individual operates implies that its effects may extend beyond the organizational functions traditionally examined. The patterns of interference that characterize the expression of stereotype threat are generally undesirable features which could seemingly arise and influence outcomes in areas other than testing and performance assessment. It is this very possibility that serves as the impetus for the present research, and to which focus is now directed.

## Stereotype Threat at Learning: Rationale, Applications, and Implications

As implied by Figure 1 and Schmader et al. (2008), the experience of stereotype threat can theoretically emerge in any instance where a negative domain-relevant stereotype exists capable of producing dissonance amongst a person's concept of self, group, and ability that he/she is motivated to overcome. Although perhaps self-evident, the stereotypes which could trigger this imbalance (e.g., "Women struggle with mathematics," "Whites are not as naturally athletic as Blacks," "Lower income individuals are not very intelligent," etc.) do not simply appear during performance episodes and then vanish; such stereotypes are persistent features of a domain and can influence virtually any related functional pursuit within its purview (e.g., Stangor, 2000; Stangor & Lange, 1994). By this rationale then, stereotype threat may manifest in numerous occasions and across myriad circumstances. Among the many potential applications of the theory though, the implication of working memory efficiency as the primary gateway through which stereotype threat exerts influence suggests that *learning* efforts may be particularly sensitive to threat effects.

**Conceptual background for stereotype threat at learning.** Learning is often generically defined as a relatively permanent change in knowledge, skill, or behavior brought about through experience (e.g., Weiss, 1990; Wexley & Latham, 1991). This conceptualization, however, somewhat oversimplifies the nuanced and multifaceted nature of what learning "means;" for example, at different points in the history of psychology, learning has been approached from perspectives of classical and operant conditioning, observational and social learning/modeling, rote memorization of letter/digit strings, insight learning, and latent learning of complex procedures (Kosslyn & Rossberg, 2004). In an effort to better orient the diversity of such learning investigations, a number of researchers suggest that it is desirable to focus on the

outcomes of the learning process that are of interest to the research question/domain as a means of better aligning empirical efforts and to capture the full spectrum of learning-relevant behaviors (cf., Gagne, 1984; Glaser, 1990; Messick, 1984). In the spirit of this recommendation, Kraiger et al. (1993) offer a relatively simple yet comprehensive categorization system useful for describing the different forms and specifications of learning outcomes one could assess during learning or training experiences (Figure 3). These authors' classification scheme describes three broad classes of learning outcomes: cognitive (procurement and synthesis of information/knowledge), skill-based (development of procedural/behavioral routines or understanding), and affective (attitudinal, motivational, or dispositional changes). Each of these three learning outcomes is also further divided into associated categories that characterize the processes, constructs, or targets associated with that outcome.

Kraiger et al.'s (1993) depiction of learning outcomes serves as a useful guiding framework for examining the potential impact of stereotype threat at learning. As noted previously, there is strong evidence to suggest that stereotype threat interferes with an individual's cognitive functioning primarily by forcing him/her to allocate limited attentional resources from working memory towards suppressing task-irrelevant responses that stem from the presence of a negative domain stereotype (e.g., Beilock et al., 2006; Beilock et al., 2007; Schmader, 2010; Schmader et al., 2009; Schmader & Johns, 2003). To the extent that working memory capacity/processes are related to relevant cognitive, skill-based, and/or affective learning outcomes, there would be sufficient reason to believe that the generalizable experience of stereotype threat could negatively influence learning efforts—and indeed, there is evidence supporting the relation between working memory and components of all three of these learning outcomes. For example, individuals with lower working memory capacity appear to be less



*Figure 3*. Classification system for learning outcomes (adapted from Kraiger, Ford, & Salas, 1993)

proficient at maintaining and adhering to appropriate task goals (i.e., affective outcomes) in performance situations with greater task interference than individuals with higher working memory (Kane, Bleckley, Conway, & Engle, 2001; Kane & Engle, 2003; Rosen & Engle, 1997; see also Daily, Lovett, & Reder, 2001). Similarly, there is a large body of research demonstrating that working memory contributes heavily to both synthesizing declarative facts, concepts, and information into longer-term memory stores (i.e., cognitive outcomes) as well as integrating those pieces into procedural relations (i.e., skill-based outcomes) (e.g., Baddeley, 2001; Budd, Whitney, & Turley, 1995; Cantor & Engle, 1993; Just & Carpenter, 1992; Rosen & Engle, 1997; see also Feldman Barrett et al., 2004, for a review).

In the domain of reading comprehension, for instance, Whitney, Ritchie, and Clark (1991) found that persons with less available working memory tend to draw more surface-level (versus deeper) interpretations of written text and do so much earlier while reading than individuals with higher working memory. Presumably this occurs because the former's reduced working memory capacity does not enable them to maintain enough information in an active, readily accessible state for a long enough time to draw more comprehensive inferences (e.g., Engle, 2002). This form of reading comprehension, however, has long been demonstrated to be an ineffective and inefficient means of interpreting written information. Research indicates that readers who employ more thematic or structural approaches to reading (i.e., attempt to deduce major themes and inferences from text) are better able to recall and make use of more information from narrative passages than readers who process that same text serially (i.e., sentence-by-sentence) (e.g., Loman & Mayer, 1983; Marshall & Glock, 1979; Meyer, Brandt, & Bluth, 1980; Meyer & Rice, 1982; Reder & Anderson, 1980). Thus, a primary implication of these research streams is that individuals with less available working memory-whether as a result of individual differences in capacity or "artificial" reductions through situational primes (e.g., Beilock & Carr, 2005)—may be at a much larger disadvantage when it comes to achieving cognitive, skill-based, and/or affective learning outcomes in complex, self-directed learning environments (e.g., students taking online classes, workers learning complicated procedures from an instructional manual, decision-makers attempting to synthesize technical reports, etc.).

Although there is reasonable evidence to postulate that stereotype threat is detrimental to all three of the learning outcomes shown in Figure 3, the present study focuses on the attainment
of cognitive learning outcomes as the primary domain of interest for three reasons. First, the centrality of working memory processes to the acquisition of declarative and procedural knowledge is a hallmark of nearly all models of higher-order cognition (e.g., Anderson et al., 2004; Baddeley, 2000; Just & Carpenter, 1992; Kieras & Meyer, 1997; Newell, 1990); thus, deficiencies in related cognitive learning outcomes represent the most likely victim of any decrement in working memory generated by stereotype threat. Second, despite its broader range of applicability, stereotype threat theory has gained its greatest traction in the area of cognitive ability and intelligence testing (cf., Nguyen & Ryan, 2008). While there is some debate regarding the magnitude and empirical relations among measures of working memory and cognitive ability (see Ackerman, Beier, & Boyle, 2005, and responses from Kane, Hambrick, & Conway, 2005, and Oberauer, Schulze, Wilhelm, & Süß, 2005) there is general agreement that the concepts are highly interrelated (Beier, & Ackerman, 2005; Carroll, 1993; Engle, 2002; Jensen, 1998; Kyllonen & Christal, 1990; Turner & Engle, 1989). The generalization of stereotype threat from cognitive performance outcomes to cognitive learning outcomes, therefore, represents a logical extension of the theory—and one which also carries implications for disentangling how stereotype threat's effects at learning may influence stereotype threat at performance. Lastly, fluid intelligence/working memory is often considered among the most important factors in the success of professional/educational learning experiences, especially in situations that are complex or demanding (Deary, Strand, Smith, & Fernandes, 2007; Gottfredson, 1997; Jaeggi et al., 2008; Neisser et al., 1996; Rohde & Thomspon, 2007; te Nijenhuis, van Vianen, & van der Flier, 2007). As such, person or situation factors which reduce working memory and influence fluid intelligence are likely to have a substantial impact on the acquisition and expression of domain-relevant information in the learning environment.

Within the subset of cognitive learning outcomes, Kraiger et al. (1993) describe three categories of learning constructs which could serve as potential targets for examining effects of stereotype threat at learning. The first, *declarative knowledge*, reflects the encoding of data "chunks" comprised of context-free facts, statements, or assertions—or, as it is colloquially referenced, information about "what" (i.e., "3 + 4 = 7," "Lincoln's Gettysburg Address was delivered in 1863," etc., Anderson, 1996; Miller, 1956; Simon, 1974). Learning outcomes of this type are widely regarded as foundational to higher-order cognitive skill development, and it is a generally accepted dictum that the acquisition of basic declarative knowledge is a necessary condition for the development of more sophisticated procedural knowledge (e.g., Ackerman, 1986, 1987; Anderson, 1982, 1993a, 1996).

The second group of cognitive learning outcomes refers to *knowledge organization*, or the manner by which individuals represent relations/associations among concepts, facts, functions, and other knowledge objects relevant to a given task domain (e.g., Glaser, 1990; Jonassen, Beissner, & Yacci, 1993; Rowe, Cook, Hall, & Halgren, 1996; Schoenfeld & Herrmann, 1982). Synonymous with mental models, cognitive maps, schema, or conceptual frameworks (Dorsey, Campbell, Foster, & Miles, 1999; Schuelke et al., 2009), these knowledge structures serve as contextual organizers for the interpretation and acquisition of new knowledge as well as influence one's ability to make use of existing knowledge to accomplish task requirements (Ausubel, 1963; Day, Winfred, & Gettman, 2001; Kozlowski, Gully, et al., 2001; Messick, 1984; Medin et al., 2006). For this reason, some researchers suggest that the organization of knowledge structures may be of equal or greater importance than the amount or type of knowledge one possesses (Johnson-Laird, 1983; Kraiger et al., 1993; Rouse & Morris, 1986).

The final cognitive learning outcome concerns the development of *cognitive strategies*, indicative of the internalized procedures and mental activities that individuals use to facilitate the synthesis of knowledge and its application to a given task space (Kraiger et al., 1993; Prawat, 1989). The achievement of effective cognitive strategies signals that individuals have developed a deeper understanding of the relationship between their capabilities and the demands of the task environment, as well as metacognitive awareness of their thought processes and causal attributions relevant to task completion (Bereiter & Scardamalia, 1985; Kanfer & Ackerman, 1989; Pressley, Snyder, Levin, Murray, & Ghatala, 1987). In their framework, Kraiger et al. (1993) posit a sequential progression through these three cognitive learning outcomes as learners advance from early to later stages of knowledge acquisition. Individuals generally work to acquire basic declarative knowledge first, organize that knowledge into meaningful structures/mental models that provide context for drawing interpretations among relevant information, and then use that understanding to develop procedural approaches for accomplishing specific task goals.

A final important consideration for examining stereotype threat effects at learning is explicating the manner by which the learning environment is organized as such structural aspects can hold significant influence over the attainment of desired learning outcomes (Bell & Kozlowski, 2008; Iran-Nejad, 1990; Schwartz & Bransford, 1998). The present study focuses on the effects of stereotype threat during episodes of *exploratory learning*. Exploratory learning (sometimes referred to as discovery learning) is a form of active learning in which individuals are encouraged to experiment with task content in order to infer the principles, rules, and mechanisms of a given operational domain (Frese et al., 1988; Kamouri, Kamouri, & Smith, 1986; McDaniel & Schlader, 1990). As opposed to more traditional passive learning approaches

(e.g., lectures, proceduralized instruction, etc.), exploratory learning provides near complete control over the instructional environment to the learner, who bears the brunt of the responsibility for making learning decisions (e.g., choosing what content to learn and when to learn it, monitoring learning progress and adjusting strategies as necessary, etc.). This requirement promotes an inductive learning frame that necessitates engaging in more effortful metacognitive activity on the part of the learner, a critical component in the acquisition and transfer of adaptive expertise and complex skills/knowledge (Bell & Kozlowski, 2002, 2008; Ford & Kraiger, 1995; Ford, Smith, Weissbein, Gully, & Salas, 1998; Frese et al., 1988; Ivancic & Hesketh, 2000). Researchers have noted that learners can easily be overwhelmed by purely exploratory approaches and may fail to ever come into contact with the instructional material (Debowski, Wood, & Bandura, 2001; Mayer, 2004); however, the provision of even minimal guidance can often be enough to stimulate the sense-making and metacognitive efforts essential to knowledge acquisition without undermining the self-directed efforts of learners (Bell & Kozlowski, 2008; Kozlowski, Toney, et al., 2001).

As a result, such guided "constructivist" approaches to learning and training in which learners are intimately involved in the comprehension and organization of domain/tasks concepts, principles, strategies, etc. have become increasingly popular instructional methods in both professional (e.g., de Freitas & Neumann, 20009; Grand & Kozlowski, in press; Rieman, 1996) and educational (Marshall, 1996; Phillips, 1998; Steffe & Gale, 1995) domains. Given the demands placed on learners in such approaches though, exploratory learning environments may be particularly susceptible to stereotype threat effects. Many aspects of working memory have been implicated in the metacognitive functions which are central to successfully navigating exploratory learning paradigms (i.e., selective attention, error detection, inhibitory control,

Fernandez-Duque, Baird, & Posner, 2000; Shimamura, 2000). To the extent that stereotype threat disrupts available working memory capacity by stimulating task-irrelevant thoughts and emotions (Schmader et al., 2008), exploratory learning paradigms may represent one of the more probable instances in which threat-based learning decrements are likely to be experienced. Based on its relative popularity, desirable learning outcomes, and cognitively demanding nature, this particular learning environment thus marks a reasonable point of departure for examining stereotype threat at learning with both empirical and practical implications.

**Examinations of stereotype threat at learning.** Rydell and colleagues (Rydell, Rydell, & Boucher, 2010; Rydell, Shiffrin, et al., 2010) have recently published a series of findings which mark the first documented effects of stereotype threat at learning. Table 1 presents a summary of these experiments outlining the basic rationale, methodological aspects, and results for each study<sup>3</sup>. Additionally an attempt was made to map the learning outcomes assessed in each of the studies back to the classification scheme of Kraiger et al. (1993) to illustrate the scope of their research findings. Taken together, these investigations provide an ambitious point of entry into the research domain. The accumulated evidence they present illustrates an important initial application of stereotype threat theory to outcomes specific to the learning process and furthermore offers a useful means for assessing a number of methodological and conceptual considerations for continued work in the area. Additionally, they reveal a number of theoretical and methodological concerns within this research stream in need of further refinement. These issues bear direct relevance to the development of the current study and help elucidate its proposed rationale and added contribution; thus, these points of interest are highlighted below.

Although their efforts mark only an initial foray into the research domain, the manner by which learning was conceptualized, implemented, and operationalized in both of the Rydell

## Table 1

		Rydell, Rydell, & Boucher (2010)	
	Study 1	Study 2	Study 3
Description	Examined whether ST influences women's ability to learn and perform novel mathematical rules.	Examined whether ST influences women's ability to learn rules from a novel math task (MA). Previous research has shown that ST has no influence on women when solving easy MA items. However, ST may influence the learning of MA rules, which should inhibit their ability to solve even easy MA problems.	Examined whether ST influences women's ability to learn abstract logic task that utilized math principles. Also examined whether ST at initial learning inhibits transfer of learning to novel domain and implicit learning.
Procedure/ task	Participants received tutorial on how to solve math problems based on 8 novel mathematical rules presented one at a time. After first 4 rules were presented, additional instructions were presented that introduced ST for half of participants. The remaining 4 rules were then presented in the same manner as before.	2 (ST: control, ST) x 2 (introduction of ST: before learning, after learning) factorial design. Self-paced procedural tutorials were presented to all participants, with presence & placement of instructional manipulation depending on experimental condition.	Participants received ST or control instructions prior to initial learning. Focal & transfer learning tasks presented list of logic rules to learn (e.g., circle plus diamond equals flag, etc.); transfer task used new symbols, but followed same logic rules. An additional implicit learning task asked participants to indicate if a target symbol had been presented in the focal task after being primed with a new or old stimulus; individuals who learned focal task stimuli should take longer to respond to old-old than new-old pairings.

Summary of Rydell and Colleagues' Multi-study Experiments Examining Stereotype Threat Effects at Learning

Table 1 (cont'd)

	Rydell, Rydell, & Boucher (2010)			
	Study 1	Study 2	Study 3	
Learning measure	Number of rules recalled correctly	Subjective ratings of participants' written description of the steps for solving MA problems	Number of questions correctly answered on test applying logic rules ( <i>focal</i> and <i>transfer task</i> ); response time to identify target symbol ( <i>implicit learning task</i> )	
Performance measure	Number of math problems requiring the novel learned rules answered correctly	Number of easy, moderate, and difficult MA items answered correctly	N/A	
Main results	No difference in pre-instruction rule learning for ST vs. control women, but ST women recalled fewer post- instruction rules. Performance worse on problems with post-, but not pre-, instruction rules for ST women.	ST-before learning women provided less accurate descriptions and answered fewer easy problems than control; no difference on description ratings or performance for ST-after learning vs. control women.	ST women scored lower on focal and transfer task than control, & the effect of ST on transfer test was greater than on focal test. Response times longer for control vs. ST women on implicit learning task	
Learning outcome	Cognitive (Declarative knowledge)	Cognitive (Declarative knowledge)	Cognitive (Declarative knowledge)	

*Note*. ST = stereotype threat; MA = modular arithmetic (see Beilock et al., 2007).

# Table 1 (cont'd)

		Rydell, Shiffrin, et al. (2010)	
	Study 4	Study 5	Study 6
Description	Examined whether ST influences women's ability to learn more efficient, automatized, and less effortful processing strategies for completing a visual search task. <sup>†</sup>	Examined effects of ST introduced later in learning (control+ST) and effects of ST removed later (ST+release) on women's ability to learn visual search task.	Examined alternative indicator of ST's ability to influence women's ability to learn automatic visual search strategy by examining whether performance on an unrelated visual search task would be interfered with by presence of a familiar target stimuli.
Procedure/ task	Participants randomly assigned to receive ST vs. control instructions at beginning and at start of each trial block (6 blocks of 80 trials). Visual search task involved identifying whether one of five target Chinese characters was present or absent amongst either two or four displayed Chinese characters.	Participants randomly assigned to control+ST, ST+release, and control conditions and completed 8 blocks of 80 comparison trials. After block 6, ST was introduced to the control+ST group and a self-affirmation manipulation meant to reduce ST was introduced to the ST+release group.	Same learning task and design as Study 4. Following learning trials, participants presented with new visual search task where goal was to identify which of two same-color patches was more saturated. Superimposed on each patch was either a new Chinese character or one of the target characters from learning.

# Table 1 (cont'd)

	Rydell, Shiffrin, et al. (2010)			
	Study 4	Study 5	Study 6	
Learning measure	Response times (T) collected and deconstructed into separate measures of learning and performance. Measures were derived from T based on algorithmic model of serial self-terminating search in visual information processing (i.e., target items compared to display items one at a time 		Response time required to identify saturated color swatch; longer time when a familiar vs. new Chinese character presented implies individuals had learned target characters, which	
	<u>Learning</u> = comparison time per charac implies individuals are learning/automa strategies for visual search.	eter ( $C$ ). Reduction in $C$ over time ating more efficient processing	interfered with color saturation task	
Performance measure	<u>Performance</u> = base time of responses of visual search comparisons ( <i>B</i> ). Reducti performance for components of respon- perception time, motor-response time, of	other than those used to carry out on in <i>B</i> over time implies improved ding unrelated to learning (i.e., etc.).	N/A	
Main results	<i>C</i> decreased (indicating learning) for control women while remaining relatively constant (no learning) for ST women across blocks. <i>B</i> decreased for ST women (indicating performance improvement on processing components not related to visual search).	<i>C</i> decreased for women in control+ST until block 6, after which it increased; <i>C</i> remained constant for ST+release group for all blocks. <i>B</i> increased for women in control+ST group until block 6, after which it decreased; <i>B</i> decreased for women in ST+release group for all blocks.	Control women took longer to select correct color patch than ST women when target Chinese character was presented; suggests automatic processing of target characters was learned by control women and was interfering with new visual search task but not ST women	
Learning outcome	Skill-based (Compilation & Automaticity)	Skill-based (Compilation & Automaticity)	Skill-based (Compilation & Automaticity)	

<sup>1</sup>ST was elicited by instructing participants that the visual search task was diagnostic of why women tend to underperform on math tests

publications is somewhat questionable. With respect to the former two concerns, relatively little theoretical or methodological attention was directed towards the design, delivery, and/or development of the participant's learning environment—a key determinant in the effectiveness of individuals' learning experiences (e.g., Kozlowski, Toney, et al., 2001). For example, the instructional delivery mechanism through which participants were expected to learn the novel mathematical principles/rules in Studies 1 and 2 were presented as short, text-based tutorials in which participants were passively exposed to the material. While there is nothing inherently wrong with empirically examining stereotype threat effects in such a learning system, no mention was made of the implications this environment held for the expected success of learning (regardless of threat) in this context nor whether the sparse training delivery system may have been more or less susceptible to threat effects than other more feature-rich—but potentially more cognitively demanding—systems (i.e., provision of feedback, active participation, advanced organizers, etc., van Merriënboer & Sweller, 2005).

Furthermore, the experience of learning is typically viewed as a dynamic, iterative process that develops over time, exposure, and rehearsal (e.g., Anderson et al., 2004; Goldstein & Ford, 2002). While Studies 4-6 integrated a longitudinal component in their design, the presentation of novel material in a single brief episode in Studies 1-3 may not have been a rich enough context for individuals to learn the material. For example, the average number of declarative math rules correctly recalled in Study 1 was only .83 and .74 out of 8 for women in the control and stereotype threat conditions, respectively, indicating that even those participants purportedly not influenced by stereotype threat had only learned approximately 10% of the presented material. At best then, these results suggest that there may be very minor (though statistically significant) differences in *immediate* "learning" as a result of stereotype threat, but

this single exposure learning environment limits the extent to which inferences can be drawn about threat-based interferences on the overall process of cognitive knowledge acquisition. Admittedly, the influence of threat relative to the design and characteristics of the learning environment was not a primary or even secondary focus of the Rydell studies. Thus, acknowledgement and consideration of the manner by which a learner's engagement with the desired content material could impact the attainment of desired learning outcomes is a needed next step within this research stream.

Rydell and colleagues are quick to note that the operationalization of learning and the manner by which one assesses that criterion with respect to stereotype threat is a crucial matter for research in the area. On multiple occasions, the authors note that a primary reason why stereotype threat research has not extended into the learning domain is because "learning is difficult to distinguish from performance" (Rydell, Rydell, & Boucher, 2010, p. 885). This sentiment suggests that a more clearly defined conceptual framework of the primary dependent variable would be of great benefit. Without such a theoretical foundation, issues related to construct validity become a significant concern in the exploration of stereotype threat effects during learning. For instance, the Rydell studies employ both judgments of participants' procedural descriptions (Study 2) and response times (Studies 4-5) as indicators of learning both of which are fairly subjective and can be easily confounded by external characteristics (e.g., verbal/writing skills of participants in Study 2, accuracy of computational model for describing visual search task in Studies 4-5, etc.) that do not accurately reflect underlying changes in the learning. Additionally, the assessments of learning in the focal and transfer learning tasks assessed in Studies 1 and 2 could have just as easily been considered measures of task performance rather than changes in learning, an issue explored further in the following section.

Perhaps of greater import though, the learning indicators used in many of these studies were not often consistent with nor effective at demonstrating how the proposed mechanisms of stereotype threat impacted knowledge/skill acquisition efforts. For instance, Rydell, Rydell, and Boucher (2010) stated that stereotype threat affects mathematical learning "by reducing [a threatened individual's] ability to encode mathematical information into memory, not by inhibiting the ability to retrieve mathematical information from memory." Irrespective of research which suggests working memory and therefore impediments to working memory do interfere with the process of retrieving/decoding information from memory (Rosen & Engle, 1997, 1998), the use of measures which require participants to explicitly recall learned material (e.g., Studies 1 and 2) most certainly taps into memory retrieval processes and thus dilutes observations of threat's effects on encoding efficiency. Additionally, Rydell, Shiffrin et al. (2010) hypothesize in Studies 4-6 that stereotype threat inhibited women's perceptual learning because it impeded the development of more efficient visual search strategies ("popout" and character unitization, Shiffrin & Lightfoot, 1997; Shiffrin & Schneider, 1977). They later confess, though, that their choice to only model changes in response times as an indicator of learning (Studies 4-5) does not allow them to conclude whether learning of these visual search processes was impeded by the introduction of threat—instead they can simply infer that "Whatever had been learned by women in the control group, it seems not to have been learned by women under [stereotype threat]" (p. 14046).

**Research/practical implications of stereotype threat at learning.** In sum, the Rydell studies simultaneously provide an important proof of concept and demonstrate that the implementation and operationalization of learning experiences is an important consideration to the study of stereotype threat effects on knowledge/skill acquisition. Fortunately, there is well

over 100 years' worth of accumulated literature on learning behaviors, outcomes, and techniques available to inform investigations of threat at learning. The tripartite classification scheme developed by Kraiger et al. (1993) has been advanced here as one particularly useful theoretical perspective for organizing investigations of threat effects on the learning process due to its theoretically grounded and intuitive description of possible learning foci and constructs. As shown in Table 1, translating the learning outcomes assessed by Rydell and colleagues through this conceptual lens reveals that Studies 1-3 appeared to focus on threatened individuals' ability to learn declarative knowledge (facts/statements/rules about mathematical or logical operators) while Studies 4-6 investigated the acquisition of skill-based learning outcomes and the ability for individuals to compile and automate new visual search strategies (popout and character unitization). Rydell and colleagues' initial efforts therefore examine only a small portion of the possible construct space within which stereotype threat may influence learning (especially with respect to cognitive learning outcomes, the learning domain arguably most relevant to traditional applications of stereotype threat, e.g., Steele, 1997; Steele & Aronson, 1995). As such, integrating even basic frameworks of learning such as Kraiger et al. (1993) offers a useful contribution to future research efforts in this area-and reveals there is still significant empirical ground to cover with respect to understanding stereotype threat at learning.

Rydell, Rydell, and Boucher's (2010) investigation of stereotype threat effects on cognitive learning outcomes also exemplifies another unique and troubling methodological issue for continued research in this area. Specifically, although declarative knowledge acquisition serves as the foundation for the development of more advanced learning (Anderson, 1982, 1993a, 1996), the most common assessments of declarative knowledge *acquisition* (multiple-choice, true-false, or free recall tests, cf., Kirkpatrick, 1976, 1987; Kraiger et al., 1993) are identical to

those used to demonstrate stereotype threat effects on cognitive *performance* outcomes (e.g., Grand et al., 2011; Nguyen et al., 2003; Ployhart et al., 2003; Spencer et al., 1999; Steele & Aronson, 1995, etc.). This is problematic given that (1) stereotype threat hijacks available cognitive resources from working memory (Schmader et al., 2008; Schmader, 2010), (2) working memory capacity influences both the encoding (Feldman Barrett et al., 2004) and retrieval (Rosen & Engle, 1997, 1998) of declarative knowledge to/from longer-term memory stores, and (3) virtually all self-report assessments of declarative knowledge involve memory retrieval mechanisms (cf., Kirkpatrick, 1976, 1987). Therefore, pinpointing the manner by which stereotype threat affects the acquisition of declarative knowledge through traditional "learning assessments" becomes analytically untenable as such measurement approaches are ill-equipped to disentangle threat's effects on cognitive encoding, storage, and synthesis mechanisms (characteristics associated with learning, e.g., Kraiger et al., 1993) from its effects on cognitive retrieval, integration, and manipulation mechanisms (characteristics associated with performance, e.g., Schmader et al., 2008)<sup>4</sup>.

As Figure 3 highlights though, knowledge structure organization and procedural strategy formulation represent two alternative possibilities for examining the influence of stereotype threat on the acquisition of cognitive learning outcomes. Although changes in these outcomes are posited to be more "advanced" consequences of one's learning experiences (Kraiger et al., 1993) and will, to some extent, be influenced by the acquisition and retrieval of declarative knowledge (Anderson, 1982, 1993a, 1996; Anderson et al., 2004), examining these learning indicators holds certain advantages over more traditional measures of learning proficiency. For example, knowledge structures can be assessed by asking respondents to make relational ratings among relevant domain/task concepts that are presented to respondents and do not necessarily require

one to explicitly recall declarative facts (e.g., Schvaneveldt, 1990). As a result, one can lessen the "double-dipping" problem of stereotype threat effects on encoding and retrieval processes that would taint investigations focused only on declarative knowledge while still examining and interpreting the influence of stereotype threat on learning outcomes.

Additionally, the development of effective knowledge structures and cognitive performance strategies hold important implications for a variety of performance outcomes as well. At a conceptual level, the relation between performance and knowledge has been widely acknowledged. Campbell, McCloy, Oppler, and Sager (1993) characterize job performance as a multiply determined, integrative function of domain-/goal-relevant declarative knowledge, procedural knowledge/skill, and motivation. More generally, the ACT-R model of cognition (cf., Anderson, 1996; Anderson et al., 2004) posits that virtually all demonstrative instances of intellective performance result from the encoding of environmental stimuli into feature-rich information "chunks" (declarative knowledge) that feed production rules (i.e., procedural knowledge) and guide the generation of task-relevant outcomes. Inherent in both of these conceptualizations of performance, however, is the notion that the translation of basic knowledge to actionable knowledge (i.e., translating information about what to do into information about how to do it, when to do it, and why it's done in that manner) is central to performing any task. Empirical evidence of the relation between knowledge and task outcomes further supports the notion that knowledge structures and cognitive strategies are integral in determining the manner by which one contextualizes, approaches, and undertakes performance-relevant activities (e.g., Day et al., 2001; Kozlowski, Gully, et al., 2001; Medin et al., 2006; Royer, Tronsky, Chan, Jackson, & Marchant, 1999; Zentall, 1999).

The above also suggests that the influence of stereotype threat on the development of advanced cognitive learning outcomes could impact the performance *potential* of threatened individuals. For example, even if one were to find no significant performance differences between threatened and non-threatened individuals on a declarative knowledge-type assessment in a performance domain, threatened individuals may possess less well organized or proceduralized knowledge within that domain. As a result, they may take longer to do the same tasks, expend greater cognitive resources to achieve the same performance levels, or be less capable/require greater levels of investment to learn related tasks than non-threatened others. Such consequences could potentially lead to a host of undesirable outcomes in the long run, such as quicker burnout within a domain, stagnated performance growth, and fewer opportunities for advancement/promotion—especially in areas with both prevalent group stereotypes and rapidly-paced learning environments (i.e., STEM fields).

In sum, the importance of working memory to the acquisition/retention of task-relevant information, behaviors, and dispositions strongly implies that the learning efforts of threatened individuals may be undermined by the working memory decrements stemming from stereotype threat. The use of Kraiger et al.'s (1993) classification system in this context indicates a variety of possible learning outcomes which may be susceptible to threat effects and further serves as a useful organizing framework for systematically advancing research in this area beyond previous works. Although the present study focuses only on the acquisition of advanced cognitive learning outcomes, stereotype threat may impact equally important components related to skill-based (e.g., inability to adapt to effective behavioral routines, slower to automatize procedures and thus develop expertise) and affective (e.g., smaller self-efficacy improvements, greater resistance to error-based training, etc.) learning outcomes as well. Investigations of the effects of stereotype

threat at learning thus potentially carry a number of methodological and practical implications, including a better understanding of how threat experienced at performance differs or is similar to threat experienced during learning, how to improve training/learning paradigms to help targeted individuals learn and engage information from stereotyped content domains, and further explication of the boundary conditions, cognitive mechanisms, and alternative consequences of stereotype threat beyond intellective testing (cf., Nguyen & Ryan, 2008).

#### **Research Hypotheses**

The preceding description of the nature of stereotype threat and the theoretical rationale for the potential influence of stereotype threat at learning provides the bulk of the conceptual underpinnings for the hypotheses presented below. In the present study, the primary predictions of interest concern the effects of stereotype threat on participants' knowledge organization, cognitive strategy formulation, and subsequent task performance. Given that the elicitation of stereotype threat is conditioned on the subgroup comparison and stereotyped domain of interest, it is pertinent to briefly make note of these aspects. In the stereotype threat conditions, the purpose of the study will be presented to participants as an examination of why some subgroups have greater difficulty on mathematical reasoning assessments. Males and females will serve as the subgroup comparison of interest as stereotypes and empirical performance differences favoring men are commonly documented within this ability domain, especially amongst collegeeducated populations, and are commonly recognized by most individuals in Western cultures (Ackerman, Bowen, Beier, & Kanfer, 2001; Beilock et al., 2010; Halpern, 2000; Halpern et al., 2007; Hyde, Fennema, & Lamon, 1990). Furthermore, females' performance within mathematical domains has been shown to be reactive to and produce patterns consistent with stereotype threat in previous research (e.g., Spencer et al., 1999) and can be elicited even in

instances where the task itself does not explicitly involve mathematical operations/content (e.g., Jamieson & Harkins, 2007; Rydell, Shiffrin, et al., 2010).

Stereotype threat and knowledge organization. Knowledge organization represents the way in which individuals form and store relationships among declarative facts, concepts, propositions, data, and other objects within a given task domain (e.g., Jonassen et al., 1993; Koubek, Clarkston, & Chavez, 1994; Shavelson, 1972, 1974; Taber, 2000). It is generally believed that such knowledge structures reflect an individual's deeper understanding of the manner by which task requirements are fulfilled as well as the knowledge, skills, and procedures necessary to achieve them (Glaser, 1990; Rowe et al., 1996; Schoenfeld & Herrmann, 1982). Assessments of knowledge structures have most commonly been used to examine betweenperson differences in knowledge/skill acquisition and as a means to differentiate between domain experts and novices (e.g., Chi, Glaser, & Farr, 1988; Day et al., 2001; Ford & Kraiger, 1995). However, knowledge structures can also be used to investigate within-person changes in information acquisition/synthesis as individuals accumulate experience in a given content domain (Rumelhart & Norman, 1978), although longitudinal examinations of this process are rare (Ifenthaler, Masduki, & Seel, 2011; Ifenthaler & Seel, 2005; Seel, 1999).

In order to characterize the hypothesized relationship between stereotype threat and knowledge organization, it is useful to describe the manner by which knowledge structures will be operationalized and constructed in the present study. Knowledge structures can be elicited from individuals in a variety of ways. Ifenthaler et al. (2011) broadly classify these different techniques into either natural language (e.g., thinking-out-loud protocols, word association, card sorting, etc.) or graphical (concept mapping, causal diagrams, etc.) approaches. Each methodology possesses different strengths and weaknesses, though graphical data gathering and

analytic approaches have become increasingly popular due to the ease with which they can be used to quantitatively and qualitatively represent structural knowledge formations (Goldsmith, Johnson, & Acton, 1991; Schuelke, 2009). Among these graphical approaches, the Pathfinder algorithm (Schvaneveldt, 1990) and its associated statistical software (Interlink, 2011) is among the most familiar and commonly used technique in the learning and training literature and the one that will be adopted for the present study.

In brief, Pathfinder reconstructs structural networks amongst a set of concepts based on similarity or proximity ratings provided by an individual for all possible pairs of concepts in the set. Each rating is interpreted as how strongly a pair of concepts is related in a person's memory (Nagy, 1984). The Pathfinder algorithm then identifies the most parsimonious relationships shared amongst all concepts in order to form a structural network composed of nodes (concepts), links (a single relationship between two concepts), and paths (an indirect relationship between any two concepts composed of x number of links). Analytically, the algorithm operates by first linking all concepts in the network together and then removing direct links between any two concepts if a stronger indirect link (i.e., a link that passes through another node that lies between the target and destination node) exists. Two parameters can be manipulated in the network generation algorithm to control the production of links and paths among nodes (Dearholt & Schvaneveldt, 1990): r (from Minkowski's distance formula), which determines how the distance between nodes not directly linked is computed, and q, which places a limit on the number of links that can exist in a path between any two nodes in the network. Similar to previous studies employing Pathfinder (e.g., Day et al., 2001; Schuelke et al., 2009; Kozlowski, Gully, et al., 2001; Kraiger, Salas, & Cannon-Bowers, 1995), the present analyses evaluated structures in which  $r = \infty$  (the weight of a path is equal to the maximum weight of any link in the

path) and q = n-1, where *n* equals the number of concepts in the network. These parameter values tend to produce the most parsimonious structures while still allowing maximal interconnectivity among nodes in the network.

Pathfinder networks can also be used to compute two sets of statistical indices useful to characterizing the organizational efficiency of observed knowledge structures. The first of these are the *similarity* and *correlation* indices and are used to evaluate comparative similarity across different knowledge structures (e.g., a structure derived from a novice versus one derived from a domain expert, etc.). Similarity reflects the extent to which two knowledge structures share the same pattern of linkages among concepts, whereas correlation indices represent the extent to which these associations share consistent rank-order priorities (Schuelke et al., 2009; Smith-Jentsch, Mathieau, & Kraiger, 2005). In addition to these referent-based indices, the second set of numeric indices provides information about the structural characteristics of any single knowledge map. Number of links yields information about the complexity/parsimony of a knowledge structure and can be used to evaluate the degree to which particular concepts are more/less related to others (Day et al., 2001). Alternatively, *coherence* is a measure of internal consistency depicting the extent to which concepts share logical relations based broadly on assumptions of transitivity (i.e., if two concepts share similar relations with other concepts, then those two concepts should be similar to each other, Interlink, 2011). In addition to these numerical outputs, Pathfinder networks can be used to examine the manner by which concepts are *clustered* in individual's knowledge structures. The derivation of knowledge structures in Pathfinder's software incorporates features similar to those used in hierarchical clustering analyses and multidimensional scaling techniques; as a result, the relative location of concepts in a visualized Pathfinder graph can also be used to deduce semantic categorization of knowledge

(Dearholt & Schvaneveldt, 1990; Esposito, 1990). Thus, similarity, correlation, number of links, coherence and clustering will be examined as the primary knowledge structure output of interest.

Based on the rationale that stereotype threat impedes working memory efficiency by hijacking available capacity for task-irrelevant demands (i.e., Beilock et al., 2006; Beilock et al., 2007; Schmader, 2010; Schmader et al., 2009; Schmader & Johns, 2003) and working memory capacity is related to learning and related information encoding processes (cf., Feldman Barrett et al., 2004), stereotype threat should negatively influence knowledge structure formation by inhibiting threatened individuals' ability to maintain enough task-relevant information in an activated/accessible state for a sufficient duration of time needed to develop an integrated mental representation of the content. Developing coherent and proficient knowledge structures is a cognitively demanding and effortful process that requires individuals to successfully coordinate information about features, meanings, and associations simultaneously across multiple concepts. To the extent that one's capacity to actively engage and efficiently direct attentional resources towards these activities is inhibited, the formation of a comprehensive knowledge structure indicative of a deep understanding of the task domain should be affected.

As indirect evidence of this proposition, MacDonald, Just, and Carpenter (1992) and Whitney et al. (1991) both report evidence consistent with the claim that persons with lower working memory capacity are less able to maintain as many pieces of information in active attention long enough to derive alternative possible meanings for written prose or disambiguate multiple semantic interpretations, both of which are indicative that one has formed more complex associations among available information. More directly relevant to the relationship between working memory and knowledge structure formation, Cantor and Engle (1993) posited that individuals learn declarative knowledge by creating semantic memory networks containing

the underlying relational propositions among those chunks (cf., Anderson, 1996). However, in order to form such mental representations, a significant portion of those information bits must remain in an activated state during learning in order for individuals to draw associations between their shared features (Johsnon-Laird, 1983; Radavansky & Zacks, 1991). The development of these schemas facilitates task functioning in that they allow individuals to access the majority of information about a context by activating only a single mental representation containing all the underlying propositional relations among the information rather than actively retrieving each of those facts separately and independently (Reder & Anderson, 1980). Thus, well-formed knowledge structures enable one to retrieve more information faster and with less attentional demand. To the extent that working memory is related to the development of these mental models, one would expect to see differences in how readily individuals with differing levels of working memory capacity could learn and recall large amounts of rote declarative information (Johnson-Laird, 1983).

In a series of studies examining these predictions, Cantor and Engle (1993) revealed that individuals with lower working memory capacity were (A) slower at correctly identifying learned sentences as the number of common features they shared with other learned sentences increased and (B) slower at correctly identifying learned sentences as the number of sentences in the pool of related sentences to-be-learned increased. Comparatively, individuals with higher working memory were much quicker at completing (A) than their lower-memory counterparts and became faster at (B) as the learning pool increased. Although not a direct examination of mental model formation per se, these results are consistent with the claim that more comprehensive knowledge structures facilitate task functioning and individuals with lower working memory capacities may have greater difficulty maintaining enough data in working

memory during learning efforts necessary to efficiently integrate the semantic content of and relational associations among task-relevant concepts.

Thus, the inability to make use of one's full reserve of working memory capacity should disproportionately influence a threatened individual's ability to actively maintain enough information related to task goals/requirements and for the duration of time needed to develop comprehensive associations among knowledge concepts indicative of advanced learning. Consequently, working memory decrements engendered by stereotype threat should exert a stronger influence on the development of integrated, sophisticated, and well-organized knowledge structures in threatened individuals compared to non-threatened individuals. In the context of the present study, men are not expected to be influenced by the stereotype threat manipulation and thus provide a reasonable control condition against which to examine differences in knowledge structure formations between threatened and non-threatened women.

*Hypothesis 1*: The knowledge structures of females who learn under conditions of stereotype threat will be less similar to those from top performers/men than the knowledge structures of females who learn under control conditions.

*Hypothesis 2*: The knowledge structures of females who learn under conditions of stereotype threat will be less correlated with those from top performers/men than the knowledge structures of females who learn under control conditions.

*Hypothesis 3*: The knowledge structures of females who learn under conditions of stereotype threat will be less coherent than the knowledge structures of females who learn under control conditions.

*Hypothesis 4*: The knowledge structures of females who learn under conditions of stereotype threat will have significantly more links (i.e., be less parsimonious) than the knowledge structures of females who learn under control conditions.

*Hypothesis 5*: The clustering of concepts in the knowledge structures of females who learn under conditions of stereotype threat will be significantly different than that for non-threatened women; specifically, the knowledge structures of females in the stereotype threat conditions will exhibit poorer integration of related task concepts (i.e., report fewer associations between task concepts whose meanings are mutually informative and/or relevant to task performance) than the knowledge structures of women in the control condition.

Given the inherently dynamic process through which the accumulation of learning experiences and the development of cognitive learning outcomes proceeds (Anderson et al., 2004; Goldstein & Ford, 2002; Kraiger et al., 1993), investigating changes in knowledge structure formation over time marks an important contribution to the understanding of stereotype threat effects on learning. When attempting to learn a novel/complex domain task, adaptations to one's knowledge structure should emerge as individuals receive and make use of subsequent learning opportunities to better understand domain concepts and their associations (e.g., Ifenthaler et al., 2011; Jonassen et al., 1993; Kraiger et al., 1993; Rumelhart & Norman, 1978). Furthermore, to the extent that practice and learning exposures enable individuals to improve their understanding of domain concepts and how to effectively complete tasks based on those concepts, knowledge structures should show some degree of convergence towards a singular "optimal" configuration (or small subset of optimal configurations, depending on the nature of the task or the manner by which it is learned, Medin et al., 2006) over time as individuals become knowledgeable, acquire expertise, and become better at performing domain tasks (e.g., Chi et al., 1988; Day et al., 2001). However, it is widely acknowledged that early stages of learning exert significant influence on subsequent learning efforts (e.g., Anderson et al., 2004; Bell & Kozlowski, 2002; Goldstein & Ford, 2002). To the extent that threat-based working memory decrements interfere with the formation of knowledge structures early in the learning process then, the learning difficulties of threatened individuals may compound over time, resulting in more stagnant structures that are much slower to or which never converge towards models that represent efficient, high performing "expert" models.

In the present study, the knowledge structures of individuals were evaluated on three consecutive days in an attempt to evaluate differences in learning growth between threatened versus non-threatened individuals. In addition to the contributions that longitudinal examinations of knowledge structure development hold for learning researchers in general (cf., Ifenthaler et al., 2011), these results hold a number of possible implications for design and evaluation considerations of instructional systems in stereotyped domains (e.g., value added of additional training exposures, expected success/retention, etc.)—especially in environments where individuals are expected to quickly gain expertise in one knowledge/skill area before moving to more advanced applications (e.g., mathematics education, technology training, etc.).

*Hypothesis 6*: The similarity between the knowledge structures of females who learn under conditions of stereotype threat with those from top performers/men will improve at a slower rate compared to females who learn under control conditions.

*Hypothesis* 7: The correlation between the knowledge structures of females who learn under conditions of stereotype threat with those from top performers/men will improve at a slower rate compared to females who learn under control conditions.

*Hypothesis 8*: The coherence of the knowledge structures of females who learn under conditions of stereotype threat will improve at a slower rate compared to females who learn under control conditions.

*Hypothesis 9*: The number of links in the knowledge structures of females who learn under conditions of stereotype threat will increase at a faster rate (i.e., structures will become less parsimonious) compared to females who learn under control conditions.

*Hypothesis 10*: The knowledge structures of females who learn under conditions of stereotype threat will demonstrate less integration of related task concepts over time (i.e., fewer and less efficient associations between related task concepts) compared to females who learn under control conditions.

Stereotype threat and cognitive strategy. As a general learning outcome, cognitive strategy development describes the internalization and manifestation of cognitive/behavioral processes, procedures, and heuristics that direct one's efforts towards accomplishing a given goal (Anderson, 1982; Kanfer and Ackerman, 1989; Prawat, 1989). Acquiring effective cognitive strategies is a relatively advanced outcome of the learning process that usually only develops after one has developed a deeper level of understanding about a task/domain and its requirements (Kraiger et al., 1993; Sweller, Mawer, & Ward, 1983). It is perhaps not surprising to note, then, that knowledge organization and cognitive strategies are somewhat interdependent. Knowledge

structures provide broadly construed mental schema for deducing cause-effect relations within a domain space, which individuals then learn to effectively employ and integrate with task-specific conditions (i.e., rules, demands, criteria, etc.) to form cognitive strategies that direct future learning and performance efforts (e.g., Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Rees, 1982; Simon & Simon, 1978). As one's cognitive strategies are vetted and feedback about their success gathered, subsequent refinements about the relations and clustering among knowledge concepts may also occur.

Consistent with Kraiger et al.'s (1993) framework, stereotype threat conditions which adversely impact the formation of effective knowledge structures likely also impair the acquisition of effective cognitive strategies as such strategies would be based on a less sophisticated or complete understanding of the content domain. However, as implied above, cognitive strategy acquisition also requires individuals to simultaneously attend to unique features of a given task in order to incorporate those task demands/needs with one's conceptual understanding of the domain. Again, working memory functions are believed to play an integral role in this monitoring and integration procedure. Johnson-Laird (1983) succinctly elucidates this intersection:

The effects of both number of [mental] models and figure [i.e., task requirements] arise from an inevitable bottleneck in the inferential machinery: the processing capacity of working memory, which must hold one representation in a store, while at the same time the relevant information from the current premise is substituted in it. (p. 115)

The development and selection of these inferential strategies is purportedly coordinated by the executive control and episodic buffer functions of one's working memory (Baddeley, 1986; 2000). During such activities, individuals learn to activate representative knowledge structures and cognitive areas of functioning that assist in reconciling current conditions/needs in the task space (e.g., Johnson-Laird, 1983; Wraga et al., 2007). As individuals streamline this process and

become more proficient at its application over time, the coordination between knowledge structure retrieval and task demand processing also becomes a less effortful undertaking (Smith, McEvoy, & Gevins, 1999), freeing working memory resources to monitor and respond to other stimuli in the learning environment. However, threatened individuals who may be employing sparsely developed knowledge structures to begin with while also relying on limited working memory capacities to coordinate these inferential processes should be less likely to see such effective strategies emerge and/or improve during learning. A threatened individual with diminished working memory capacity should therefore have greater difficulty holding learned knowledge structures in active awareness, learning/interpreting conditional features of a task space, and integrating those pieces into effective strategies that dictate how those representations should be applied to most effectively operate within a given context.

In support of this proposition, a number of studies report strong correlations between working memory capacity and cognitive strategy use/selection (e.g., Anderson, Reder, & Lebiere, 1996; Barrouillet, Bernardin, & Camos, 2004; Barrouillet & Lépine, 2005; Dunlosky & Kane, 2007; Espy et al., 2004; Gilhooly, Logie, Wetherick, & Wynn, 1993; McNamara & Scott, 2001). For example, research on arithmetic skill development in elementary school children reveals that those with lower working memory capacities tend to have more difficulty learning and employing advanced problem-solving tactics. Furthermore, individuals with less working memory capacity are less likely to adapt to using more effective strategies in response to increased complexity in task demands (Geary, Hoard, Byrd-Craven, & DeSoto, 2004; Imbo & Vandierendonck, 2007). Interestingly, these findings are consistent with the interpretations drawn by Rydell, Shiffrin et al. (2011) that threatened individuals (whose working memory capacities are presumably diminished) tended to persist in suboptimal visual search strategies

over the course of multiple experimental trials rather than acquiring new and more efficient approaches to task completion. As such, threatened women are predicted to develop less effective and advanced cognitive task strategies than non-threatened women.

*Hypothesis 11*: Females who learn under conditions of stereotype threat condition will exhibit poorer/more basic cognitive task strategies than females who learn under control conditions.

For many task domains, cognitive strategies also reflect one's methods for identifying particular pieces of declarative knowledge and optimally interpreting that information according to procedural rules/knowledge. For example, developing expertise in a number of educational (solving mathematical word problems, completing verbal reasoning tasks, etc.) and practical (e.g., technology troubleshooting, providing task orders/direction in an emergency room, etc.) applications occurs as individuals learn how to distinguish relevant from irrelevant information, interpret its relative value to the desired performance outcome, and correctly formulate interpretations based on that information (Sweller et al., 1983). Part of the learning process involved in such tasks, therefore, is the development of strategic heuristics that explicitly orient these information gathering and combination efforts in the most efficient manner. However, these cognitive tasks may be more difficult for threatened individuals as they have less working memory capacity available to assist in information screening and evaluation activities (Schmader et al., 2003; Schmader et al., 2008). As a result, these persons may be more likely to attend to irrelevant declarative knowledge facts during learning and incorporate this information into suboptimal performance strategies. It is therefore predicted that:

*Hypothesis 12*: Females who learn under conditions of stereotype threat will develop less optimal procedural decision strategies for task completion than females who learn under control conditions.

Stereotype threat and performance. The overwhelming majority of stereotype threat research has been directed towards identifying decrements in ability test performance (cf., Nguyen & Ryan, 2008). Though stereotype threat has been shown to influence outcomes from other sensorimotor and social tasks (e.g., Stone 2002; Stone et al., 1999; Stone & McWhinnie, 2008; Kray et al., 2002), few investigations have examined threat effects on performance outcomes in computer-based simulation tasks. However, simulation use has become an exceedingly common method of assessment and training in educational and industry areas (Bell, Kanar, & Kozlowski, 2008) and thus represents an important extension and application of stereotype threat research. Advances in technology have made the design, maintenance, and implementation of realistic job/educational simulations more affordable and accessible to decision-makers than ever before (Bell & Kozlowski, 2007), leading many industries to adopt these tools as part of their everyday repertoire. For example, one survey claims that upwards of 97.5% of business schools use some form of simulation gaming as part of their standard curricula (Faria, 1998). Furthermore, it was estimated that at least 75% of organizations in the United States with more than 1,000 employees use business simulations for hiring/training purposes (Faria & Nulsen, 1996) and that between \$623 and \$712 million dollars in global revenue was generated by the simulation-based training industry in 2003 (Summers, 2004).

In the present study, a computer-based simulation task will serve as the learning and performance environment for participants. Many simulations are designed to teach, evaluate, and/or relate to skills and abilities comparable to those targeted by traditional mediums (e.g.,

spatial ability, mathematical ability, etc.); it is therefore plausible that individuals would believe that a given simulation is indicative of a particular capability if so informed even if, on its face, that simulation does not appear to be a direct measure of the ability in question (e.g., Rydell, Shiffrin, et al., 2011). Thus, the same performance discrepancies predicted by the broader stereotype threat literature on standard ability tests between threatened and non-threatened individuals are also expected in the present simulation-based task. Additionally, because stereotype threat is expected to produce learning difficulties within the task domain, it is also predicted that the performance trajectories across conditions of threat will differ over time.

*Hypothesis 13*: Females who learn under conditions of stereotype threat will demonstrate worse performance on the learned task than females who learn under control conditions.

*Hypothesis 14*: Females who learn under conditions of stereotype threat will improve their performance on the learned task at a slower rate than females who learn under control conditions.

#### **METHOD**

### **Participants**

Participants were 198 undergraduate students (M age = 19.49, SD = 2.04) from psychology courses at a large Midwestern university. All individuals were informed that the experiment spanned three consecutive days and that interested persons should only volunteer for the experiment if they were able to complete the study in its entirety. Table 2 provides a breakdown of the number of participants who completed each day of the experiment by sex and experimental condition. Given the focus on negative stereotypes towards female achievement in mathematical tasks and the within-group nature of the stereotype threat theory/predictions, women were the primary group of analytic interest; as such, they were purposely overrepresented in the sample relative to men. Across the entire sample, 79.8% of participants completed all three days of the experiment, with a slightly higher attrition rate occurring between Days 2 and 3 relative to Days 1 and 2. A two-way analysis of variance (ANOVA) on number of days attended by participants in the study revealed no significant main effects for sex (F(1,198)) = .228, ns) or condition (F(1,198) = .013, ns), nor was there a significant sex by condition interaction (F(1,198) = .820, ns). These results indicate that the observed attrition rates were not differentially influenced by the stereotype threat manipulation nor did they differ for males and females in the experimental conditions.

For their participation in the study, all individuals were compensated with course credit; as additional incentive, participants were also informed that the top 10% of performers in the experiment would receive a \$60 cash prize. Because women facing stereotype threat were expected to do more poorly in the study overall, the cash prizes were awarded to the top 10% of performers within each of the 2 (sex: male, female) x 2 (condition: stereotype threat, control)

	Stereotype Threat		Control		
-	Females	Males	Females	Males	- I otai
Day 1	71	27	74	26	198
Day 2	65	22	69	24	180
Day 3	58	22	56	22	158
% Attrition from Day 1 to Day 2	8.45%	18.52%	6.76%	7.70%	9.09%
% Attrition from Day 2 to Day 3	10.77%	0%	18.84%	8.33%	12.22%
% Attrition from Day 1 to Day 3	18.31%	18.52%	24.32%	15.38%	20.20%

Table 2Total Sample Size and Attrition Rates across Days by Sex and Experimental Condition

experimental design cells to ensure that all participants had a fair chance of earning the award. To be eligible for the cash prize, participants were required to attend all three of the scheduled sessions as described in the experimental procedure below.

### **Experimental Task**

A modified version of the Tactical Naval Decision Making system (TANDEM, Weaver, Bowers, Salas, Cannon-Bowers, 1995) was used as the experimental platform. TANDEM is a complex, dynamic, information-processing and decision task set in the context of a low fidelity radar-tracking simulation. The TANDEM task paradigm has been used to study a variety of phenomena, and has a particularly rich history in investigations of learning and self-regulation at both the individual- and team-level (e.g., Bell & Kozlowski, 2002; 2008; DeShon, Kozlowski, Schmidt, Milner, & Wiechmann, 2004; Inzana, Driskell, Salas, & Johnston, 1996; Kozlowski, Gully, et al., 2001). TANDEM is well suited for the purposes of the present study as it requires participants to learn basic declarative facts which share clearly identifiable relations as well as more advanced procedural strategies in order to effectively complete the simulation. Furthermore, the nature of the task environment and associated cognitive load creates a demanding operational environment for learners, characteristics which have been shown to exacerbate the salience of stereotype threat and thereby improve the likelihood of observing its effects on learning outcomes (cf., Steele, 1997; Steele & Davies, 2003).

The primary objective of TANDEM is to earn points by accurately and quickly identifying, evaluating, and making decisions regarding what action to take against targets that appear on one's computer screen. Participants are presented with a circular radar display which shows multiple targets in motion around a central radar-tracking station (Figure 4). Targets are either present on the screen from the beginning of a trial or appear ("pop-up") after some period of time elapses. Within the radar space are two defensive perimeters: an inner perimeter clearly marked on screen, and an outer perimeter that is not visible. In the current task design, the location of the outer perimeter could be approximated by expanding the display window ("zooming out") to 256 NM (nautical miles) and locating a ring of six stationary targets designated as "markers." The marker targets were identical in appearance to all other contacts except they did not move and engaging would them would neither earn nor lose points.

Points were gained for every target correctly prosecuted and lost for every target incorrectly prosecuted or which crossed into one of the defensive perimeters. To prosecute a contact and earn points, participants needed to (1) "hook" a target by selecting it with the mouse cursor, (2) view and interpret cues that provided information about various decision-relevant characteristics of that target (i.e., Speed, Direction of Origin, Countermeasures, etc.), (3) make three subdecisions based on those cue values, and (4) indicate a final engagement decision



Figure 4. TANDEM graphical user interface

based on the selected subdecisions. To earn points, all three subdecisions and the final engagement decision needed to be correct; making even one of these decisions incorrectly resulted in a loss of points.

As shown in Table 3, each subdecision (corresponding to a target's Type, Class, and Intent) was informed by three cues whose values were uniquely associated with a single subdecision outcome. For example, when making the Type subdecision, participants could classify the target as an Aircraft, Surface, or Submarine vessel. A target whose Speed was greater than 35 knots, Altitude/Depth greater than zero feet, and communication time between one and forty seconds was classified as an Air vessel; similarly, a Speed between 25 and 34 knots, an

Subdecision	Subdecision Outcomes	Identifying Cues & Cue Values			
		Speed	Altitude/Depth	Communication Time	
Type	Air	$\geq$ 35 knots	> 0 feet	1-40s	
	Surface	25-34 knots	0 feet	41-80s	
	Sub	0-24 knots	< 0 feet	81-120s	
		Countermeasures	Signal Strength	Maneuvering Pattern	
Class	Civilian	None	Moderate	Code Foxtrot	
	Unknown	Inactive	Indistinct	Code Echo	
	Military	Jamming	Weak	Code Delta	
		Identification	Direction of Origin	Response	
Intent	Peaceful	Prince	Green Beach	Authorized	
ment	Unknown	Golf	Blue Lagoon	Inaudible	
	Hostile	Tango	Orange Bay	Invalid	

Table 3Subdecision Outcomes and Relevant Identifying Information Cues/Values in TANDEM

*Note.* Some targets possessed cue values for the Class and Intent subdecisions which identified them as Unknown; however, participants were not permitted to classify a target's Class or Intent as Unknown.

Altitude/Depth of zero feet, and a communication time between 41 and 80 seconds were indicative of a Surface contact. The same protocol was used to identify a target's Class and Intent, though each of these subdecisions possessed only two possible outcomes (Civilian or Military for the Class subdecision, Peaceful or Hostile for the Intent subdecision). Once all three subdecisions were made, participants could make the final engagement decision for a target based on the "rules of engagement" shown in Table 4. The rules indicated how a target should be prosecuted according to its Type, Class, and Intent; thus, a target whose Type was Air, Class was Civilian, and Intent was Peaceful should be Warned, whereas a target that was a Sub, Military, Hostile should be Marked.
Clear	Warn	Mark
Air, Military, Peaceful	Air, Civilian, Hostile	Air, Military, Hostile
Surface, Civilian, Peaceful	Air, Civilian, Peaceful	Surface, Military, Hostile
Surface, Military, Peaceful	Surface, Civilian, Hostile	Sub, Civilian, Hostile
Sub, Military, Peaceful	Sub, Civilian, Peaceful	Sub, Military, Hostile

Table 4Rules of Engagement for Determining Final Engagement Decisions

A comprehensive operations manual was available to participants prior to each trial that contained all the relevant information needed to operate TANDEM. The information in the manual could be categorized into three major topic areas: basic gameplay information, cue value interpretation, and task strategies. The sections describing basic gameplay included information on the computer functions needed to operate the task (i.e., how to hook targets, access cue menus, etc.), scoring rules, and the task objectives/background context (i.e., engaging contacts quickly and accurately, etc.). Material pertaining to cue value interpretation included the cue values needed to make accurate subdecisions, the rules of engagement, and the order in which task decisions needed to be made. This section also indicated that the radar sometimes provided conflicting/ambiguous cue values for a target (i.e., two cue values supported one decision while the remaining cue value supported a different decision) and that, in such cases, the option supported by the majority of cues should be selected. Lastly, details about task strategies included advanced/tactical aspects of the task related to perimeter identification (using the zoom functions, locating/using marker targets to identify the invisible defensive perimeter) and target prioritization (gauging contact speed and location relative to perimeters, switching between defensive perimeters, cost of perimeter breach).

### Procedure

**Online signup.** Individuals were recruited and registered to participate in the study through the Psychology Department subject pool website. In addition to basic information about the experiment, the online study description/recruitment materials described incentives for participation and indicated that participants would be eligible to receive a monetary award for completing all portions of the study. Approximately 15 participants were permitted to sign up for a single experimental session at a time, though efforts were made to ensure that each session contained both male and female participants. Each experimental session was assigned to either the control or stereotype threat condition with attempts to maintain a balanced sample size across both conditions. After signing up for the experiment, individuals completed an online consent form (Appendix A) as well as a short questionnaire containing a small number of background/demographic items and a survey on math domain identification. In total, the sign-up procedure and questionnaire lasted approximately 10-15 minutes. Upon completing the online portion of the experiment, participants were directed to attend the computer lab where they would play the TANDEM simulation on the first day of their scheduled experimental session. A reminder e-mail was sent to all participants approximately one week before the date of their lab session which provided the dates, times, and room number in which the in-person portion of the experiment would be held.

**Experimental sessions.** An overview of the sequencing and timing of the experimental sessions is presented in Table 5. Each session took place over three consecutive days. At the beginning of Day 1, individuals completed an additional informed consent containing information about the lab portion of the experiment and their rights as research participants (Appendix B). Following the consent procedure, participants completed a computerized

Day	Activity	Time
	Informed consent	
	Working memory assessment	20 mins
	Introductory presentation	8 mins
Day 1	Familiarization trial	3 mins
2491	Practice trials (6)	8 mins (48 mins)
	Performance trial	10 mins
	Post-trial measurement	25 mins
	Familiarization trial	3 mins
Day 2	Practice trials (6)	8 mins (48 mins)
	Performance trial	10 mins
	Post-trial measurement	25 mins
	Familiarization trial	3 mins
Day 3	Practice trials (6)	8 mins (48 mins)
-	Performance trial	10 mins
	Post-trial measurement	25 mins

Table 5Summary of Experimental Session Sequence and Timings

Note. Days 1-3 occurred consecutively.

assessment of working memory; once all participants finished the assessment, a short, automated training presentation was displayed. A brief familiarization trial followed which enabled participants to familiarize themselves with the TANDEM interface and the experimenter to explain the rules and procedures participants should follow for completing the experiment. Next, participants completed six practice trials with TANDEM where they learned to engage targets using the radar interface; at set intervals between these trials, instructions containing both the

experimental manipulations and guided exploratory learning recommendations were presented to participants. Once participants had completed the practice rounds, a single performance trial was completed; scores from the performance trials were later used to determine the winner of the monetary awards. Lastly, individuals were asked to complete a series of post-trial measures. At the end of Days 1 and 2, participants were provided with a reminder slip that indicated the date, time, and location of their next session. At the end of Day 3, participants were debriefed and the manner by which the monetary awards would be distributed was described (Appendix C). In its entirety, the lab portion of the experiment took approximately 5 hours to complete. The full experimental protocol for the experimental sessions is provided in Appendix D.

*Task introduction and familiarization trial.* The automated introductory presentation shown to all participants at the beginning of Day 1 was projected on a large screen at the front of the computer lab. The narrated video lasted approximately eight minutes and described the purpose of TANDEM, the sequence of events in the study, procedural rules for the lab, and how to operate the task interface and manual. The experimental manipulation instructions were also presented for the first time near the start of the presentation.

Following the introductory training video, participants completed a short familiarization trial using the TANDEM task environment. During this period, individuals were permitted access to the online instruction manual for 30 seconds and then completed a one minute trial that enabled participants to practice using the manual, starting TANDEM, and selecting, viewing information, and inputting target decisions on the radar screen. Participants were informed that the purpose of the familiarization trial was simply to orient themselves with the computer equipment and how to perform these common task operations. No feedback was given regarding performance or activities during the familiarization trial.

*Practice trials.* Following the familiarization period, participants engaged in six practice trials each day, for a total of 18 trials across the course of the experimental session. Each practice trial followed a standard progression of two minutes for studying the task manual, five minutes for hands-on practice with the radar interface, and one minute to review post-trial feedback about performance during that trial (Appendix E). Scores for the practice trials were computed based on scoring algorithms made available to participants in the online manual; 100 points were earned for every target correctly identified and engaged (i.e., all three subdecisions and engagement decision correct) while 100 points were deducted for every misidentification/incorrect prosecution and every target that crossed into either the inner or outer defensive perimeters.

Each practice trial consisted of a set of targets which comprised the task scenario. Participants were presented with the same scenario for all trials on a single day, though a different scenario was used across days. Thus, participants saw the same set of targets in all six practice trials on Day 1, with different scenarios used for each of the practice trials on Days 2 and 3. Each practice trial consisted of 21 valid targets (plus six marker targets) distributed at various locations on the radar screen, fourteen of which were visible at the beginning of the trial and seven of which were pop-up targets. The pop-up targets appeared one at a time every 27 seconds from the start of the trial.

Although the specific cue values (cf., Table 3) for a given target were generated randomly, significant consideration was given to the distribution of target decision outcomes and the manner by which targets were programmed to behave. Specifically, the construction of each practice scenario was standardized to ensure that:

- A. The total number of targets representing each of the possible Type, Class, and Intent outcomes was approximately equal across all three days
- B. The total number of targets representing each of the final engagement decision outcomes as well as the unique combinations of Type, Class, and Intent outcomes (cf., Table 4) was approximately equal across all three days
- C. The same number of targets crossed the inner (4) and outer (7) perimeters in each scenario
- D. Five of the seven pop-up targets would cross a defensive perimeter if not prosecuted (3 crossed the outer, 2 crossed the inner) in each scenario
- E. The speed and location of targets were such that participants could realistically prosecute all "high priority" targets (i.e., those that would cross a defensive perimeter and cost points) in a single scenario

(A) and (B) in the list above were particularly important both for analytic purposes and to ensure that, on average, participants could practice interpreting and making decisions using all possible cue values during the practice scenarios. The middle column of Table 6 shows the distribution of subdecision and final engagement outcome characteristics across all targets in the three practice scenarios; for example, across all 63 viable targets constructed for the practice scenarios, 23 were designed to be Cleared, 21 to be Warned, and 19 to be Marked. Similarly equal distributions were achieved for the Type, Class, and Intent subdecisions as well. In sum, across all the practice trials, participants were provided with equal opportunity to practice identifying and classifying all possible target characteristics.

*Performance trial.* At the completion of each day's practice trials, participants engaged in a single performance trial; the performance trial was similar to the practice trials,

Target Characteristics	Practice Trials	Performance Trials
Final Engagement		
Clear	23	36
Warn	21	48
Mark	19	42
Туре		
Air	19	42
Surface	22	42
Sub	22	42
Class		
Civilian	32	72
Military	31	54
Intent		
Peaceful	33	60
Hostile	30	66

Table 6 Distribution of Target Characteristics Across all Scenarios for Practice Trial Targets (n = 63) and Performance Trial Targets (n = 126)

*Note.* Cell values represent total number of targets possessing a given characteristic across all practice and performance trials, respectively

though more difficult. Procedurally, the length of the trial was increased from five minutes to eight minutes and participants only received one minute to view the task manual prior to the scenario. Additionally, the following changes were introduced to increase the complexity and cognitive demands of the task (Bell, 2002): (1) the number of prosecutable targets was increased (from 21 to 42); (2) the number of pop-up targets in the scenario was increased (from 7 to 13); (3) scoring was changed such that more points were deducted for targets crossing the inner and outer perimeters (from 100 points to 150 points); (4) more targets were created which crossed a defensive perimeter (from 11 to 18); and (5) the number of pop-up targets that appeared close to a defensive perimeter were increased (from 5 to 8). Participants received instructions describing many of these critical differences before beginning the performance trial. Of relevance to the present study, these changes were designed so that even if high-capacity threatened individuals were capable of "brute forcing" their way through the learning trials by memorizing the correct cue decisions for contacts, they would still experience difficulties during the performance trial because they would not have developed parsimonious and efficiently organized knowledge structures or well-rehearsed cognitive task strategies that would facilitate performance in a more challenging context.

*Exploratory learning recommendations.* During each day's practice trials, participants were shown a set of exploratory learning recommendations at set intervals; Figure 5 depicts the sequencing of instructional delivery between practice trials for all three days of the experimental session. All instructional text was presented visually on-screen as well as audibly through headphones worn by each participant throughout the experiment. The duration for which the instructions were displayed was controlled by a timer that would automatically advance participants to the next stage of the experiment after it had elapsed.

The content for the exploratory learning recommendations was adapted from Bell (2002) and provided participants with reflective questions intended to stimulate learning and exploration of the procedures required to perform TANDEM. The recommendations focused on three key areas relevant to task completion: gathering and interpreting information, monitoring defensive perimeters, and prioritizing targets/maximizing score (full text of the recommendations are provided in Appendix F). During the initial presentation of the recommendations described the intent of the learning recommendations and suggestions for how to focus participants' learning efforts:

This page presents a list of questions that you can use to guide your learning activities within the task. You may find it useful to focus more of your early



*Figure 5*. Sequencing of exploratory learning recommendations and manipulation instructions during daily practice rounds

learning efforts in the radar control simulation on how to effectively and efficiently gather and interpret information in order to make accurate decisions. As you become more skilled at accurately processing targets, you may wish to shift your focus to learning how to monitor defensive perimeters and prioritize targets in order to maximize your score.

For all subsequent presentations of the learning guidelines, the audio text was slightly altered to

encourage participants to reflect on their performance and consider which areas may need greater

attention during their practice with the task:

Take a moment to assess how you are currently doing on the radar control task. If you are still having difficulty correctly processing targets, you may find it useful to focus more time on learning how to interpret information to make accurate decisions. If your accuracy is improving, you may wish to consider learning more about how to monitor your defensive perimeters and prioritize targets to maximize your score. The exploratory learning recommendations were always presented prior to the delivery of the experimental manipulation instructions preceding practice trials 1 and 5 each day and were displayed on-screen for 50 seconds before terminating.

*Experimental manipulation.* The delivery of the experimental manipulation text was similar to that of the exploratory learning recommendations. A timer-controlled page containing the manipulation instructions was presented to participants at specified intervals. Participants were asked to read the text on screen and/or listen along with the instructions as they were narrated to them through their headphones. To facilitate the pace of the experiment and maintain participant engagement, two versions of the stereotype threat and control condition manipulation instructions were created. The longer, full-length version of the instructions was presented at the start of each day prior to the first practice trial, while a shorter, reduced-length version of the instructions was on Day 1, in which the longer manipulation was presented during the introductory video and the shortened version presented prior to beginning practice trial 1). The instructions for both the stereotype threat and control conditions are presented in Appendix G and H, respectively.

The instructional text was based directly on experimental manipulations used in previous investigations of stereotype threat (e.g., Beilock et al., 2007; Rydell, Shiffrin, et al., 2010; Spencer et al., 1999). The stereotype threat manipulation text incorporated a number of features shown to enhance the cognitive imbalance elicited by the presence of negative group-ability stereotypes. First, participants in the stereotype threat condition were instructed that the purpose of the experiment was to examine possible explanations for why women tend to perform more poorly than men on math problems like those found on the SAT or ACT. Individuals were further informed that one reason for this finding was that women may have more difficulty

distinguishing relevant information needed to solve a problem from irrelevant/distracting information, and that TANDEM was designed to examine how these skills develop differently in men and women. The intent of this instructional element was to logically connect the behaviors exhibited in TANDEM to the sex-stereotyped domain of math; similar instructional paradigms have been successfully used to induce stereotype threat effects even in experimental tasks that are not directly indicative of traditional math performance/ability (Beilock et al., 2007; Rydell, Shiffrin et al., 2010).

Second, the diagnosticity/normative value of the task was emphasized to participants in the stereotype threat condition by reiterating that the task was capable of detecting differences in the above stereotyped skills. Research by Steele and colleagues (Steele, 1997; Steele & Aronson, 1995; Steele & Davies, 2003) has found that such information can increase the perceived risk of failure and likelihood of adhering to the negative stereotype by threatened individuals. Third, prior to the performance trials, participants in the stereotype threat condition were reminded that their goal was to score as many points as possible and that the top performers on these trials would be eligible for monetary rewards. Research has shown that similar performance approach perspectives are inconsistent with the failure avoidance mindset typically primed by negative stereotypes and can stimulate self-regulatory mismatches for participants in the stereotype condition that could contribute to poor learning/performance behaviors (e.g., Grimm et al., 2009; Jamieson & Harkins, 2007). Finally, all individuals were asked to input their sex into the computer after receiving the first on-screen presentation of the manipulation instructions and just prior to beginning the first practice trial each day. A number of researchers (Ambady et al., 2001; Danaher & Crandall, 2008; McGlone & Aronson, 2006; Shih et al., 1999; Shih et al., 2006; Steele and Aronson, 1995; Yopyk & Prentice, 2005) report that such group saliency reminders

can heighten awareness of one's group affiliation, thereby making it easier for threatened individuals to draw the negative group-ability propositional relation that triggers stereotype threat (but see Stricker and Ward, 2004).

Steele and Davies (2003) note it is imperative that cues which might otherwise elicit threat in the control/comparison condition be removed or minimized to the greatest extent possible in order to obtain an accurate test of threat effects. As such, participants in the control condition were informed that the purpose of the experiment was simply to examine individual differences in learning and problem-solving skills (e.g., Steele & Aronson, 1995; Rydell, Shiffrin et al., 2010) and made no mention of sex or sex differences. Furthermore, no indication of the diagnosticity of the task was provided. Lastly, control condition participants were informed that the performance trials should be viewed as opportunities to demonstrate their newly learned skills in a more challenging environment.

#### Measures

Over the course of the study, participants were asked to complete a number of individual difference measures either online (Appendix I) or in person (Appendix J). Additionally, several measures were computed from data directly recorded in the TANDEM program

**Demographics.** Participants provided basic demographic information about their age, sex, and proficiency with English in an online questionnaire prior to arriving to the lab. Additionally, participants were asked to indicate their handedness and overall familiarity/experience with playing video games.

**Cognitive ability.** Given its documented relation to learning and performance within the TANDEM task paradigm (Kozlowski, Gully, et al., 2001), a measure of cognitive ability was gathered as a potential control variable in the online questionnaire as well. Participants were

asked to report their highest score achieved on the SAT or ACT, which were verified through the university registrar. Research has demonstrated that scores on these tests typically possess a large *g* component and are generally internally consistent (e.g., Frey & Detterman, 2004; Koenig, Frey, & Detterman, 2008).

**Math domain identification.** Identification with a domain characterizes an individual's perceptions of the attractiveness, importance, and relevance to self of one's performance in a particular area of functioning (Steele, 1997). As noted previously, a number of researchers have suggested that the individuals most susceptible to stereotype threat effects are those who care about or are strongly invested in the domain/area where the stereotype threat applies (Crocker et al., 1998; Steele & Aronson, 1995; Steele & Davies, 2003). Although heightened domain identification may exacerbate the experience of stereotype threat, meta-analytic evidence suggests that stereotype threat effects can still emerge for threatened individuals who do not report being strongly identified with a given domain (Nguyen & Ryan, 2008).

Consequently, data were collected on participant's math domain identification as a potential control variable using Smith and White's (2001) nine-item Domain Identification Measure. The item content focused on respondents' enjoyment, interest, and performance in mathematics (e.g., "I have always done well in Math", "How much is Math to the sense of who you are?"); the internal consistency reliability of the scale was  $\alpha = .91$ .

**Working memory.** Working memory capacity was also assessed for use as a possible control variable using a modified version of the automated Operation Span (OSPAN) task (Unsworth, Heitz, Schrock, & Engle, 2005). The automated OSPAN task was administered inperson electronically to each participant using the E-Prime 2.0 software package (Psychology Software Tools, 2012; www.pstnet.com)<sup>5</sup>. OSPAN requires individuals to memorize and later

recall a series of stimuli (letters, words, etc.) while solving simple mathematics problems (Turner & Engle, 1989). Prior to administration of the OSPAN measure, participants proceeded through a guided training exercise that provided instructions on the operational procedures of the task and opportunities to familiarize themselves with memorizing/recalling letters, solving the math problems, and then memorizing/recalling letters while solving math problems just as they would during actual administration of the memory task.

Each set of items on the OSPAN began by presenting participants with a relatively simple mathematical operation (e.g., (9/3) + 2 = ?, (3\*2) - 1 = ?) which they were to solve in their heads and then click on-screen once they knew the answer. The amount of time participants were given to answer each question was based on the average response time needed to solve the math items during the practice trials; if participants took longer than their average response time plus 2.5 SD, the task would automatically advance and counted that trial as an error. Following the math operation, a single number was then shown near the top of a new screen along with two buttons labeled true and false which participants used to indicate whether that number was the correct answer to the math problem just viewed. After providing their answer, a random letter (e.g., "F") was then shown on-screen for 800ms. This procedure repeated until three to seven math operation-letter presentation pairings were completed. Once all the pairings in a sequence were administered, a screen containing 12 letters arranged in a 3x4 grid was displayed; participants were asked to recall the letters in the same order in which they were presented during the pairings by selecting the appropriate letters and then clicking a button to submit their answer. In total, respondents completed three memorization sets from each of the five sizes of operationletter pairings (i.e., three sets of three operation-letter pairings, three sets of four operation-letter pairings, etc.); thus, individuals saw a total of 75 letters and math problems. The specific math

operations, letters, and order in which the set sizes were presented were randomized across participants.

The recommendations of Conway et al. (2005) were followed to compute respondents' scores from the OSPAN task. First, participants were required to accurately answer 85% of the math problems they attempted and have no more than 15% speed errors across all items. Second, to compute the final OSPAN scores, the number of letters correctly recalled in the correct serial position was summed across all 15 item sets (i.e., a participant who saw the sequence of letters RSTLNE, but recalled RETSNL would receive a score of 3 for their memory span on the item set as the letters R, T, and N were recalled in the correct serial position). Although some researchers advocate computing OSPAN scores based on whether the entire sequence of letters is reported in the correct order (i.e., a participant who saw the sequence of letters RSTLNE, but recalled RETSNL would receive a score of 0 for their memory span on the item set as the letters S, L, and E were not reported in the correct sequence, cf., Turner & Engle, 1989), previous research shows that this approach can lead to poorer construct/criterion validity and reliability of the working memory span measures (Conway et al., 2005). Thus, performance on the OSPAN was calculated using the partial credit/serial position algorithm described above; scores on the measure could vary from 0 (no letters correctly recalled in the correct location in any item set) to 75 (all letters correctly recalled in the correct location on all item sets).

**Metacognitive activity.** As an alternative measure of cognitive strategy development (Kraiger et al., 1993), individual's self-reported metacognition was measured at the end of each day using a 12-item measure developed by Ford et al. (1998) and adapted by Bell (2002) to specifically fit within the context of the TANDEM task paradigm. The questions were administered through an online survey system during the post-trial measurement period. Each

question asked participants to indicate the extent to which they consciously reflected on their learning and performance activities during the task (e.g., "As I performed in the practice trials, I evaluated how well I was learning the skills of the simulation," "When my methods were not successful, I experimented with different procedures for performing the task"), with responses given on a 5-point scale (*1—Never, 2—Rarely, 3—Sometimes, 4—Frequently, 5—Constantly*). Coefficient alphas for the measure were  $\alpha = .86$ , .91, and .95 for the assessments on Days 1, 2, and 3, respectively; test-retest reliability between assessment periods was also reasonably strong (r = .68 between Days 1 and 2, r = .83 between Days 2 and 3, and r = .59 between Days 1 and 3).

Manipulation checks. Two separate scales were administered to participants through an online questionnaire administered at the end of Day 3 in an attempt to assess the efficacy of the stereotype threat manipulation and participants' belief that TANDEM was related to mathematical ability, respectively. As a check on the stereotype threat manipulation, a 7-item self-report measure of perceived stereotype threat adapted from Ployhart et al. (2003) was administered. The measure asked participants the extent to which they agreed with statements regarding negative perceptions/expectations about their gender's performance in the experimental task (e.g., "A negative opinion exists about how members of my gender should perform on this type of task."). Although Steele (1997) contends that individuals need not be consciously aware that they are under the influence of a negative stereotype to experience its effects, previous research has found that threatened individuals often do perceive this threat and that such self-report measures may be a useful means to assess the saliency of stereotype threat across experimental conditions (Grand et al., 2011). Lastly, a simple two-item measure was constructed to examine participant's belief about the feasibility of TANDEM's relatedness to mathematical ability (e.g., "The radar control task assesses skills related to mathematical

ability."). The coefficient alphas for the perceived stereotype threat and manipulation check measures were  $\alpha = .67$  and .74, respectively.

**Declarative knowledge.** An 11-item, multiple-choice test of declarative knowledge pertaining to TANDEM adapted from Bell and Kozlowski (2002) was completed by participants during the post-trial measurement period at the end of each day. The test questions focused exclusively on basic content concerning the interpretation of cue values ("If a target's characteristics are Speed = 35 knots and Altitude/Depth = 15 feet, which of the following actions should you take?"), and thus provided a measure of the extent to which individuals were able to learn and retrieve foundational knowledge about the task. In an attempt to minimize participant's reliance on memory of past administrations when answering the items (e.g., Lievens, Reeve, & Heggestad, 2007), a different set of test items was given to participants each day. The additional items for Days 2 and 3 were constructed by altering the item stems and responses from the Day 1. A single item from the Day 1 knowledge test was removed during analyses due to an error in the question response options, leaving only 10 items for this assessment period. The final Cronbach's alpha coefficients for the test versions administered on Day 1 ( $\alpha$  = .60), Day 2 ( $\alpha$  = .65), and Day 3 ( $\alpha$  = .69) were all moderate and typical for a dichotomously scored assessment.

**Knowledge structure assessment.** To assess the development of knowledge structures, participants were asked to provide proximity ratings indicating perceived similarity between 16 concepts identified as critical to performance in TANDEM at the end of each day (Table 7). Participants provided their ratings using an online survey system. A detailed set of instructions (Appendix K) was provided to participants prior to beginning the rating task that described the purpose of the assessment and recommendations about how the rating task should be completed (cf., Goldsmith et al., 1991). When providing ratings, participants were presented with two

Focus	Concept	Description
	1. Identify contact Type as Air	Classifying a target as an Aircraft
	2. Identify contact Type as Surface	Classifying a target as a Surface vessel
	3. Identify contact Type as Submarine	Classifying a target as a Submarine
	4. Identify contact Class as Civilian	Classifying target as a Civilian craft
	5. Identify contact Class as Military	Classifying target as a Military craft
Decision-	6. Identify contact Intent as Peaceful	Classifying target as a Peaceful craft
making	7. Identify contact Intent as Hostile	Classifying target as a Hostile craft
	8. Make decision to Clear contact	Making final engagement decision to Clear the target
	9. Make decision to Warn contact	Making final engagement decision to Warn the target
	10. Make decision to Mark contact	Making final engagement decision to Mark the target
	11. Gain/lose points	Objective indicator of task performance
	12. Zoom out/zoom in	Changing radar display resolution
	13. Monitor inner perimeter	Tracking potential boundary intrusions around the smaller visible defensive perimeter
Procedural/ Strategic	14. Monitor outer perimeter	Tracking potential boundary intrusions around the larger invisible defensive perimeter
	15. Find/engage pop-up targets	Identification/prosecution of new targets
	16. Prioritize targets (engage targets likely to cross perimeter first)	Identification/prosecution of high priority targets likely to cross a defensive perimeter

Table 7TANDEM Knowledge Concepts with Descriptions

concepts and asked to indicate how related they were to one another using a 9-point scale (1 not at all related to 9—highly related). A proximity rating was provided for every unique pairwise combination of concepts, resulting in 16\*(16-1)/2 = 120 ratings per knowledge structure<sup>6</sup>. To ensure that participants did not see the same ordering of concept pairs across days, concept pairs were presented twelve at a time in random order on each survey page and the order in which survey pages was presented was randomized.

The knowledge concepts rated by participants were adapted from those employed in previous research examining knowledge structures using the TANDEM task environment (Kozlowski, Gully, et al., 2001; Kraiger et al., 1995). Unlike previous investigations, however, an explicit focus of the present study was directed towards investigating differences in the manner by which individuals learned and formed mental representations of concepts/information relevant to decision-making in the task, as opposed to the acquisition of broader operational gameplay procedures (e.g., hooking contacts, gathering information, monitoring feedback, etc.). Consequently, the concepts shown in Table 7 were purposefully constructed to represent two distinct foci; those that dealt with outcomes related to decision-making in the task and those that concerned more procedural/strategic aspects.

An advantage of this stimulus set is that it permits a number of options for visually and statistically interpreting the structural relations reported by individuals. At the coarsest level, the extent to which distinctive decision-making and procedural/strategic concept clusters emerge in individual's knowledge structures provides insight into participants' ability to distinguish these two aspects of task performance. However, more subtle distinctions can also be examined by considering the specific pattern of relations that emerge among the decision-making concepts. As can be inferred from Table 3 presented earlier, similarity among decision-making concepts could be based on the degree to which concepts correspond to the same subdecision/decision outcome (e.g., Concepts 1-3 in Table 7 all refer to making the Type subdecision, while Concepts 4 and 5 correspond with the Class subdecision). A knowledge structure demonstrating this pattern of

clustering would be indicative of learning based on *feature* in which decision-relevant information is related on the basis of decision class.

Alternatively, a relational pattern among decision-making concepts based on the extent to which a given subdecision outcome (e.g., Air, Civilian, etc.) is indicative of a particular final engagement decision outcome (e.g., Clear, Warn, Mark) would be indicative of learning based on *functional* similarity. To better clarify this structural pattern, consider the distribution of each Type, Class, and Intent outcome across the three possible final engagement decision outcomes (e.g., Table 4). Based on this distribution, the probability with which any given subdecision outcome is associated with a particular final engagement outcome can be calculated (Table 8). Functionally, these probabilities indicate the likelihood that one would Clear, Warn, or Mark a target given that the target has a particular Type, Class, or Intent. For example, if a target is identified as Civilian, there is a 67% chance that the correct final engagement decision for that target is Warn, regardless of its Type or Intent; similarly if a target is identified as an Aircraft, there is only a 25% chance the correct final engagement decision would be Clear. The relative values of these probabilistic relations therefore represent the degree to which a subdecision outcome is functionally informative of the correct final engagement decision outcome. Consequently, participants who come to learn the rules of engagement effectively might generate knowledge structures which demonstrate a more functional organization of decision-critical information such that each Type, Class, and Intent outcome is more strongly associated with (i.e., seen as more similar to) its most probable final engagement decision as opposed to concepts with which it shares similar features.

In addition to these qualitative descriptors, the Pathfinder algorithm and software (Interlink, 2012; http://www.interlinkinc.net/) was used to compute the four quantitative metrics

Subdesision	Subdecision	<b>Final</b> ]	Engagement Out	tcomes
Subdecision	Outcomes	Clear	Warn	Mark
	Air	.25	.5	.25
Туре	Surface	.5	.25	.25
	Sub	.25	.25	.5
Class	Civilian	.17	.67	.17
Class	Military	.5	0	.5
T., 4 4	Peaceful	.67	.33	0
Intent	Hostile	0	.33	.67

Table 8Relative Probabilities Shared between Subdecision Outcomes and FinalEngagement Decision Outcomes

*Note.* Values in a single row should add to 1 (within rounding error). The value within each cell can be interpreted as the probability with which a single subdecision outcome is associated with a final engagement decision outcome based on the rules of engagement established for participants (Table 4).

of network structural quality noted earlier; two indices (similarity and correlation) provide information about the relatedness among different network structures, while the remaining two indices (coherence and number of links) provide descriptive information about a single network's composition. With respect to the relatedness metrics, similarity is a measure of the correspondence in links between networks; it is formally computed as the number of links held in common between any two networks divided by the total number of unique links in the structures (Goldsmith & Davenport, 1990). Alternatively, the correlation index measures the degree to which the concept-pair ratings in two networks/proximity matrices covary (Schuelke et al., 2009). These operationalizations have led some researchers to characterize similarity as a measure of agreement and correlation as a measure of consistency between two networks (e.g., Webber et al., 2000). Both structural similarity and correlation range from 0 to 1, with higher values indicating that the two contrasted structures share more links in common (similarity) or similar strengths of relations among concepts (correlation).

With respect to the descriptive indices, measures of structural coherence are based on the assumption that relatedness between a pair of concepts can be predicted by the relations of those concepts to other concepts in the network (Interlink, 2012). Specifically, coherence is calculated by first correlating, for each pair of concepts, the proximities between those concepts and all others in the network/proximity matrix. This "indirect" measure of relatedness is then correlated with the original proximity data to produce the coherence measure. Coherence values range from 0 to 1, with higher values indicating that the raw proximity/relatedness data is consistent with the indirect relatedness data inferred from the proximity matrix. The final measure, number of links in a network, is simply a count of the number of links retained in knowledge structure following application of the Pathfinder network algorithm.

A meaningful referent network was required in order to compute both the similarity and correlation indices described above. As described in Hypotheses 1, 2, 6, and 7, the referent structures of interest in the present study were those of males and top scoring performers on the TANDEM task. To compose these referent networks, the proximity ratings provided by males at each day were averaged together to form three separate proximity matrices and, subsequently, three separate networks representing the average male knowledge structure at the end of each day. A similar process was followed using the proximity ratings provided by the top 15 highest-scoring participants across all three performance trials<sup>7</sup>. The observed similarity and correlation indices at Day 1 were then computed by comparing the Day 1 knowledge structure of each participant to the averaged Day 1 knowledge structures of males and top performers separately; likewise, the observed indices for Days 2 and 3 were computed by comparing participants' Day 2

structure to the averaged males/top performers' Day 2 structures and participants' Day 3 structure to the averaged males/top performers' Day 3 structures, respectively. Consequently, the relatedness metrics computed in the present study reflect the extent to which participant's knowledge structures were more or less similar to/correlated with those of males and top performers at the same point in time.

Lastly, network structure metrics based on traditional graph theory analytic techniques were also computed for exploratory analyses (see Watts, 1999). These measures included the average shortest path lengths between all pairs of nodes (L), network diameter (D); the single longest path length across all nodes), and the clustering coefficient (C; the probability that two neighbors of a randomly chosen node will themselves be neighbors, indicative of a network's "clumpiness," Watts & Strogatz, 1998). Two additional indices were also computed in an attempt to quantify the extent to which structures exhibited feature similarity and functional similarity as described previously. For the former, the shortest path lengths between all concepts within a given subdecision were first computed (e.g., L computed separately for only Type concepts, only Class concepts, and only Intent concepts) and then averaged together to provide a measure of the average shortest path lengths among feature concepts ( $L_{\text{feature}}$ ). For the latter, the shortest path lengths between each subdecision outcome and its most strongly associated final engagement decision were computed (e.g., L between Civilian and Warn, L between Peaceful and Clear, etc.) and then averaged together to provide a measure of the average shortest path lengths among function concepts ( $L_{\text{function}}$ ). All graph theoretic computations were performed in MATLAB version 7.14 (MathWorks, 2012) using the MatlabBGL library (http://dgleich.github.com/matlab-bgl/).

**Strategic learning behaviors.** Task strategy was operationalized using two sets of variables. First, the amount of time participants spent looking at the manual pages containing basic gameplay, cue value, and task strategy information was recorded for every trial. The amount of time spent reading materials on each of these task manual sections provides an indication of how participants structured their learning efforts during information acquisition phases and thus the approach they took to learning the task space. In general, individuals should be expected to spend less time on the basic gameplay manual pages and more time on the cue value and task strategy pages as more experience is gained with the task across trials.

In addition to information acquisition, the manner by which individuals interacted with the radar interface and performed certain operations in TANDEM was examined as an indicator of task comprehension/strategic learning. As participants gain experience and learn how various task mechanisms interact with the rules of engagement (i.e., how points are gained/lost) to dictate performance, they should begin to formulate implicit heuristics which guide their selection of targets and gameplay behaviors in TANDEM. The online task manual presented to participants contains information which describes the most effective strategy for prioritizing targets in order to maximize task performance; this strategy is depicted in graphical form in Figure 6. In short, this strategic approach suggests that individuals start by first locating-but not engaging—the marker targets which outline the invisible outer perimeter. Once this critical boundary is identified, individuals can begin to prioritize target engagement by attempting to clear those targets most likely to breach this outer perimeter or the visible inner perimeter. While working to clear these targets, individuals should continually be on the lookout for pop-up and other high priority targets which threaten to cross either perimeter. Consequently, a number of behavioral indicators were measured which reflected participants adherence to these advanced



Figure 6. Cognitive strategy heuristic for TANDEM performance

task strategies, including the number of marker targets engaged (fewer engaged is indicative of better strategic task performance), number of times an individual zoomed the radar screen in/out (more is indicative of better strategic task performance), and the number of high priority targets processed (i.e., pop-up targets and targets which would cross a defensive perimeter if not engaged; more engaged is indicative of better strategic task performance).

**Decision-making strategy.** At its core, TANDEM is a multiple-cue decision-making task. Participants are provided with a number of different pieces of information (e.g., the identifying cues and cue values shown in Table 3) that must be interpreted and integrated in order to make a series of decisions about the identity of and appropriate course of action against a target (cf., Table 4). Consequently, each cue viewed by a participant contributes some unique informational value to those decisions; sampled across multiple targets and combinations of cue values, these decision "weights" can be reconstructed empirically and used to draw inferences about the manner by which individuals combine information in order to prosecute targets in TANDEM. Furthermore, given that any specific cue value is directly associated with a known and veridical decision outcome (e.g., a Speed value of 115 knots is indicative of a specified Type outcome, etc.), these decision weights can be interpreted as the extent to which participants have correctly learned to distinguish, evaluate, and integrate task-relevant information in order to make accurate task decisions.

Two pieces of data were extracted from the TANDEM game files in order to evaluate participants' decision-making heuristics. First, the content and associated meaning of the informational cues examined by participants for each processed target was collected; thus, data on which cues were viewed and what information those cues conveyed were gathered for every target processed by each participant. Lastly, the specific outcomes selected for each of the Type, Class, Intent, and Final Engagement decisions for every target processed were recorded.

**Task performance.** Performance scores in TANDEM are a combination of both effective procedural decision-making and strategic target selection. Consequently, task

performance was broken into these unique components and analyzed separately to provide a more accurate picture of how participants were performing on the task. The number of targets engaged correctly and incorrectly and the number of targets that crossed the inner and outer defensive perimeters were gathered for each trial in the game; additionally, performance scores based on the algorithms denoted previously were computed as an overall indicator of task effectiveness.

#### **RESULTS**

### **Descriptive Statistics and Data Cleaning**

Means, standard deviations, and intercorrelations for all study variables are presented in Table 9. Prior to performing all analyses, the integrity of participants' OSPAN and knowledge structure data was evaluated to improve the quality of the dataset. As described previously, participants' performance on the mathematics portion of the OSPAN was assessed to screen participants who may not have been adequately attending to the processing component of the task and/or were using that time to rehearse the letters rather than solve the math problems (Turner & Engle, 1989; Unsworth et al., 2005). Seven and five participants failed to reach the 85% accuracy and speed error criteria, respectively; additionally, a computer recording error resulted in complete loss of OSPAN data for one participant. Consequently, OSPAN scores for these 13 participants (6.6% of the total sample) were not included in the dataset.

The number of links produced in participants' knowledge structures was also examined for each day to identify individuals who were likely not attending to the proximity rating task seriously. Specifically, any network containing 120 links was not included in subsequent knowledge structure analyses—a network with 120 links reflects a structure in which each concept is linked to all other concepts and, by extension, a participant who provided the same numeric rating for all 120 pairwise concept comparisons during the rating task. This procedure resulted in the removal of seven participants (4.4% of the total 3-day sample), six of whom were women (four from the stereotype threat condition and two from the control condition). A computer error resulted in the loss of knowledge structure data at Day 2 for 11 participants in the stereotype threat condition (8 females, 3 males), though data from Days 1 and 3 for these participants were used in subsequent analyses.

Variable	М	SD	1	2	3	4	5	6	7	8	9
1. Sex <sup>a</sup>	.27	.44									
2. Condition <sup>b</sup>	.49	.50	.02								
3. ACT	23.55	3.63	.07	.14							
4. OSPAN	56.59	12.54	.19	08	.39						
5. Video game experience	2.47	1.32	.42	.03	.11	.16					
6. Math domain identification	2.99	.87	.21	12	.21	.15	.18				
7. Perceived stereotype threat	2.92	.61	46	.27	04	12	14	03			
8. Metacognitive activity (T1)	3.82	.57	.19	12	.28	.23	.27	.17	05		
9. Metacognitive activity (T2)	3.87	.68	.13	20	.12	.11	.29	.18	.03	.68	
10. Metacognitive activity (T3)	3.86	.79	.11	24	.12	.10	.20	.13	.00	.60	.83
11. Knowledge test (T1)	.70	.21	.05	.03	.31	.25	.14	.13	02	.33	.25
12. Knowledge test (T2)	.78	.18	.12	17	.19	.23	.10	.11	01	.27	.34
13. Knowledge test (T3)	.76	.20	.11	22	.21	.22	.16	.21	.03	.33	.40
14. Total Points (P1)	-2186	835	.28	02	.43	.18	.25	.08	18	.32	.26
15. Total Points (P2)	-1411	1283	.24	03	.48	.32	.24	.25	09	.27	.31
16. Total Points (P3)	-780	1468	.26	13	.49	.34	.31	.25	12	.31	.31
17. Number targets correct (P1)	6.56	3.99	.21	.03	.42	.22	.26	.07	10	.40	.28
18. Number targets correct (P2)	9.93	6.04	.19	04	.46	.31	.22	.21	08	.25	.29
19. Number targets correct (P3)	13.00	7.03	.22	12	.47	.30	.31	.22	10	.29	.30
20. Number targets incorrect (P1)	10.93	5.03	24	.01	28	09	16	08	.11	11	14
21. Number targets incorrect (P2)	8.05	6.20	22	.06	37	28	17	23	.05	21	25
22. Number targets incorrect (P3)	6.35	6.07	23	.19	35	34	24	22	.12	25	27
23. Total perimeter intrusions (P1)	17.48	2.46	11	.10	21	07	10	01	.19	19	16
24. Total perimeter intrusions (P2)	15.99	3.26	16	05	32	13	21	14	.09	22	21
25. Total perimeter intrusions (P3)	14.45	4.20	21	03	45	22	23	19	.07	23	22
26. Avg basic manual time (L1)	21.86	18.77	09	12	34	20	17	.04	06	35	31
27. Avg basic manual time (L2)	18.62	15.79	06	.10	16	.02	.00	14	.00	04	22
28. Avg basic manual time (L3)	21.43	17.97	12	04	06	03	.00	02	.09	01	03

Table 9Means, Standard Deviations and Interrcorrelations for Study Variables

Variable	M	SD	1	2	3	4	5	6	7	8	9
$\frac{1}{29} \text{ Avg cue manual time (I 1)}$	84.47	14 45	- 02	04	_ 10	01	06	- 01	, 00	12	06
30 Avg cue manual time (L2)	81 15	23.01	02	.04	19	.01 - 18	- 12	- 03	10	- 10	.00
31 Avg cue manual time $(L3)$	74 62	34 55	- 03	31	24	- 13	- 06	- 03	- 04	- 12	- 08
32 Avg strategy manual time (L1)	41.33	20.24	- 04	04	.24	11	.00	08	.01	.23	.00
33 Avg strategy manual time (L2)	51.03	24 32	- 02	- 06	02	07	03	11	02	05	02
34. Avg strategy manual time (L3)	47.38	29.73	18	08	.01	13	06	06	.11	04	04
35. Coherence (T1)	.29	.27	.00	02	.33	.13	.17	.09	06	.24	.15
36. Coherence (T2)	.34	.31	04	08	.23	.21	.13	.20	02	.21	.17
37. Coherence (T3)	.32	.38	17	02	.17	.10	.04	.05	.10	.20	.20
38. Number of links (T1)	29.38	13.46	.01	.03	03	01	01	.02	01	05	.01
39. Number of links (T2)	31.66	17.47	01	.05	.07	.12	06	.01	09	.09	.00
40. Number of links (T3)	35.05	19.61	03	09	.11	.16	11	.05	09	.13	.10
41. Correlation $(T1)^{c}$	.38	.24	.02	02	.51	.23	.22	.16	03	.27	.20
42. Correlation (T2) <sup>c</sup>	.45	.27	.04	09	.44	.24	.19	.26	05	.21	.23
43. Correlation (T3) <sup>c</sup>	.46	.29	.00	05	.38	.25	.16	.17	.08	.20	.18
44. Similarity (T1) <sup>c</sup>	.16	.07	.10	01	.39	.34	.17	.09	02	.25	.15
45. Similarity (T2) <sup>c</sup>	.20	.10	.16	09	.26	.20	.27	.19	04	.18	.20
46. Similarity $(T3)^{c}$	.22	.14	.21	01	.26	.20	.36	.17	01	.25	.21
47. Clustering coefficient (T1)	.15	.09	.05	03	.09	.09	.06	.10	06	.14	.17
48. Clustering coefficient (T2)	.17	.10	.05	.01	.23	.26	.14	.10	06	.28	.17
49. Clustering coefficient (T3)	.19	.12	.04	.00	.33	.18	.03	.10	10	.24	.15
50. Diameter (T1)	4.77	1.37	05	03	.01	.04	01	.01	08	11	12
51. Diameter (T2)	4.50	1.54	16	.02	02	12	15	05	.11	17	05
52. Diameter (T3)	4.16	1.62	12	02	19	16	02	09	.05	16	07
53. Avg. path length (T1)	2.32	.46	03	.01	.05	.05	03	01	08	08	11
54. Avg. path length (T2)	2.27	.53	11	.03	02	12	11	.00	.11	17	02
55. Avg. path length (T3)	2.12	.51	05	.03	13	17	.03	01	.07	18	08

Variable	М	SD	1	2	3	4	5	6	7	8	9
56. Avg. feature path length (T1)	2.14	.78	.05	06	09	07	.04	03	.04	06	.00
57. Avg. feature path length (T2)	2.07	.75	03	.06	03	02	07	05	.05	08	07
58. Avg. feature path length (T3)	1.95	.63	.09	.08	13	08	.04	.01	.05	07	10
59. Avg. function path length (T1)	2.09	.48	.03	.08	07	10	.03	.02	.03	10	08
60. Avg. function path length (T2)	2.01	.64	05	.17	04	16	11	08	.13	13	08
61. Avg. function path length (T3)	1.89	.60	.11	.02	07	10	.07	.12	06	15	11

Correlations in bold are significant at p < .05

<sup>a</sup>Dummy-coded variable (Female = 0, Male = 1)

<sup>b</sup>Dummy-coded variable (Control = 0, Stereotype threat = 1)

<sup>c</sup>Referent knowledge structure for computations was the averaged knowledge structure of the top 15 highest-scoring performers at the same point in time

*Note.* The code in parentheses following each variable name indicates when measurement was taken. Specifically, T1-T3 refer to the end of Days 1-3; P1-P3 refer to performance trials on Days 1-3; and L1-L3 refer to learning trials on Days 1-3.

Table 9 (cont'd)

Variable	10	11	12	13	14	15	16	17	18	19	20
1. Sex											
2. Condition											
3. ACT											
4. OSPAN											
5. Video game experience											
6. Math domain identification											
7. Perceived stereotype threat											
8. Metacognitive activity (T1)											
9. Metacognitive activity (T2)											
10. Metacognitive activity (T3)											
11. Knowledge test (T1)	.19										
12. Knowledge test (T2)	.23	.47									
13. Knowledge test (T3)	.38	.46	.64								
14. Total Points (P1)	.19	.45	.38	.40							
15. Total Points (P2)	.21	.51	.58	.49	.67	_					
16. Total Points (P3)	.30	.47	.62	.59	.60	.82	—				
17. Number targets correct (P1)	.20	.50	.33	.35	<b>.8</b> 7	.59	.53	—			
18. Number targets correct (P2)	.18	.50	.57	.44	.66	.95	.79	.62	—		
19. Number targets correct (P3)	.28	.47	.58	.56	.58	.78	.94	.55	.82		
20. Number targets incorrect (P1)	08	30	29	34	83	55	47	59	49	43	—
21. Number targets incorrect (P2)	17	45	55	50	57	88	71	43	78	63	.56
22. Number targets incorrect (P3)	29	38	57	55	49	69	87	38	65	78	.44
23. Total perimeter intrusions (P1)	14	11	17	13	26	21	24	16	22	21	17
24. Total perimeter intrusions (P2)	16	25	19	19	36	51	43	37	40	38	.18
25. Total perimeter intrusions (P3)	16	30	37	31	43	56	65	37	46	50	.28
26. Avg basic manual time (L1)	27	22	24	11	23	22	25	29	22	23	.02
27. Avg basic manual time (L2)	21	29	37	30	30	36	35	25	29	32	.21
28. Avg basic manual time (L3)	01	.03	05	.01	07	.07	06	04	.07	06	.08

Variable	10	11	12	13	14	15	16	17	18	19	20
29. Avg cue manual time (L1)	.07	.15	.04	06	.08	04	06	.21	.03	02	.07
30. Avg cue manual time (L2)	02	23	08	14	33	21	20	31	18	18	.31
31. Avg cue manual time (L3)	05	28	23	18	24	38	32	20	35	31	.15
32. Avg strategy manual time (L1)	.11	.05	.12	.20	.01	.16	.15	03	.12	.15	.00
33. Avg strategy manual time (L2)	.06	.15	.07	.19	.24	.13	.05	.26	.09	.00	21
34. Avg strategy manual time (L3)	07	.02	.12	04	01	04	03	01	.00	03	12
35. Coherence (T1)	.14	.11	.18	.29	.15	.30	.34	.12	.29	.35	10
36. Coherence (T2)	.11	.17	.30	.17	.20	.34	.31	.19	.35	.32	17
37. Coherence (T3)	.14	.24	.27	.30	.12	.29	.29	.16	.32	.34	04
38. Number of links (T1)	.02	.03	.07	.02	.06	.05	.06	.05	.03	.00	01
39. Number of links (T2)	03	.04	.04	.01	.07	.02	01	.12	.04	02	.02
40. Number of links (T3)	.14	.12	.03	.02	.07	.00	.05	.09	.01	.03	.07
41. Correlation (T1)	.16	.31	.26	.31	.29	.43	.47	.29	.44	.50	15
42. Correlation (T2)	.12	.28	.33	.32	.30	.41	.49	.28	.42	.50	21
43. Correlation (T3)	.11	.33	.46	.41	.32	.47	.48	.30	.49	.49	19
44. Similarity (T1)	.12	.33	.25	.30	.29	.40	.39	.30	.38	.39	17
45. Similarity (T2)	.06	.22	.29	.29	.25	.35	.45	.25	.33	.43	21
46. Similarity (T3)	.13	.22	.35	.31	.33	.41	.42	.28	.40	.41	28
47. Clustering coefficient (T1)	.18	.14	.16	.27	.15	.24	.25	.15	.17	.20	05
48. Clustering coefficient (T2)	.07	.11	.15	.15	.18	.23	.22	.21	.22	.20	08
49. Clustering coefficient (T3)	.14	.20	.13	.15	.17	.22	.23	.16	.23	.20	02
50. Diameter (T1)	16	12	09	04	13	12	16	15	11	10	.02
51. Diameter (T2)	.05	03	03	.00	22	15	09	25	13	08	.14
52. Diameter (T3)	06	09	08	04	15	10	06	14	05	.03	.07
53. Avg. path length (T1)	16	08	07	03	06	06	11	09	04	05	03
54. Avg. path length (T2)	.08	02	.00	.04	18	08	01	21	06	.00	.09
55. Avg. path length (T3)	10	07	01	.00	12	02	01	12	.02	.06	.01

Variable	10	11	12	13	14	15	16	17	18	19	20
56. Avg. feature path length (T1)	04	05	13	09	03	06	08	06	05	07	02
57. Avg. feature path length (T2)	04	05	11	11	12	17	13	12	16	16	.07
58. Avg. feature path length (T3)	06	12	13	12	08	13	09	12	16	09	04
59. Avg. function path length (T1)	12	23	19	19	15	16	21	17	16	17	.05
60. Avg. function path length (T2)	03	10	15	16	16	15	15	17	15	13	.07
61. Avg. function path length (T3)	09	02	.02	.01	03	.01	03	04	01	01	01

Table 9 (cont'd)

Variable	21	22	23	24	25	26	27	28	29	30	31
1. Sex											
2. Condition											
3. ACT											
4. OSPAN											
5. Video game experience											
6. Math domain identification											
7. Perceived stereotype threat											
8. Metacognitive activity (T1)											
9. Metacognitive activity (T2)											
10. Metacognitive activity (T3)											
11. Knowledge test (T1)											
12. Knowledge test (T2)											
13. Knowledge test (T3)											
14. Total Points (P1)											
15. Total Points (P2)											
16. Total Points (P3)											
17. Number targets correct (P1)											
18. Number targets correct (P2)											
19. Number targets correct (P3)											
20. Number targets incorrect (P1)											
21. Number targets incorrect (P2)											
22. Number targets incorrect (P3)	.72										
23. Total perimeter intrusions (P1)	.09	.16									
24. Total perimeter intrusions (P2)	.13	.17	.26								
25. Total perimeter intrusions (P3)	.40	.31	.26	.62							
26. Avg basic manual time (L1)	.12	.13	.27	.22	.33						
27. Avg basic manual time (L2)	.38	.36	.16	.13	.15	.04					
28. Avg basic manual time (L3)	10	.05	.00	.03	.05	.07	.16				

Variable	21	22	23	24	25	26	27	28	29	30	31
29. Avg cue manual time (L1)	.05	02	07	.12	.20	16	06	.10			
30. Avg cue manual time (L2)	.18	.15	03	.14	.19	.16	13	05	.14		
31. Avg cue manual time (L3)	.36	.18	.18	.18	.33	.13	.12	25	.02	.29	
32. Avg strategy manual time (L1)	11	04	08	20	22	10	.02	09	38	04	21
33. Avg strategy manual time (L2)	13	03	.03	11	11	.06	05	.04	.06	32	03
34. Avg strategy manual time (L3)	02	.00	.23	.17	.06	08	06	.15	10	13	.05
35. Coherence (T1)	24	22	13	18	28	11	.04	15	13	15	14
36. Coherence (T2)	32	25	04	08	19	09	14	12	.04	10	07
37. Coherence (T3)	27	23	06	05	12	06	15	10	.06	.11	07
38. Number of links (T1)	07	07	11	02	09	08	.04	.21	.08	.09	.07
39. Number of links (T2)	.00	.02	08	01	02	02	.03	.13	.11	.13	.17
40. Number of links (T3)	.02	05	22	03	06	02	09	04	.05	.16	.11
41. Correlation (T1)	33	33	20	24	35	19	13	13	08	18	22
42. Correlation (T2)	35	38	16	17	33	18	16	11	03	16	24
43. Correlation (T3)	40	38	22	20	31	19	23	03	03	16	24
44. Similarity (T1)	33	28	13	23	31	15	16	04	08	21	28
45. Similarity (T2)	28	34	03	23	34	18	09	01	05	21	26
46. Similarity (T3)	34	33	08	23	29	15	08	.16	05	33	33
47. Clustering coefficient (T1)	23	21	16	19	23	15	05	.07	.02	01	03
48. Clustering coefficient (T2)	18	17	11	16	19	12	05	.07	.12	01	.04
49. Clustering coefficient (T3)	18	18	25	13	20	12	12	01	.02	.02	08
50. Diameter (T1)	.11	.13	.14	.07	.18	.09	08	24	13	.03	.12
51. Diameter (T2)	.07	.05	.07	.18	.10	.06	03	18	11	.14	08
52. Diameter (T3)	.09	.08	.14	.13	.13	.04	.12	05	.05	.09	.15
53. Avg. path length (T1)	.06	.12	.13	.02	.13	.07	07	28	09	.00	.05
54. Avg. path length (T2)	.02	02	.08	.14	.06	.03	07	17	10	.09	12
55. Avg. path length (T3)	.01	.04	.18	.09	.07	.03	.07	02	03	.04	.06
Variable	21	22	23	24	25	26	27	28	29	30	31
------------------------------------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
56. Avg. feature path length (T1)	.11	.05	.05	04	.10	.10	15	13	06	.06	.07
57. Avg. feature path length (T2)	.14	.06	.07	.09	.11	.00	05	12	08	.04	10
58. Avg. feature path length (T3)	.07	.04	.15	.08	.13	.02	.04	11	.05	03	05
59. Avg. function path length (T1)	.14	.25	.12	.05	.08	.09	03	13	06	.04	.07
60. Avg. function path length (T2)	.10	.15	.14	.11	.08	.04	.03	06	11	.05	12
61. Avg. function path length (T3)	01	.05	.07	06	.02	.04	01	.05	.00	03	.04

Table 9 (cont'd)

												_
Variable	32	33	34	35	36	37	38	39	40	41	42	
29. Avg cue manual time (L1)												-
30. Avg cue manual time (L2)												
31. Avg cue manual time (L3)												
32. Avg strategy manual time (L1)												
33. Avg strategy manual time (L2)	.08											
34. Avg strategy manual time (L3)	.01	.34										
35. Coherence (T1)	.28	06	.08									
36. Coherence (T2)	.16	05	.18	.59	—							
37. Coherence (T3)	.14	20	.02	.54	.75							
38. Number of links (T1)	10	.00	.03	08	.04	.02						
39. Number of links (T2)	11	11	15	.12	.17	.17	.28					
40. Number of links (T3)	13	11	11	.15	.21	.22	.46	.51	—			
41. Correlation (T1)	.27	.02	.06	.72	.59	.50	08	.06	.12			
42. Correlation (T2)	.22	03	.07	.56	.70	.56	.01	.02	.07	.78		
43. Correlation (T3)	.16	.00	.03	.50	.63	.58	.02	.06	.07	.75	.89	
44. Similarity (T1)	.20	.11	.08	.43	.40	.33	12	03	.04	.64	.50	
45. Similarity (T2)	.12	.00	.01	.27	.25	.21	12	20	26	.44	.57	
46. Similarity (T3)	.14	.14	.04	.23	.24	.13	11	18	29	.48	.57	
47. Clustering coefficient (T1)	.09	.03	.01	.21	.16	.18	.72	.24	.31	.18	.21	
48. Clustering coefficient (T2)	.01	08	06	.35	.39	.32	.15	.74	.35	.32	.33	
49. Clustering coefficient (T3)	03	17	03	.38	.38	.39	.29	.43	.71	.37	.38	
50. Diameter (T1)	.06	05	.00	.16	.16	.10	58	17	15	.06	.06	
51. Diameter (T2)	.09	11	.12	05	04	.08	.00	58	10	10	10	
52. Diameter (T3)	.04	07	.09	.00	.04	.11	24	18	54	12	07	
53. Avg. path length (T1)	.06	06	02	.21	.20	.14	70	15	21	.10	.10	
54. Avg. path length (T2)	.10	05	.19	03	.01	.10	07	71	15	04	02	
55. Avg. path length (T3)	.09	.01	.13	01	.03	.08	33	30	71	07	01	

Variable	32	33	34	35	36	37	38	39	40	41	42
56. Avg. feature path length (T1)	09	.09	.06	17	18	23	36	07	10	14	16
57. Avg. feature path length (T2)	.09	.11	.27	30	34	38	19	40	20	23	28
58. Avg. feature path length (T3)	03	01	.20	27	28	29	15	23	38	24	23
59. Avg. function path length (T1)	.02	05	01	.00	02	04	50	15	21	12	13
60. Avg. function path length (T2)	.01	01	05	19	18	12	10	50	16	14	18
61. Avg. function path length (T3)	.04	.09	11	13	17	20	15	30	55	10	10

Table 9 (cont'd)

Variable	43	44	45	46	47	48	49	50	51	52	53
29. Avg cue manual time (L1)											
30. Avg cue manual time (L2)											
31. Avg cue manual time (L3)											
32. Avg strategy manual time (L1)											
33. Avg strategy manual time (L2)											
34. Avg strategy manual time (L3)											
35. Coherence (T1)											
36. Coherence (T2)											
37. Coherence (T3)											
38. Number of links (T1)											
39. Number of links (T2)											
40. Number of links (T3)											
41. Correlation (T1)											
42. Correlation (T2)											
43. Correlation (T3)											
44. Similarity (T1)	.49	—									
45. Similarity (T2)	.50	.27	—								
46. Similarity (T3)	.63	.30	.77	—							
47. Clustering coefficient (T1)	.18	.20	12	.09							
48. Clustering coefficient (T2)	.34	.23	.14	.18	.37	—					
49. Clustering coefficient (T3)	.38	.28	.09	.11	.36	.53					
50. Diameter (T1)	.05	.16	03	10	56	14	09				
51. Diameter (T2)	08	08	20	26	19	59	20	.22	—		
52. Diameter (T3)	13	09	14	24	21	23	53	.19	.33		
53. Avg. path length (T1)	.06	.18	.01	05	60	07	11	.92	.17	.26	
54. Avg. path length (T2)	.00	04	11	14	19	66	22	.24	.94	.32	.19
55. Avg. path length (T3)	06	05	.02	06	26	25	60	.21	.30	.93	.29

Variable	43	44	45	46	47	48	49	50	51	52	53
56. Avg. feature path length (T1)	14	05	.01	02	47	07	21	.40	03	06	.39
57. Avg. feature path length (T2)	31	16	01	10	33	47	38	.04	.31	03	.02
58. Avg. feature path length (T3)	24	20	.00	03	18	29	50	01	.04	.22	05
59. Avg. function path length (T1)	10	18	09	03	49	13	15	.52	.12	.14	.56
60. Avg. function path length (T2)	13	18	24	12	19	51	22	.03	.61	.17	.03
61. Avg. function path length (T3)	16	07	02	04	08	27	53	.05	.16	.56	.09

Table 9 (cont'd)

Variable	54	55	56	57	58	59	60	61
29. Avg cue manual time (L1)								
30. Avg cue manual time (L2)								
31. Avg cue manual time (L3)								
32. Avg strategy manual time (L1)								
33. Avg strategy manual time (L2)								
34. Avg strategy manual time (L3)								
35. Coherence (T1)								
36. Coherence (T2)								
37. Coherence (T3)								
38. Number of links (T1)								
39. Number of links (T2)								
40. Number of links (T3)								
41. Correlation (T1)								
42. Correlation (T2)								
43. Correlation (T3)								
44. Similarity (T1)								
45. Similarity (T2)								
46. Similarity (T3)								
47. Clustering coefficient (T1)								
48. Clustering coefficient (T2)								
49. Clustering coefficient (T3)								
50. Diameter (T1)								
51. Diameter (T2)								
52. Diameter (T3)								
53. Avg. path length (T1)								
54. Avg. path length (T2)								
55. Avg. path length (T3)	.32							

Variable	54	55	56	57	58	59	60	61
56. Avg. feature path length (T1)	04	01						
57. Avg. feature path length (T2)	.32	.01	.37					
58. Avg. feature path length (T3)	.03	.24	.27	.54				
59. Avg. function path length (T1)	.13	.19	.32	.09	.05	_		
60. Avg. function path length (T2)	.64	.18	09	.23	01	.23		
61. Avg. function path length (T3)	.18	.63	02	.03	.20	.19	.31	

#### **Manipulation Check**

As noted in the description of the experimental manipulation, participants in the stereotype threat condition were told that the purpose of the study was to examine sex differences in information acquisition skills and that such skills may be a reason for male-female performance discrepancies on mathematical assessments (Beilock et al., 2007; Rydell, Shiffrin et al., 2010). Two sets of items were included at the end of Day 3 to examine whether this manipulation led female participants to believe that mathematical ability was related to performance in TANDEM and/or whether perceptions of felt stereotype threat differed across study conditions. A one-sample *t*-test revealed that females' overall response to the items asking about the relevance of mathematical ability to performance in TANDEM was significantly below the mid-point of the scale (M = 2.76, SD = .85; t(157) = 3.19, p < .01). Furthermore, responses to this check did not tend to differ between females in the control (M = 2.86, SD = .84) and experimental (M = 2.66, SD = .86) conditions (t(112) = 1.22, ns), suggesting that women generally did not believe that mathematical ability was important to the experimental task. Nevertheless, females in the stereotype threat condition (M = 3.29, SD = .50) did report significantly higher levels of perceived stereotype threat than females in the control condition (M= 2.89, SD = .51) (t(112) = 4.18, p < .001, d = .79). Thus, although females were not generally convinced that TANDEM tapped skills related to mathematical ability, the stereotype threat manipulation did appear to produce the desired reaction in the targeted group.

### **Knowledge Structure Analyses**

Hypotheses 1 through 10 investigated the influence of stereotype threat on the development of knowledge structures; specifically, Hypotheses 1-5 examined the main effects of stereotype threat on the similarity, correlation, coherence, number of links, and clustering

patterns of female knowledge structures, while Hypotheses 6-10 examined changes in these indices across each day (i.e., interactive effect of condition and day on knowledge structure development). The structure of the data was such that multiple (three) knowledge structure observations were nested within subjects; consequently, multilevel random coefficient modeling (MRCM) was used to examine the predicted effects for each set of knowledge structure indices. MRCM offers a number of advantages for analyses involving longitudinal/nested data, such as the ability to produce growth estimates when data is missing without the need for imputation and the flexibility to allow different residual covariance patterns in the data to account for nonindependence among repeated measures (Bryk & Raudenbush, 1987).

Unless otherwise noted, the following procedures were followed to test Hypotheses 1-10 (cf., Bliese & Ployhart, 2002). First, a model (Model 1) in which only the time variable was included was fit to the data:

$$DV_{ti} = \pi_{0i} + \pi_{1i}(Day_{ti}) + r_{ti}$$
(1)  
$$\pi_{0i} = \beta_{00} + u_{0i}$$
  
$$\pi_{1i} = \beta_{10} + u_{1i} ,$$

where  $DV_{ti}$  is the dependent variable of interest at time *t* for individual *i*;  $\pi_{0i}$  is individual *i*'s overall mean on the DV across observations;  $\pi_{1i}$  is the amount by which the DV changes for individual *i* at each time point *t*;  $r_{ti}$  is residual variance in the DV for individual *i* at time *t*;  $\beta_{00}$  is the overall mean of the sample on the DV;  $u_{0i}$  estimates residual variance in individual *i*'s standing on the DV relative to the sample mean;  $\beta_{10}$  represents the average change in the DV over time for the sample; and  $u_{1i}$  estimates residual variance in individual *i*'s change in the DV over time relative to the sample. The results of Model 1 (specifically the estimates of  $r_{ti}$ ,  $u_{0i}$ , and  $u_{1i}$ ) can be used to provide an overall estimate of the proportion of variance attributable to within-person ( $\sigma^2$ ) versus between-person ( $\tau$ ) sources through the calculation of the intraclass correlation coefficient (ICC =  $\tau / (\tau + \sigma^2)$ ; Bryk & Raudenbush, 1992). In the present model, between-person variance in both slopes (change in the DV over time,  $\tau_{11}$ ) and intercepts (mean level of the DV controlling for slope variation,  $\tau_{00}$ ) are computed, and thus the relative proportion of variance attributable to within-person and between-person sources can be examined for both of these parameters.

Note that the categorical time variable (Day) was mean-centered prior to entry into the Level-1 MRCM equation; thus, the value of the  $\beta_{00}$  intercept reflects the overall sample mean for the DV estimated from the regression model collapsed over time. In most applications of MRCM to longitudinal analyses, one often codes the repeated measures variable such that the model intercept term(s) reflect the sample's mean standing on the DV relative to the first observation (e.g., code Day 1 as 0, Day 2 as 1, Day 3 as 2, etc.). However, the significance test of the model intercept term when the time variable has been mean centered is equivalent to the statistical test of a between-subjects factor in a repeated measures ANOVA—which is desirable in the present study given that many of the hypotheses propose simple main effects/between-group differences in the DV of interest.

Next, the main and interaction effects of the stereotype threat manipulation were modeled by adding the Condition variable to both Level-2 equations (Model 2):

$$DV_{ti} = \pi_{0i} + \pi_{1i} (Day_{ti}) + r_{ti}$$
(2)  
$$\pi_{0i} = \beta_{00} + \beta_{01} (Condition_i) + u_{0i}$$
  
$$\pi_{1i} = \beta_{10} + \beta_{11} (Condition_i) + u_{1i}$$

 $\beta_{01}$  represents the main effect of the experimental condition manipulation on the DV for individual *i*, while  $\beta_{11}$  models the interaction effect of the experimental manipulation on changes in the DV over time for individual *i*. The categorical Condition variable was dummy-coded such that the control condition (coded as 0) served as the reference variable for comparison with the stereotype threat condition (coded as 1). Consequently, the value and direction of  $\beta_{01}$  indicates the extent to which individuals in the stereotype threat condition differed from individuals in the control condition averaged across time, whereas the  $\beta_{11}$  coefficient reflects the average degree by which individuals in the stereotype threat condition changed over time relative to individuals in the control condition.

An additional model was run if the  $\beta_{01}$  or  $\beta_{11}$  coefficients achieved significance in Model 2 by adding control variables to the appropriate Level-2 equation(s) to determine whether the effects of stereotype threat remained significant after accounting for relevant between-person predictors (Model 2A). Based on the zero-order correlations presented in Table 9 and the conceptual rationale outlined previously, cognitive ability (ACT scores), working memory (OSPAN scores), and video game experience were included as control variables. The interaction between condition and math domain identification was also considered for inclusion as a Level-2 control variable, as previous researchers have argued that the influence of stereotype threat on

performance outcomes is generally greatest for individuals who are highly identified with the domain of interest (Crocker et al., 1998; Steele & Aronson, 1995; Steele & Davies, 2003). However, because participants generally did not perceive mathematical ability as relevant to completing the objectives of TANDEM, the meaning/significance of this interaction in the present study is questionable. Nevertheless, the Model 2A MRCM equations were defined as:

$$DV_{ti} = \pi_{0i} + \pi_{1i} (Day_{ti}) + r_{ti}$$
(3)  

$$\pi_{0i} = \beta_{00} + \beta_{01} (Condition_i) + \beta_{02} (ACT_i) + \beta_{03} (OSPAN_i) + \beta_{04} (Games_i) + \beta_{05} (MathID_i^*Condition_i) + u_{0i}$$
(3)  

$$\pi_{1i} = \beta_{10} + \beta_{11} (Condition_i) + \beta_{12} (ACT_i) + \beta_{13} (OSPAN_i) + \beta_{14} (Games_i) + \beta_{15} (MathID_i^*Condition_i) + u_{1i}$$

Note that because all study hypotheses only concerned differences in the knowledge structures of female participants (as they were the intended target of the stereotype threat manipulation), the above analyses only include data from female participants. As such, the continuous control variables for Model 2A were mean-centered based on the female sample prior to entry into the MRCM equations to facilitate interpretation of their regression coefficients.

In sum, the analytic approach for testing the effects of stereotype threat on knowledge structure formation was as follows:

- <u>Model 1</u>: Fit a model including only the time variable at Level-1 to examine the proportion of variance in slopes and intercepts attributable to between- and within-person sources.
- <u>Model 2</u>: Add the condition effect to both Level-2 equations; evaluate the direction and significance of the  $\beta_{01}$  coefficient to examine the main effect of stereotype threat on the

outcome of interest (e.g., DV differed between individuals in the stereotype threat versus control conditions on average) and the direction and significance of the  $\beta_{11}$  coefficient to determine whether changes in the DV over time differed between conditions.

• <u>Model 2A</u>: If either  $\beta_{01}$  or  $\beta_{11}$  are significant, add control variables to the Level-2 model to determine whether the influence of stereotype threat contributed above and beyond relevant between-person predictors.

All MRCM models were computed in R version 2.15 (R Development Core Team, 2012) using the *lme4* package (Bates, Maechler, & Bolker, 2011). Since the tests for the predicted main and interactive effects were conducted within the same single MRCM model (Model 2 and/or Model 2A), the results below are organized by knowledge structure outcome rather than hypothesis ordering for convenience.

**Knowledge structure similarity.** Hypotheses 1 and 6 examined differences in the similarity of knowledge structures produced by threatened versus non-threatened females to those produced by males and top performers<sup>8</sup>; Table 10 summarizes the results of the MRCM models for these hypotheses. With respect to similarity with male knowledge structures, Model 1 indicated that 53% of the variance in mean similarity and 43% of the variance in similarity change over time could be attributed to between-person differences. However, the results of Model 2 indicated that neither the main effect ( $\beta_{01} = -.01$ , *ns*) nor interaction effect ( $\beta_{11} = -.01$ , *ns*) of stereotype threat on structural similarity was significant, suggesting that the knowledge structures for females in both experimental conditions were equally similar to male knowledge structures on average and that the rate of change in similarity did not significantly differ across conditions of stereotype threat. With respect to the similarity with the top performers' knowledge

Table 10

Referent	Madal			Param	eter est	timates		
Structure	Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$
	Model 1							
	Similarity <sub>ti</sub> = $\pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	10		01		005	006	004
	$\pi_{0i} = \beta_{00} + u_{0i}$	.18		.01		.005	.000	.004
Males	$\pi_{1i} = \beta_{10} + u_{1i}$							
Iviales	Model 2							
	Similarity <sub>ti</sub> = $\pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	10	01	02	01	005	006	004
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	.19	01	.02	01	.005	.000	.004
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	Model 1							
	Similarity <sub>ti</sub> = $\pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	10		02		003	005	002
	$\pi_{0i} = \beta_{00} + u_{0i}$	.10		.02		.005	.005	.002
Top 15	$\pi_{1i} = \beta_{10} + u_{1i}$							
100 15	Model 2							
	Similarity <sub>ti</sub> = $\pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	10	- 01	02	- 01	003	005	002
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	.17	01	.02	01	.005	.005	.002
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							

MRCM Parameter Estimates for Female's Knowledge Structure Similarity with Males and the Top 15 Performers (Hypotheses 1 & 6)

structures, approximately 61% of the variance in mean similarity and 41% of the variance in similarity change over time was attributable to between-person differences. The results from Model 2 revealed that although females' knowledge structures tended to become more similar to top performers over time on average ( $\beta_{10} = .02, p < .05$ ), there were no differences in knowledge structure similarity across condition ( $\beta_{01} = -.01, ns$ ) nor did changes in similarity to top performers over time differ between conditions ( $\beta_{11} = -.01, ns$ ). In sum, neither Hypothesis 1 nor Hypothesis 6 was supported.

**Knowledge structure correlation.** Hypothesis 2 proposed that the knowledge structures of threatened females relative to those produced by males and top performers would be less correlated on average than the knowledge structures of non-threatened females, while Hypothesis 7 postulated that this correlation would increase more slowly over time for threatened females than for non-threatened females. The results of the MRCM models for these hypotheses are summarized in Table 11. Using male knowledge structures as the referent, Model 1 revealed that a significant portion of the variance in females' mean correlation index was attributable to between-person factors (84%), while the amount of change in knowledge structure correlations across days tended to vary less across individuals (31% of variance in slopes attributable to between-person variables). Results from Model 2 revealed that, on average, the knowledge structures of females and males did become more strongly correlated over time ( $\beta_{10} = .03, p$ < .05); however, the knowledge structures of threatened females were not significantly less correlated on average ( $\beta_{01} = -.03$ , *ns*) nor did they converge towards male knowledge structures at a significantly slower rate ( $\beta_{11} = -.02$ , ns). Using top performers as the referent again indicated that average knowledge structure correlations varied substantially across females (ICC = .83), with rates of change tending to be less variable (ICC = .32). Similar to the previous set of analyses, on average female knowledge structures tended to become more strongly correlated with top performers over time ( $\beta_{10} = .04, p < .05$ ); again, however, the main ( $\beta_{01} = -.02, ns$ ) and interaction effects ( $\beta_{11} = -.01$ , ns) for Condition failed to reach significance, indicating that stereotype threat did not influence the average structural correlation with top performers or changes in correlation across days. Consequently, the pattern of results failed to support either Hypothesis 2 or Hypothesis 7.

Table 11

Referent	Model			Param	eter est	timates		
Structure	Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ00	$\tau_{11}$
	Model 1							
	$\operatorname{Corr}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Day}_{ti}) + r_{ti}$	12		03		010	052	005
	$\pi_{0i} = \beta_{00} + u_{0i}$	.42		.02		.010	.032	.005
Males	$\pi_{1i} = \beta_{10} + u_{1i}$							
Iviales	Model 2							
	$\operatorname{Corr}_{ti} = \pi_{0i} + \pi_{1i} \left( \operatorname{Day}_{ti} \right) + r_{ti}$	12	02	02	02	010	052	005
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	.43	05	.05	02	.010	.035	.005
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	Model 1							
	$\operatorname{Corr}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Day}_{ti}) + r_{ti}$	40		04		011	054	005
	$\pi_{0i} = \beta_{00} + u_{0i}$	.40		.04		.011	.034	.005
Top 15	$\pi_{1i} = \beta_{10} + u_{1i}$							
100 10	Model 2							
	$\operatorname{Corr}_{ti} = \pi_{0i} + \pi_{1i} \left( \operatorname{Day}_{ti} \right) + r_{ti}$	41	- 02	04	- 01	011	055	005
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	.71	02	.07	.01	.011	.055	.005
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							

MRCM Parameter Estimates for Female's Knowledge Structure Correlation with Males and the Top 15 Performers (Hypotheses 2 & 7)

**Knowledge structure coherence.** Hypothesis 3 and Hypothesis 8 examined differences in the coherence of knowledge structures produced by females in the stereotype threat versus control conditions as well as differences in changes to knowledge structure coherence over time (Table 12). Computation of the ICCs revealed that 73% of the variance in females' knowledge structure coherence and 23% of the variance in coherence growth rates was attributable to differences at the individual level. The addition of the between-subject Condition variable in Model 2 revealed that in general, the coherence of females' knowledge structures did not significantly improve over time ( $\beta_{10} = .03$ , *ns*); furthermore, no significant differences between

Model -			Param	eter est	imates		
Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ00	$\tau_{11}$
Model 1							
$Coherence_{ti} = \pi_{0i} + \pi_{1i}(Day_{ti}) + r_{ti}$	27		02		026	070	008
$\pi_{0i} = \beta_{00} + u_{0i}$	.32		.02		.020	.070	.008
$\pi_{1i} = \beta_{10} + u_{1i}$							
Model 2							
$Coherence_{ti} = \pi_{0i} + \pi_{1i}(Day_{ti}) + r_{ti}$	24	05	02	02	026	070	000
$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	.34	05	.05	02	.020	.070	.008
$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							

Table 12MRCM Parameter Estimates for Female's Knowledge Structure Coherence(Hypotheses 3 & 8)

mean levels of coherence ( $\beta_{01} = -.05$ , *ns*) or changes in coherence over time ( $\beta_{11} = -.02$ , *ns*) were observed between conditions of stereotype threat, thus failing to support Hypotheses 3 and 8.

Number of knowledge structure links. The next set of predictions examined differences in the number of links present in the knowledge structures of threatened versus control females (Hypothesis 4) and the manner by which the number of network links changed over time across condition (Hypothesis 9). Table 13 presents the results of the MRCM analyses for this outcome. Overall, both the average number of links in females' knowledge structures (ICC = .44) and the rate of change in number of links across days (ICC = .08) did not vary dramatically across individuals. The final Model 2 analyses revealed that, on average, female participants' knowledge structures tended to become more interconnected over time, growing by approximately 3 links each day ( $\beta_{10} = 2.96$ , p < .05). However, the average number of links in the knowledge structure of females under stereotype threat did not significantly differ from those in the control condition ( $\beta_{01} = 3.12$ , *ns*); furthermore, there were no differences between

Table 13

Madal			Param	eter est	timates		
Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	$\tau_{00}$	$\tau_{11}$
Model 1							
$\text{Links}_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	22 52		2 20		162.2	120.4	14.07
$\pi_{0i} = \beta_{00} + u_{0i}$	32.32		5.59		102.2	129.4	14.77
$\pi_{1i} = \beta_{10} + u_{1i}$							
Model 2							
$Links_{ti} = \pi_{0i} + \pi_{1i}(Day_{ti}) + r_{ti}$	21.01	2 1 2	2.00	00	167 1	120.05	15 21
$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	31.01	3.12	2.90	.90	102.1	130.03	13.31
$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							

*MRCM Parameter Estimates for Number of Links in Female's Knowledge Structures* (*Hypotheses 4 & 9*)

conditions in the number of links added per day ( $\beta_{11} = .90, ns$ ). In sum, no differences in the average number of knowledge structure links nor in the number of knowledge structure links added over time were observed for females between conditions of stereotype threat, therefore failing to support Hypotheses 4 and 9.

**Knowledge structure clustering.** Hypotheses 5 and 10 predicted that differences in the average compositional form of knowledge structures as well as the development of functional relations among knowledge structure concepts over time would emerge between females learning TANDEM under conditions of stereotype threat versus females in the control condition. To examine differences among clustering patterns, two sets of visual representations for participants' knowledge structure were computed using the Pathfinder software and qualitative comparisons of their form were performed. To investigate whether the experimental manipulation exerted an effect on the development of structural relationships amongst concepts on average (Hypothesis 5), a single proximity matrix composed of the average pair-wise

similarity ratings amongst concepts collapsed across days was computed for females in each of the experimental conditions. This procedure resulted in the creation of two networks representing the knowledge structures for females in each condition averaged across all time points (Figure 7). To investigate changes in knowledge structure development over time, an averaged proximity matrix was also computed for female participants in each condition at each day, resulting in six knowledge structures depicting the average network at each day separately for females in the stereotype threat and control conditions (Figures 8-10).

In large part, Figure 7 illustrates that the knowledge structures of females in the stereotype threat and control conditions were remarkably alike when averaged across time. Both groups appeared to draw a distinction between the decision-making and procedural/strategic concepts, with gaining/losing points (Points) generally serving as the logical connection between these two sets of clusters. Control condition females appeared to more strongly associate monitoring of the inner perimeter (MonInn) with scoring points in the task, whereas threatened females generally associated monitoring of the outer perimeter (MonOut) with this aspect of the task. Control females also seemed to associate the prioritization of critical targets (Priority) primarily with monitoring the inner defensive perimeter, while stereotype threat females associated prioritization with finding/engaging pop-up targets (PopUp)—perhaps suggesting that stereotype threat females were more actively seeking out pop-up targets and/or willing to immediately prosecute new targets rather than focusing their attention on protecting a particular defensive perimeter.

Evidence for average differences in knowledge organization on the basis of feature and functional similarity was mixed. Females in both conditions seemed to draw equally distinctive functional relationships with respect to the Intent subdecisions; the classification of a target as



*Figure* 7. Knowledge structures for female participants in the stereotype threat and control conditions averaged across days



Figure 8. Average knowledge structures for female participants in the stereotype threat and control conditions at end of Day 1



Figure 9. Average knowledge structures for female participants in the stereotype threat and control conditions at end of Day 2



Figure 10. Average knowledge structures for female participants in the stereotype threat and control conditions at end of Day 3

Peaceful (idPeac) or Hostile (idHost) were both related with their most probable Final Engagement outcomes (Clear and Mark, respectively; see Table 8). The relations amongst many of the other decision-making concepts, though, were more ambiguous. However, this pattern of results was perhaps not entirely surprising given that the outcomes of the Intent subdecision are among the most diagnostic/informative regarding how to apply the rules of engagement (Table 4) in order to make the correct Final Engagement decision. As shown in Table 8, the Intent subdecision contained only two possible outcomes (Peaceful or Hostile), each of which was only associated with two (as opposed to all three) Final Engagement decision outcomes; further, one of those Final Engagement options was always twice as likely to be correct (e.g., Clearing a Peaceful contact was likely to be correct 67% of the time, while Warning that target was likely to be correct only 33% of the time). As such, the classification of a target as either Peaceful or Hostile carried with it a relatively high degree of certainty/informative value regarding the correct Final Engagement decision compared with the Type and Class subdecisions, suggesting that a strong functional association among the Intent and Final Engagement concepts may have been easier for participants to infer. Regardless, the lack of noticeable differences in the aggregate structural relations of females between conditions does not lend support to the predictions of Hypothesis 5.

Although comparison of the aggregated knowledge structures provides a broad overview of the manner by which females in both conditions perceived relations among task-relevant concepts and information, it does not account for the fact that individuals' knowledge structures were likely to change as they gained more experience within the task domain. Hypothesis 10 therefore sought to examine whether the development and growth in the knowledge structures of females under stereotype threat differed from that of females in the control condition. To this end,

Figures 8-10 reveal a number of intriguing differences in the pattern of knowledge structure development between these two groups of female learners. By the end of Day 3, a distinctive pattern had developed in the knowledge structure of female participants in the control condition that was not present in the structure of stereotype threat females. Namely, the acquisition of points emerged as a central hub in the network of control females (similar to the knowledge structures of male participants), which was subsequently linked to all three Final Engagement decision outcomes (Clear, Warn, Mark). In turn, each of these Final Engagement outcomes was linked to its single most probable Class and Intent subdecision outcome (e.g., Mark related to the identification of a Hostile target, Warn related to the identification of a Civilian target, etc.; see Table 8). Lastly, although the outcomes related to the Type subdecision (idAir, idSurf, and idSub) were not associated with a particular Final Engagement decision, they were each most strongly/directly related to gaining/losing points.

One plausible interpretation of this structural pattern is that, by the end of Day 3, control females had developed an efficient/simplified heuristic for making Final Engagement decisions that they used for earning points in the task. Before describing the specific form of this heuristic, it is helpful to again consider Table 8 and the relative probabilities between each subdecision outcome and the three Final Engagement outcomes in order to understand why it is an efficient and effective means by which to make decisions in the present version of TANDEM. As was detailed above, the Intent subdecision was arguably the most easily interpretable subdecision for helping an individual determine how to apply the rules of engagement in order to produce the correct Final Engagement decision. After this, the Class subdecision was likely the next most informative subdecision. Similar to the Intent subdecision, there were only two possible Class outcomes from which to choose (Civilian and Military), and one of these outcomes (Civilian)

was highly diagnostic of the correct Final Engagement decision. In contrast, the Type subdecision possessed three possible outcomes (Air, Surface, Sub), each of which could be associated with any of the three Final Engagement decisions with probabilities that were not vastly different from one another (e.g., 50% of Air targets were likely to be Warned, 25% Cleared, and 25% Marked); consequently, integrating the Type subdecision into one's Final Engagement choice was likely the most difficult part of the decision procedure.

Returning to the implicit heuristic implied by the control condition females' knowledge structures then, by Day 3, learners in this group appeared to have made the diagnostic functional connections between the Class and Intent subdecisions for a target and the Final Engagement decision which those pieces of information suggested was most likely correct and which would earn them points in the task. Once these pieces of information about a target are known, the correct Final Engagement decision becomes substantially easier and, more often than not, can be made correctly regardless of whether an individual has learned the relatively more difficult functional relationships for the Type subdecision (note that the person must still make the Type subdecision *correctly* in order to earn points, but in many cases they do not need to functionally *integrate* that information in order to make the correct Final Engagement decision). More specifically, the information presented in Tables 4 and 8 indicate that if a person follows the heuristic:

- 1. If a target is Civilian, Warn it; if target is Military, go to Step 2
- 2. Choose whatever Final Engagement decision is most probable based on the targets' Intent,

then the correct Final Engagement decision will be made, on average, 84% of the time without even needing to consider the target's Type (Step 1 will lead to correct Final Engagement

decision in 67% of occasions, whereas Step 2 will lead to the correct Final Engagement decision in 100% of occasions).

While simply examining the knowledge structures for control condition females does not indicate that these participants were following this decision-making heuristic (the results of Hypothesis 12 present a more detailed examination of this possibility), the observed changes in the pattern of network concepts over time is consistent with learners in this group acquiring this highly efficient/effective decision process. Control females appeared to have learned the relatively easier functional relationships for the Intent subdecisions by the end of Day 1, though they had yet to fully make sense of the remaining subdecision outcomes. By Day 2, the importance of the three Final Engagement outcomes in relation to scoring points had been established and learners seemed to be drawing more clearly interpretable associations between the Final Engagement and Class outcomes. Finally, at the end of Day 3, the structural relations consistent with the decision heuristic outlined above had been achieved. Also of note, an organized structure amongst the procedural/strategic task concepts did not emerge until Day 3 as well, perhaps indicating that control condition females delayed learning/practicing these functions until they had developed more expertise with the foundational decision-making components of the TANDEM task environment.

In contrast, the pattern of concept relations in the knowledge structures of females in the stereotype threat condition differed rather substantially from that described above. Unlike control condition females, there was never a point in time where stereotype threat females' average daily knowledge structure exhibited a pattern in which gaining/losing points was associated with all three Final Engagement decisions, which in turn were then associated with their single most diagnostic Class/Intent outcomes. Furthermore, stereotype threat females' knowledge structures

were the only networks in which gaining/losing points was not always the most interconnected node in the knowledge structure on any given day. In fact, at Day 1, gaining/losing points was most strongly related to monitoring the outer perimeter, perhaps implying that females under threat were too concerned early on with gaining/losing points as a result of procedural/strategic aspects of the task (e.g., ensuring targets did not cross the invisible outer perimeter) rather than the manner by which the more fundamental target engagement/decision-making processes affected task performance. A greater focus on these advanced task concepts is further evidenced by the fact that relations among the procedural/strategic task concepts had already begun to exhibit a logical structure by the end of Day 2, which was earlier than what was observed for control condition females.

Of final interest, it did appear that threatened females were generally inferring appropriate functional relationships between the various subdecision and Final Engagement outcome concepts; however, the pattern of structural relations suggests that this group of learners may have been doing so in a less efficient manner. More specifically, the knowledge structures of stereotype threat learners were more likely to possess relational patterns in which a single subdecision outcome (e.g., identification of target as Hostile) was related to multiple Final Engagement outcomes (e.g., Warn and Mark). As noted above in the description of the efficiency decision heuristic, the key interpretative inference that is needed when an individual identifies a target as possessing a particular Class or Intent is "What is the *single* most likely Final Engagement decision for the target based on its classification?," not "What are *all* the possible Final Engagement decision outcomes for the target based on its classification?" That is, if an individual identifies a target as Hostile, then it is relatively more efficient/diagnostic to know that the most probable engagement action is to Mark that contact rather than knowing that a Hostile

contact could be either Warned or Marked. The clustering pattern in which multiple engagement outcomes were linked to a single subdecision outcome was observed once in the Day 2 knowledge structure (identification of Hostile targets) and twice in the Day 3 knowledge structure (identification of Hostile targets and identification of Military targets) of threatened females. Although these functional relations are not necessarily "wrong," they do suggest that threatened female learners may have been focusing their learning efforts on memorizing the entire distribution of final engagement decisions rather than attempting to learn the seemingly more efficient heuristic approach noted above.

To summarize, a number of differences were observed in the progression of female's knowledge structure development over time across conditions of stereotype threat. In general, females who did not experience the stereotype threat manipulation appeared to organize information related to decision-making in TANDEM in a manner consistent with an efficient and reasonably effective decision heuristic for scoring points in the task. Furthermore, there was tentative evidence that these female learners may have also delayed learning/practicing more advanced strategic aspects until these more fundamental processes were learned. Alternatively, the knowledge structures of females experiencing stereotype threat during learning appeared to be organized in a less efficient manner, and were instead consistent with an approach in which individuals attempted to learn through brute memorization rather than identifying the most informative/diagnostic relations among concepts. Threatened females may have also been somewhat more likely to attempt learning the more advanced procedures of task performance earlier during learning activities well before they had effectively learned more basic task concepts. On the basis of this evidence, the predictions of Hypothesis 10 were largely supported.

Exploratory analyses using the graph theoretic metrics were also conducted to examine mean differences in the composition of female knowledge structures between conditions and across time. The results of these MRCM analyses are presented in Table 14. Overall, the pattern of results revealed no significant main effects or interaction effects between groups. On the whole, the results revealed that the knowledge structures of all female learners tended to become more tightly interconnected over time (e.g., average shortest path length between node pairs decreased, diameter of network decreased, and clustering coefficient increased). Additionally, this trend appeared to equally influence the distances among both feature (links between concepts from the same subdecision) and functional (links between subdecision concepts and most probable Final Engagement outcome) network relations.

### **Cognitive Strategy Analyses**

**Strategic learning behaviors.** Hypothesis 11 proposed that females learning under stereotype threat would exhibit poorer/more basic task strategies than females in the control condition. The analytic approach used to evaluate this prediction was essentially identical to that outlined for the knowledge structure analyses. As noted in the Methods section, a number of variables were recorded and analyzed to evaluate the strategic learning/performance demonstrated by individuals in the task. Consequently, the results in this section are organized into three areas. The first, *knowledge acquisition behaviors*, presents analyses from participants' use and study with the online task manual prior to each TANDEM trial. The second section, *task practice behaviors*, examines data from participants' actions within the game that are representative of strategic learning and performance; these data includes the number of marker targets engaged, the number of times participants zoomed their radar screen in/out to help monitor defensive perimeters, the number of high priority targets engaged, and the total number

Table 14

Dependent	Madal	Parameter estimates									
Variable	Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$			
	Model 1										
	$L_{ti} = \pi_{0i} + \pi_{1i}(\mathrm{Day}_{ti}) + r_{ti}$	2 25		10		100	070	002			
Avg.	$\pi_{0i} = \beta_{00} + u_{0i}$	2.23	_	10		.190	.070	.002			
Shortest	$\pi_{1i} = \beta_{10} + u_{1i}$										
Path	Model 2										
Length	$L_{ti} = \pi_{0i} + \pi_{1i}(\mathrm{Day}_{ti}) + r_{ti}$	2 28	- 04	- 08	- 04	190	071	003			
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	2.20	04	00	04	.170	.071	.005			
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$										
	Model 1										
	$D_{ti} = \pi_{0i} + \pi_{1i}(\mathrm{Day}_{ti}) + r_{ti}$	4 56		- 27		1 77	548	034			
	$\pi_{0i} = \beta_{00} + u_{0i}$	1.50		-•27		1.,,	.5 10	.054			
Network	$\pi_{1i} = \beta_{10} + u_{1i}$										
Diameter	Model 2				07	1 77					
	$D_{ti} = \pi_{0i} + \pi_{1i}(\mathrm{Day}_{ti}) + r_{ti}$	1 61	- 16	24			554	035			
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	4.04	10	24	07	1.//	.554	.055			
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$										
	Model 1										
	$C_{ti} = \pi_{0i} + \pi_{1i}(\mathrm{Day}_{ti}) + r_{ti}$	17		02		006	004				
	$\pi_{0i} = \beta_{00} + u_{0i}$	•1 /	_	.02		.000	.004	.000			
Cluster	$\pi_{1i} = \beta_{10} + u_{1i}$										
Coeff.	Model 2										
	$C_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	16	01	03	01	006	004	.000 <sup>†</sup>			
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	.10	.01	.01 <b>.02</b>	.01	.000	5 .004				
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$										

MRCM Parameter Estimates for Graph Theoretic Metrics (Exploratory Analyses)

Table 14 (cont'd)

DV	Model	Parameter estimates						
		$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$
Avg. Feature Path Length	Model 1							
	$L\text{-}Feat_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	2.04		10		.281	.226	.046
	$\pi_{0i} = \beta_{00} + u_{0i}$							
	$\pi_{1i} = \beta_{10} + u_{1i}$							
	Model 2							
	$L-Feat_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	2.03	.02	12	.05	.283	.227	.046
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$							
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
Avg. Function Path Length	Model 1	1.98		12		.242	.100	.011
	$L-Func_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$							
	$\pi_{0i} = \beta_{00} + u_{0i}$							
	$\pi_{1i} = \beta_{10} + u_{1i}$							
	Model 2							
	$L-Func_{ti} = \pi_{0i} + \pi_{1i} \left( \text{Day}_{ti} \right) + r_{ti}$	1.95	.07	08 <sup>†</sup>	06	.222	.055	.012
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$							
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							

<sup>†</sup>Signficant at p < .10

of targets engaged. Lastly, the section on *self-regulation* presents results from the self-reported metacognitive activity scale completed at the end of each day. The same progression of MRCM equations used for the knowledge structure analyses (Equations 1 through 3) were fit for each of these dependent variables, and the direction and significance of the main effect ( $\beta_{01}$ ) and cross-level interaction ( $\beta_{11}$ ) terms for the Condition variable were examined. Given that the primary focus of these analyses was to determine whether significant differences between the experimental conditions for females were observed during learning activities, data from the

practice trials (6-18 observations per person) rather than the performance trials (1-3 observations) were used in the analyses for the task manual and practice behavior variables listed above.

**Knowledge acquisition behaviors.** To facilitate interpretation of how participants spent their time studying the online task manual prior to each learning trial, pages within the manual were coded into categories according to the type of information each page conveyed (basic gameplay, cue value, and task strategy information). An additional category (null) was defined that included time spent on the introductory menu page and/or the task exit screen. To provide an overall perspective on the manner by which individuals organized their task manual study time, the cumulative average and average amount of time spent on the various manual sections were plotted for the control and stereotype threat female learners for each trial (Figures 9 and 10, respectively). Figure 9 clearly shows that the study patterns of females in both conditions were fairly similar throughout the Day 1 and Day 2 learning trials; however, a marked decrease was observed in the average amount of time stereotype threat females spent studying the manual during the final Day 3 learning trials. Of additional interest, Figure 10 indicated that the relationship between time/trial and the amount of time spent studying certain sections of the task manual was likely not a strict linear function. Consequently, a quadratic time term was added as random effect to the Level-1 equation of the subsequent MRCM analyses to better model these observations in the subsequent analyses.

Table 15 presents the results from the MRCM analyses for each of the manual sections as well as the overall time spent studying the manual. As expected, a comparison of the Model 1  $\beta_{00}$  coefficients across each of the manual sections revealed that on average, learners spent just over half of their available study time (57%) on the cue value manual pages and an additional one quarter (25%) of their time reading the task strategy pages. An examination of the linear



Figure 11. Cumulative average time spent viewing manual pages during learning trials



Figure 12. Average time spent viewing manual pages during learning trials

Dependent Variable	Model	Parameter Estimates						
		$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\beta_{20}$	$\beta_{21}$	
Time spent on basic manual pages	Model 1	4.04	_	60		.09	_	
	$Basic_{ti} = \pi_{0i} + \pi_{1i}(Trial_{ti}) +$							
	$\pi_{2i}(\mathrm{Trial}_{ti})^2 + r_{ti}$							
	$\pi_{0i} = \beta_{00} + u_{0i}$							
	$\pi_{1i} = \beta_{10} + u_{1i}$							
	$\pi_{2i} = \beta_{20} + u_{2i}$							
	Model 2A <sup>a</sup>	3.02	2.79	60	.12	.13	11	
	$Basic_{ti} = \pi_{0i} + \pi_{1i}(Trial_{ti}) +$							
	$\pi_{2i}(\mathrm{Trial}_{ti})^2 + r_{ti}$							
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$							
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	$\pi_{2i} = \beta_{20} + \beta_{21}$ (Condition <sub>i</sub> ) + $u_{2i}$							
	Model 1	68.59		-2.28		10		
	$\operatorname{Cue}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) +$							
	$\pi_{2i}(\mathrm{Trial}_{ti})^2 + r_{ti}$							
	$\pi_{0i} = \beta_{00} + u_{0i}$							
Time spent on cue value manual pages	$\pi_{1i} = \beta_{10} + u_{1i}$							
	$\pi_{2i} = \beta_{20} + u_{2i}$							
	Model 2A <sup>a</sup>	71.49	-8.07	-1.10	-2.59	05	10	
	$\operatorname{Cue}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) +$							
	$\pi_{2i}(\mathrm{Trial}_{ti})^2 + r_{ti}$							
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$							
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	$\pi_{2i} = \beta_{20} + \beta_{21}(\text{Condition}_i) + u_{2i}$							

 Table 15

 MRCM Parameter Estimates for Time Spent on Task Manual Sections (Hypothesis 11)
Table 15 (cont'd)

Dependent	Madal	Parameter Estimates							
Variable	Widdel	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\beta_{20}$	$\beta_{21}$		
Time spent on	Model 1 Strat <sub><i>ii</i></sub> = $\pi_{0i} + \pi_{1i}(\text{Trial}_{ii}) + \pi_{2i}(\text{Trial}_{ii})^2 + r_{ti}$ $\pi_{0i} = \beta_{00} + u_{0i}$ $\pi_{1i} = \beta_{10} + u_{1i}$ $\pi_{2i} = \beta_{20} + u_{2i}$	29.50		82		27			
strategy manual pages	Model 2A <sup>a</sup> Strat <sub><i>ii</i></sub> = $\pi_{0i} + \pi_{1i}(\text{Trial}_{ii}) + \pi_{2i}(\text{Trial}_{ii})^2 + r_{ti}$ $\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$ $\pi_{2i} = \beta_{20} + \beta_{21}(\text{Condition}_i) + u_{2i}$	32.27	-4.99	39	87	28	01		
Time spent on	Model 1 Null <sub>ti</sub> = $\pi_{0i} + \pi_{1i}(\text{Trial}_{ti}) + \pi_{2i}(\text{Trial}_{ti})^2 + r_{ti}$ $\pi_{0i} = \beta_{00} + u_{0i}$ $\pi_{1i} = \beta_{10} + u_{1i}$ $\pi_{2i} = \beta_{20} + u_{2i}$	10.43		.82		.01			
manual pages	Model 2A <sup>a</sup> Null <sub>ti</sub> = $\pi_{0i} + \pi_{1i}(\text{Trial}_{ti}) + \pi_{2i}(\text{Trial}_{ti})^2 + r_{ti}$ $\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$ $\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$ $\pi_{2i} = \beta_{20} + \beta_{21}(\text{Condition}_i) + u_{2i}$	10.04	17	.99	47	.04	07		

Table 15 (cont'd)

Dependent	Madal		Pa	rameter	Estimat	tes	
Variable	Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\beta_{20}$	$\beta_{21}$
	Model 1						
$Total_{ti} = \pi_{0i} + \pi_{1i}u$ $\pi_{2i}(Trial_t)$ $\pi_{0i} = \beta_{00} + u_{0i}$ $\pi_{1i} = \beta_{10} + u_{1i}$	$\text{Total}_{ti} = \pi_{0i} + \pi_{1i}(\text{Trial}_{ti}) +$			-2.78	_	27	_
	$\pi_{2i}(\mathrm{Trial}_{ti})^2 + r_{ti}$	112.6					
	$\pi_{0i} = \beta_{00} + u_{0i}$						
	$\pi_{1i} = \beta_{10} + u_{1i}$						
spent on	$\pi_{2i} = \beta_{20} + u_{2i}$						
manual	Model 2A <sup>a</sup>						
pages	$\text{Total}_{ti} = \pi_{0i} + \pi_{1i}(\text{Trial}_{ti}) +$						
	$\pi_{2i}(\mathrm{Trial}_{ti})^2 + r_{ti}$	117 2	_10.9	-1 08	-3 80	- 15	- 29
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	11/,4	-10.7	-1.00	-5.00	15	29
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$						
	$\pi_{2i} = \beta_{20} + \beta_{21}(\text{Condition}_i) + u_{2i}$						

Coefficient estimates in bold are significant at p < .05

<sup>†</sup>Signficant at p < .10

<sup>a</sup>Model includes control variables (coefficients not printed for ease of presentation)

main effect for trial ( $\beta_{10}$ ) on total time spent studying the manual revealed that, on average, all female learners tended to spend less time reading the manual over time; this was true for each manual section except the null pages, which demonstrated a slight increase over time. Consistent with the trends shown in Figure 9, the significant negative coefficient for the quadratic trial variable ( $\beta_{20}$ ) on total study time and task strategy study indicated that the decrease in the amount of time spent viewing the manual tended to become more exaggerated during the later study trials.

Analysis of the main effects of the stereotype threat manipulation revealed that stereotype threatened female participants spent significantly more time on the basic manual pages ( $\beta_{01}$  =

2.79, p < .05), less time on the cue value pages ( $\beta_{01} = -8.07, p < .05$ ), and less total time studying overall ( $\beta_{01}$  = -10.9, p < .05). Interaction effects between stereotype threat and the linear time variable were also observed for time spent on the cue value pages ( $\beta_{11} = -2.59$ , p < .05), task strategy pages ( $\beta_{11} = -87, p < .05$ ), and total time spent studying ( $\beta_{11} = -3.80, p < .05$ ); in all cases, the direction of the effect indicated that stereotype threatened females tended to spend less time studying these pages at each trial than control condition females. Lastly, a significant interaction between stereotype threat and the quadratic trial variable was observed for both the amount of time spent studying the basic gameplay section ( $\beta_{21} = -.11, p < .05$ ) as well as overall study time ( $\beta_{21} = -.29$ , p < .05). In the former case, Figure 10 shows that control condition females tended to spend slightly longer on the basic gameplay pages early on in the task, which tapered away throughout the trials; alternatively, stereotype threatened females spent less time on these pages early on and did not tend to revisit them later in the study. The significant quadratic interaction for overall study time reflects the noted drop-off in study time observed for stereotype threat participants following Day 2 relative to control condition females (Figure 9).

In sum, analysis of the task manual data largely supported the predictions of Hypothesis 11. The significant mean differences found for both the overall time spent studying and time spent studying the critical cue value/task strategy portions of the manuals reflected poorer learning behaviors on the part of stereotype threat individuals. Although not predicted as such, the unique pattern of variation in study time over the course of the learning trials for stereotype threat participants was also consistent with predictions from stereotype threat theory, though perhaps more so with its influence on motivation rather than cognition.

Task practice behaviors. Results from the MRCM analyses and graphs contrasting control and stereotype threatened females on the four focal task practice behaviors are presented in Table 16 and Figure 11, respectively. ICCs for variance in the intercept terms were relatively moderate across the four task practice behaviors, ranging from .39 to .55; however, variation in slopes was virtually nonexistent (ICCs ranging from .01 to .02), indicating that changes in time for the modeled variables were highly similar across all females. In general, the MRCM analyses revealed that the number of marker targets engaged ( $\beta_{10} = .02, p = .08$ ), zoom activities performed ( $\beta_{10} = .42, p < .05$ ), high priority targets engaged ( $\beta_{10} = .15, p < .05$ ), and total number of targets engaged ( $\beta_{10} = .30, p < .05$ ) tended to increase across task trials for all females. However, no main or interaction effect of stereotype threat achieved significance for any of the task practice variables, indicating that females in both groups were typically engaging in similar practice behaviors during the learning trials. As can be seen in Figure 11, it appeared that the number of marker targets engaged over time may have been changing at a different rate for the different groups, suggesting that modeling a quadratic time variable might improve model parameter estimates. Although the quadratic time coefficient achieved significance in this model  $(\beta_{20} = -.01, p < .05, \text{ indicating that the number of marker targets engaged initially increased})$ rapidly and then decreased in later trials), the main effect and both the linear and quadratic interaction effects of stereotype threat failed to achieve significance.

In sum, analyses of the task practice behaviors did not support the predictions of Hypothesis 11. Threatened females were not more likely to engage marker targets, ignore defensive perimeters, or fail to prosecute high priority targets than control condition females. Contrary to the results observed with participants' use of the task manual, both groups also

Dependent	Madal			Param	eter es	timates		
Variable	Middel	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$
	Model 1						.641	
Number of Marker	$Marker_{ti} = \pi_{0i} + \pi_{1i}(Trial_{ti}) + r_{ti}$	65		02		945		.006
	$\pi_{0i} = \beta_{00} + u_{0i}$	.05		.02		.775		
	$\pi_{1i} = \beta_{10} + u_{1i}$							
Targets	Model 2							
Engaged	$Marker_{ti} = \pi_{0i} + \pi_{1i}(Trial_{ti}) + r_{ti}$	55	21	.02 <sup>†</sup>	00	945	635	006
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$		.21		.00	.943	.033	.000
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	Model 1					62.7	40.6	
	$\operatorname{Zoom}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) + r_{ti}$	9.49		48				57
	$\pi_{0i} = \beta_{00} + u_{0i}$		_	.40				.37
Number of Zoom	$\pi_{1i} = \beta_{10} + u_{1i}$							
Actions	Model 2	0.07	1.04	.42		62 7	40.7	
	$\operatorname{Zoom}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) + r_{ti}$				.11			57
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	0.97	1.00			02.7	40.7	.37
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	Model 1							
	$\operatorname{HiPrior}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) + r_{ti}$	4 89		16		1 1 5	1 44	027
Number	$\pi_{0i} = \beta_{00} + u_{0i}$	<b>T.</b> 07		.10		1.10	1.77	.027
of High Priority	$\pi_{1i} = \beta_{10} + u_{1i}$							
Priority – Targets Engaged.	Model 2							
	$\operatorname{HiPrior}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) + r_{ti}$	4 81	16	15	01	1 1 5	1 45	027
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	1.01	.10	•15	.01	1.13	1.43	.027
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							

 Table 16

 MRCM Parameter Estimates for Task Practice Behaviors (Hypothesis 11)

Table 16 (cont'd)

Dependent	Madal	Parameter estimates						
Variable	Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$
	Model 1							
	$TotEng_{ti} = \pi_{0i} + \pi_{1i}(Trial_{ti}) + r_{ti}$	10.6		28		4.64	3.63	.036
Total	$\pi_{0i} = \beta_{00} + u_{0i}$							
Number	$\pi_{1i} = \beta_{10} + u_{1i}$							
Targets	Model 2							
Engaged	$TotEng_{ti} = \pi_{0i} + \pi_{1i}(Trial_{ti}) + r_{ti}$	10.2	55	20	02	161	2 50	026
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	10.5	.55	.30	05	4.04	5.59	.030
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
Coefficient	estimates in bold are significant at $p <$	.05						

Significant at p < .10

appeared to be exerting equivalent levels of effort in their practice behaviors given that no significant mean or longitudinal differences in the total number of targets engaged were found. In short, stereotype threat and control condition females appeared to engage in highly similar practice activities relative to the advanced/strategic aspects of the task.

Self-regulation. Kraiger et al. (1993) propose that heightened metacognitive awareness is a hallmark of advanced cognitive learning. As a final investigation of female learners' strategic task learning then, results from the self-reported metacognitive activity measure assessed at the end of each day were analyzed (Table 17). Computation of the ICCs revealed that approximately 81% of the variance in mean metacognitive activity and 47% of the variance in change in metacognitive activity over time was attributable to between-person factors. Adding the experimental condition variable to the Level-2 equation revealed a significant main ( $\beta_{01} = -.20$ , p

< .05) and interaction effect ( $\beta_{11} = -.11, p < .05$ ) of stereotype threat on metacognitive activity. The direction of these coefficients indicated that females in the stereotype threat group tended to



Figure 13. Female's average task practice behaviors across learning trials



Figure 13 (cont'd).

Medal	Parameter estimates								
Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$		
Model 1									
$Metacog_{ti} = \pi_{0i} + \pi_{1i}(Day_{ti}) + r_{ti}$	2 77		01		075	22	066		
$\pi_{0i} = \beta_{00} + u_{0i}$	3.77		.01		.075	.32	.000		
$\pi_{1i} = \beta_{10} + u_{1i}$									
Model 2									
$Metacog_{ti} = \pi_{0i} + \pi_{1i}(Day_{ti}) + r_{ti}$	3 97	20	06	11	075	31	063		
$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	3.07	20	.00	-,11	.075	.51	.005		
$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$									
Model 2A	3.84	21	.02	08	.074	.27	.050		

Table 17MRCM Parameter Estimates for Female's Metacognitive Activity (Hypothesis 11)

Coefficient estimates in bold are significant at p < .05

report lower levels of metacognitive activity overall and that this average tended to decrease each day relative to control condition females (though the addition of the control variables negated this significant interaction). Thus, this pattern of results generally supports the predictions of Hypothesis 11 as well.

Summarizing across the analyses above, stereotype threat appeared to exert a demonstrative influence on female learners' knowledge acquisition strategies and the manner by which these individuals organized/focused their efforts during study time with the task manual. However, this apparent lack of engagement during the knowledge acquisition phase exerted little influence on the task practice behaviors of threatened females relative to their control counterparts. The experimental manipulation did appear to exert a negative influence on females' self-regulatory metacognitive activities though, which was compounded over time. In aggregate, the accumulated evidence was largely supportive of the predictions advanced in Hypothesis 11.

Decision-making strategy. Hypothesis 12 focused directly on the decision-making threat condition learned/relied on less optimal procedural decision strategies/heuristics for task completion than non-threatened women. To examine this hypothesis, a policy capturing approach (Aiman-Smith, Scullen, & Barr, 2002; Karren & Barringer, 2002) was employed. Policy capturing is a regression-based procedure which assesses how individuals or groups of individuals differentially weigh the importance of relevant informational cues when making an evaluation or decision. Within the organizational research literature, the methodology has been used to examine the extent to which individuals differentially value information about fit on perceptions of satisfaction (Kristof-Brown, Jansen, & Colbert, 2002), work behaviors on ratings of overall job performance (Rotundo & Sackett, 2002), and compensation packages on job pursuit intentions (Cable & Judge, 1994), among other applications. Policy capture studies require respondents to make a series of preference ratings/choices based on various combinations of decision-relevant information that are presented. For example, Kristof-Brown et al. (2002) asked participants to provide evaluations of perceived work satisfaction given information about the degree of person-job (PJ: low, medium, high), person-group (PG: low, medium, high), and person-organization (PO: low, medium, high) fit they would experience in a given organization. For that study, individuals provided ratings of perceived work satisfaction for 27 scenarios constructed by crossing all combinations of cues and cue values (3 cues with 3 values each).

Analytically, individuals' decisions/responses given a set of cues for a scenario are then regressed onto the specific cue values provided for that scenario; once aggregated across all decisions, regression coefficients for each informational cue are generated which reflect the relative importance of a particular piece of information/cue to an individual's choices. This information can be used in an idiographic manner to interpret the decision processes of one

person in particular or combined across people to draw nomothetic conclusions about individuals' general information processing tendencies (Aiman-Smith et al., 2002). Within the latter approach, cluster analytic techniques may be employed to empirically group individuals into categories of like decision-makers (e.g., managers who tend to favor information about task performance vs. counter-productivity in ratings of job performance, Rotundo & Sackett, 2002) or MRCM analyses can be used to test a priori hypotheses about the influence of between-person variables on differential cue weighting (Kristof-Brown et al., 2002).

The goal of Hypothesis 12 was most similar to the second of these approaches; that is, the primary prediction concerned whether the learning strategies of threatened versus non-threatened females led to differential weighting of the informational value/relevance of a target's Type, Class, and Intent (i.e., the informational sources/cues) in the critical Final Engagement decision for a target. Unlike most real-world decisions in which the correct decision is often unknown or ambiguous, an objectively correct engagement decision existed for every target which participants prosecuted in the task; consequently, each of the Type, Class, and Intent outcomes possessed an "optimal" cue weighting that indicated its diagnostic/informative value to a particular Final Engagement decision (similar to what is summarized in Table 8). These optimal weights could thus be extracted from the task and compared to the decision weights produced by learners. As a result, the focus for the present set of analyses was to examine whether A) the decision weights developed by female learners in the stereotype threat condition significantly differed from those in the control condition; and B) whether the decision weights developed by female learners in the stereotype threat condition were further from the optimal set of decision weights than their control condition counterparts. Note that to evaluate this hypothesis, only females' target/decision data from the three performance trials were used. Using data from only

the three performance trials was far more computationally tractable and therefore likely to lead to better model convergence and more accurate parameter estimates than using data from the 18 learning trials.

Because the dependent variable (Final Engagement decision) was a categorical variable with more than two levels (Clear, Warn, Mark), multinomial logistic regression was used for all analyses. In multinomial logistic regression, one sets a single level of the dependent variable to serve as the referent against which the other levels of the dependent variable are contrasted. Thus, k -1 regression equations are computed for the set of predictors, where k equals the number of categorical levels in the dependent variable. Given that there was no clearly logical/best choice among the Final Engagement categories to serve as the referent level, the decision was made to use the Clear category for these purposes. Consequently, two regression equations were modeled for each target: the first tested the likelihood that a given predictor was more strongly related to making a Warn as opposed to Clear decision, while the other tested the likelihood that a given predictor was more strongly related to making a Mark as opposed to Clear decision. Additionally, the categorical Type, Class, and Intent predictors necessitated the creation of dummy coded variables in order to be properly included in the regression model; k -1 dummy variables were therefore also needed for each categorical predictor. For these purposes, two dummy variables were created for the three-level Type (Air, Surface, Sub) variable using the Air subdecision outcome as the referent, while a single dummy variable was created for both the Class and Intent subdecisions in which Civilian and Peaceful subdecision values served as the referent categories, respectively.

Two separate policy capture analyses using multinomial logistic regression were required in order to evaluate the propositions stated above—one to extract the optimal decision weights

from the task and the other to estimate participants' observed decision weights. The data required to compute the first of these models were the objectively correct Type, Class, Intent, and Final Engagement decision for the targets constructed for the performance trials. Note that because the optimal weighting configuration for targets does not vary over time, it was not necessary to include the effect of time in the policy capture regression model. Subsequently, a simple singlelevel multinomial logistic regression analysis was performed on the data:

$$\log \left(\frac{P(\text{Final Engagement}_{\text{Warn}})}{P(\text{Final Engagement}_{\text{Clear}})}\right) = \beta_0 + \beta_1(\text{Type}_{\text{Surface}}) + \beta_2(\text{Type}_{\text{Sub}}) + (10)$$

$$\beta_3(\text{Class}_{\text{Military}}) + \beta_4(\text{Intent}_{\text{Hostile}}) + e_j$$

$$\log \left(\frac{P(\text{Final Engagement}_{\text{Mark}})}{P(\text{Final Engagement}_{\text{Clear}})}\right) = \beta_0 + \beta_1(\text{Type}_{\text{Surface}}) + \beta_2(\text{Type}_{\text{Sub}}) + \beta_3(\text{Class}_{\text{Military}}) + \beta_4(\text{Intent}_{\text{Hostile}}) + e_j$$

Logistic regression models will not converge or produce parameter estimates for data in which all observations of a particular predictor have the same outcome and/or can be perfectly classified into one category of the dependent variable (i.e., complete or quasi-separation exists in the data, Heinze & Schemper, 2002). This issue arises in the analysis of the optimal decision weights as a number of the predictors have a zero probability of being associated with a particular level of the dependent variable (e.g., Hostile targets are never Cleared, see Table 4, Table 8; as a result, perfect categorization/complete separation exists in the data for this variable). Consequently, Firth's bias reduction method was applied to the estimate of the multinomial logistic regression models in order to generate the optimal decision weights (Firth, 1993). The Firth corrected multinomial logistic models were run in R version 2.15 (R Development Core Team, 2012) using the *pmlr* package (Colby, Lee, Lewinger, & Bull, 2010).

With respect to the computation of the policy capture analyses using the observed data from participants, the structure of the data was such that multiple targets were nested within multiple days, which were nested in individuals. As a result, multinomial logistic MRCM was the preferred analytic approach for conducting the regression analyses. However, rather than attempt to fit a 3-level model to the data (which would substantially increase the computational intensity and complexity required for interpreting the decision weight estimates), the trial/time variable was collapsed into the Level-1 equation to create a 2-level model. This effectively removes estimation of the random effect for time from the MRCM equation and therefore does not allow for random variation in factors at the time-level to influence the relationship between information cue and engagement decision for individuals in the sample. Given that there was no reason to suspect that any one day/time point in the experiment was systematically different than any other day/time point in the study, this assumption seemed reasonable and, more importantly, unlikely to have a significant influence on the estimate of the decision weights. Thus, the final MRCM model was specified as follows (note that for ease of presentation, the separate logit link functions for the multinomial dependent variable categories are not presented):

Final Engagement<sub>*ti*</sub> = 
$$\pi_{0i} + \pi_{1i}$$
(Type<sub>Surface</sub>) +  $\pi_{2i}$ (Type<sub>Sub</sub>) + (4)  
 $\pi_{3i}$ (Class<sub>Military</sub>) +  $\pi_{4i}$ (Intent<sub>Hostile</sub>) +  $\pi_{5i}$ (Day) +  $\pi_{6i}$ (Day\*Type<sub>Surface</sub>) +  $\pi_{7i}$ (Day\*Type<sub>Sub</sub>) +  $\pi_{8i}$ (Day\*Class<sub>Military</sub>) +  $\pi_{9i}$ (Day\*Intent<sub>Hostile</sub>) +  $r_{ti}$ 

 $\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$ 

 $\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$ 

 $\pi_{2i} = \beta_{20} + \beta_{21}(\text{Condition}_i) + u_{2i}$ 

$$\pi_{3i} = \beta_{30} + \beta_{31}(\text{Condition}_i) + u_{3i}$$

$$\pi_{4i} = \beta_{40} + \beta_{41}(\text{Condition}_i) + u_{4i}$$
  

$$\pi_{5i} = \beta_{50} + \beta_{51}(\text{Condition}_i) + u_{5i}$$
  

$$\pi_{6i} = \beta_{60} + \beta_{61}(\text{Condition}_i) + u_{6i}$$
  

$$\pi_{7i} = \beta_{70} + \beta_{71}(\text{Condition}_i) + u_{7i}$$
  

$$\pi_{8i} = \beta_{80} + \beta_{81}(\text{Condition}_i) + u_{8i}$$
  

$$\pi_{9i} = \beta_{90} + \beta_{91}(\text{Condition}_i) + u_{9i}$$

Of greatest relevance to the present hypothesis, the fixed effect intercept terms  $\beta_{10}$ through  $\beta_{40}$  indicate the average decision weight observed in the entire female sample for each of the dummy-coded information variables collapsed across days, while the slope terms  $\beta_{11}$ through  $\beta_{41}$  reflect the extent to which the average decision weights differed for stereotype threat women relative to the control condition women. Similarly, the fixed effect intercept terms  $\beta_{60}$ through  $\beta_{90}$  indicate the average change in decision weighting across days observed in the entire female sample for each of the dummy-coded information variables, while the slope terms  $\beta_{61}$ through  $\beta_{91}$  reflect the extent to which changes in decision weighting over time differed for stereotype threat women relative to the control condition women. The multinomial logistic MRCM analyses were conducted using HLM version 6.08 (Scientific Software International, 2009).

Of final note, the Type, Class, and Intent outcomes used as the informational cues for the Final Engagement decision were also themselves decisions made by participants (as opposed to veridical information provided by the environment). Since participants were required to determine the correct Type, Class, and Intent for each target, it is likely that individuals may have based a number of their Final Engagement decisions on objectively "incorrect" information if they made the wrong classification for any of these subdecisions. While such errors would prevent an individual from making the actually correct Final Engagement decision for a target, they do not influence estimation of the decision weights in the policy capture analysis as these analyses simply examine the consistency with which a given piece of information is associated with a particular decision outcome. That is, the *accuracy* of individuals' Type, Class, and Intent classification decisions for any particular target is entirely independent of whether the participant made *optimal* decisions based on the information they had available.

For example, consider a target whose correct classification was Air, Civilian, Hostile, yet a participant classified the target as Surface, Military, Peaceful and subsequently chose to Clear the target (which is the correct Final Engagement decision for a Surface, Military, Peaceful target, see Table 8). In this case, the policy capture analyses would indicate that the person had made the optimal decision, even though the information used to make that decision—as well as the Final Engagement decision itself—was objectively inaccurate. In theory then, a participant could fail to correctly prosecute any targets in the game yet still learn to make optimal decisions so long as they made the correct engagement decision based on the information they had available. Consequently, the policy capture analyses do not provide insight into whether individuals or groups of individuals were more or less accurate in their decisions (which would influence performance), but rather whether they had learned to correctly interpret cue values. In short, participants could be optimal decision-makers without necessarily being accurate, but they could not be accurate without also being optimal.

Figures 14 through 21 summarize the results of the policy capture analyses. The y-axis for each graph reflects the decision weight associated with (i.e., the likelihood of selecting) either the Warn or Mark engagement decision relative to selecting the Clear engagement decision for each of the dummy-coded predictor variables at each day in the experiment. Thus, the graph in Figure 12 reflects the relative likelihood of Warning rather than Clearing a target if that target was a Surface as opposed to some other Type of vessel over time. Negative decision weights indicate that individuals were less likely to Warn/Mark rather than Clear targets given the informational cue, while positive values indicate that individuals were more likely to Warn/Mark rather than Clear targets given the informational cue. The graphs also depict the optimal weighting criteria for each of the Final Engagement by cue value combinations. Note that the direction and magnitude of the optimal decision weights perfectly mirrors the pattern of probabilities summarized in Table 8. For example, Table 8 indicated there was a higher probability that the correct Final Engagement decision for any given Surface target would be Clear (50%) as opposed to Warn (25%); this same pattern is reflected in the negative optimal decision weight coefficient ( $\beta$  = -4.65) shown in Figure 14. Of final note, for purposes of the present study, the actual numerical value of the decision weight carries no substantive meaning for evaluating Hypothesis 12. The interpretation of greater interest is the overall direction (i.e., positive or negative) and the relative differences in the decision weights produced by stereotype threat and control females to the optimal decision weights.

An examination of the main effects of stereotype threat on female learners' procedural decision-making effectiveness revealed that the only significant difference in the average decision weights across the groups was for the Surface cue; specifically, stereotype threat females tended to have more difficulty distinguishing whether to both Warn ( $\beta_{21} = .81$ , p < .05;



Weighting of Type Cue (Surface) on Warn Decision Relative to Clear Decision

*Figure 14*. Observed and optimal decision weights for Type cue (Surface) on decision to Warn rather than Clear targets for stereotype threat females and control females at each day



Weighting of Type Cue (Surface) on Mark Decision Relative to Clear Decision

*Figure 15*. Observed and optimal decision weights for Type cue (Surface) on decision to Mark rather than Clear targets for stereotype threat females and control females at each day



Weighting of Type cue (Sub) on Warn Decision Relative to Clear Decision

*Figure 16*. Observed and optimal decision weights for Type cue (Sub) on decision to Warn rather than Clear targets for stereotype threat females and control females at each day



Weighting of Type Cue (Sub) on Mark Decision Relative to Clear Decision

Figure 17. Observed and optimal decision weights for Type cue (Sub) on decision to Mark rather than Clear targets for stereotype threat females and control females at each day



Weighting of Class Cue (Military) on Warn Decision Relative to Clear Decision





Weighting of Class Cue (Military) on Mark Decision Relative to Clear Decision

*Figure 19.* Observed and optimal decision weights for Class cue (Military) on decision to Mark rather than Clear targets for stereotype threat females and control females at each day



Weighting of Intent Cue (Hostile) on Warn Decision Relative to Clear Decision





Weighting of Intent Cue (Hostile) on Mark Decision Relative to Clear Decision

*Figure 21*. Observed and optimal decision weights for Intent cue (Hostile) on decision to Mark rather than Clear targets for stereotype threat females and control females at each day

Figure 14) or Mark ( $\beta_{21} = .94$ , p < .05; Figure 15) Surface targets relative to Clearing those targets than control females. Marginally significant main effects were also observed for the decision weights involving whether to Warn versus Clear Military targets ( $\beta_{41} = .76$ , p = .09; Figure 18) and whether to Mark versus Clear Hostile targets ( $\beta_{51} = -1.65$ , p = .06; Figure 21); in both cases, stereotype threat women had more difficulty than control condition females reaching the appropriate decision. No significant differences in changes in decision weights over time were observed across the two experimental conditions, indicating that both groups of females were generally learning the procedural decision aspects of target engagement at the same rate.

Although the MRCM policy capture analyses provide insight into relative differences in learning between the two conditions, of perhaps greater interest is whether one group of learners was more effective at learning the optimal procedural decision making strategies than the other. As one might expect, all individuals appeared to become more optimal decision-makers over time; that is, the observed decision weights for all female learners—regardless of experimental condition—tended to become closer to optimal as more experience was accrued in the task. However, control condition females appeared closer to achieving the optimal decision weights than stereotype threat females for virtually all engagement decision and informational cue combinations. In fact, the overlap in the 95% confidence intervals between the observed and optimal decision weights at each time point indicated that control condition females had achieved near perfect optimality by Day 3 in their interpretation of a target's Intent and Class while the stereotype threat condition, female learners in the control condition nearly always selected the most probable Final Engagement decision associated with a given Intent or Class by the end

of the study. Furthermore, although they had not yet achieved optimality, the confidence intervals in the decision weights for stereotype threat and control condition females for the Type decisions on Day 3 did not overlap for three of the four observed decision weights, indicating that control condition females were generally making more probabilistically appropriate Final Engagement decisions based on the Type cues by this point as well. This pattern of results is remarkably similar to those described in the analysis for the knowledge structure clustering observed across conditions, thus lending a degree of support to the notion that control condition female learners appeared to be more proficient at learning efficient decision heuristics within the task than stereotype threat females.

In sum, although the significance tests for the main and interaction effects of stereotype threat observed in the estimated MRCM coefficients did not reveal dramatically different patterns of procedural decision-making strategy development between conditions of female learners, the comparison of participants' observed decision weights to the optimal decision weights inferred from the task revealed a number of insights. Of specific note, control condition females were generally closer to optimal throughout the entire experiment and even achieved optimality for certain informational cues by the end of the study. Thus, the overall patterns of the policy capture analyses generally support the predictions advanced in Hypothesis 12.

## **Task Performance**

Hypotheses 13 and 14 examined the impact of learning under conditions of stereotype threat on the demonstration of effective performance on the learned task. In the present study, individuals participated in both practice/learning trials plus a final performance trial at the end of each day. The final performance trials were different from the learning trials each day and were designed to be more challenging for all participants; however, they should have been particularly

more difficult for individuals who had not effectively learned the procedural and/or strategic aspects of the task. As mentioned in the description of the task environment, performance scores in TANDEM were a function of the number of targets engaged correctly, the number of targets engaged incorrectly, and the number of targets which crossed either of the two defensive perimeters; subsequently, these variables were analyzed with data from the performance trials and the learning/practice trials using the same MRCM equations outlined in Equations 1-3.

Table 18 present the results of the MRCM analyses for data from the learning/practice trials. With respect to the performance outcomes measured during the practice trials, no significant main effects or interactions over time were observed for the stereotype threat manipulation. On average, all female learners tended to engage approximately the same number of targets correctly each trial, with the number of correct engagements increasing by approximately one every three trials ( $\beta_{10} = .36, p < .05$ ). Interestingly, the increase in number of correct engagements did not necessarily correspond with a similarly sized decrease in the number of incorrect engagements per trial ( $\beta_{10} = -.07, p < .05$ ), indicating that the number of incorrect target engagements was relatively steady throughout the learning trials. Additionally, although very few of the four targets designed to cross the inner perimeter during the practice trials ever did so ( $\beta_{00} = -.50$ , p < .05), participants rarely engaged and/or improved their performance over time at engaging/prosecuting the seven targets designed to cross the outer perimeter during the practice trials ( $\beta_{00} = 6.58$ ,  $\beta_{10} = -.05$ , both ps < .05). Consequently, the relatively small increase in the number of points scored per trial ( $\beta_{10} = 53.7, p < .05$ ) was largely attributable to individuals' improvement in the procedural decision-making aspects of task performance (i.e., making correct engagement decisions) rather than improvements in strategic

Dependent	Madal			Param	eter es	timates		
Variable	Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$
	Model 1							
	$\operatorname{Crct}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) + r_{ti}$	176		24		2.51	5 1 2	056
Number Correct Engaged	$\pi_{0i} = \beta_{00} + u_{0i}$	4.70		.34		2.31	5.15	.050
	$\pi_{1i} = \beta_{10} + u_{1i}$							
	Model 2							
	$\operatorname{Crct}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) + r_{ti}$	4 85	- 17	36	- 04	2 51	5 17	056
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	<b>T.0</b> 5	1/	.50	04	2.31	5.17	.050
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
N7 1	Model 1					3.74	3.84	
	$\operatorname{Incrct}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) + r_{ti}$	5.16		07				046
	$\pi_{0i} = \beta_{00} + u_{0i}$			•••		5.71	5.01	.010
Number Incorrect	$\pi_{1i} = \beta_{10} + u_{1i}$							
Engaged	Model 2	A 93		07	.00	3.74	3.82	
	$\operatorname{Incrct}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Trial}_{ti}) + r_{ti}$		.47					.047
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$							,
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	Model 1							
	$\text{InnPen}_{ti} = \pi_{0i} + \pi_{1i}(\text{Trial}_{ti}) + r_{ti}$	.50		05	_	.785	.161	.002
Number	$\pi_{0i} = \beta_{00} + u_{0i}$							
Crossed	$\pi_{1i} = \beta_{10} + u_{1i}$							
Inner Perimeter	Model 2							
	$\text{InnPen}_{ti} = \pi_{0i} + \pi_{1i}(\text{Trial}_{ti}) + r_{ti}$	.56	12	05	.00	.785	.160	.002
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$		.12					
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							

Table 18MRCM Parameter Estimates for Performance Outcomes Measured duringLearning/Practice Trials (Hypothesis 13 & 14)

Table 18 (cont'd)

Dependent	Madal		Parameter estimates						
Variable	Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$	
	Model 1						.389		
	$\text{OutPen}_{ti} = \pi_{0i} + \pi_{1i}(\text{Trial}_{ti}) + r_{ti}$	6.56		06		.372		007	
Number	$\pi_{0i} = \beta_{00} + u_{0i}$	0.50		00				.007	
Crossed Outer Perimeter	$\pi_{1i} = \beta_{10} + u_{1i}$								
	Model 2								
	$\text{OutPen}_{ti} = \pi_{0i} + \pi_{1i}(\text{Trial}_{ti}) + r_{ti}$	6.58	03	05	01	.372	.391	007	
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$							.007	
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$								
	Model 1			50 1		0.2-4	1.8e5	1094	
	$Score_{ti} = \pi_{0i} + \pi_{1i}(Trial_{ti}) + r_{ti}$	746							
	$\pi_{0i} = \beta_{00} + u_{0i}$	-/40		52.1		9.264		1904	
Task	$\pi_{1i} = \beta_{10} + u_{1i}$								
Score	Model 2								
	$Score_{ti} = \pi_{0i} + \pi_{1i}(Trial_{ti}) + r_{ti}$	720	52.6	537	2 25	0 201	1 805	1008	
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	-/20	-32.0	55.7	-3.33	9.2e4	1.865	1990	
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$								

Coefficient estimates in bold are significant at p < .05

target selection.

Alternatively, results from the performance trial data (Table 19) revealed significant main effects of stereotype threat on the number of targets incorrectly engaged ( $\beta_{01} = 1.74, p < .05$ ) and marginally significant effects on the number of targets correctly engaged ( $\beta_{01} = -1.52, p = .06$ ) and total number of points scored ( $\beta_{01} = -305, p = .08$ ). The pattern of these results indicate that, on average, females who learned under conditions of stereotype threat tended to make more incorrect engagement decisions while making slightly fewer correct engagements than control condition females, leading to generally lower performance scores. Furthermore, although all

Dependent	Madal			Param	eter est	timates		
Variable	Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$
	Model 1							
	$\operatorname{Crct}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Day}_{ti}) + r_{ti}$	0.03		2 80		6.02	21.0	5 8/
	$\pi_{0i} = \beta_{00} + u_{0i}$	9.05	_	2.09	_	0.92	21.0	5.04
Number	$\pi_{1i} = \beta_{10} + u_{1i}$							
Engaged	Model 2A <sup>a</sup>							
	$\operatorname{Crct}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Day}_{ti}) + r_{ti}$	10.0	1.50	3.82	1.60	6 80	16.4	5 87
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	10.0	-1.52		-1.00	0.89	10.4	5.07
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	Model 1				·			
	$\text{Incrct}_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	9.18		_7 78		11 33	20.90	4 00
	$\pi_{0i} = \beta_{00} + u_{0i}$			-2.20		11.55	20.70	ч.00
Number	$\pi_{1i} = \beta_{10} + u_{1i}$							
Engaged	Model 2A <sup>a</sup>	Q 16			1 59	10.92	17.78	
	$\operatorname{Incrct}_{ti} = \pi_{0i} + \pi_{1i}(\operatorname{Day}_{ti}) + r_{ti}$		1 74	-3 10				3 49
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	0.10	1./ 4	-5.10	1.57	10.72		5.47
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	Model 1							
	$\text{InnPen}_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	94		- 19		1 1 9	76	135
Number	$\pi_{0i} = \beta_{00} + u_{0i}$	•/1		-,17		1.17	.70	.155
Crossed	$\pi_{1i} = \beta_{10} + u_{1i}$							
Inner Perimeter	Model 2							
Perimeter	$\text{InnPen}_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	1.06	- 25	10	18	1.19	.75	133
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	1.00	0					.155
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							

Table 19MRCM Parameter Estimates for Performance Outcomes Measured during PerformanceTrials (Hypothesis 13 & 14)

Table 19 (cont'd)

Dependent	Madal	F		Parameter estimates				
Variable	Moder	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$
	Model 1							
	$\text{OutPen}_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	0.00		63		1.13	1.81	017
Number	$\pi_{0i} = \beta_{00} + u_{0i}$	9.90		05				.947
Crossed Outer	$\pi_{1i} = \beta_{10} + u_{1i}$							
	Model 2							
Perimeter	$\text{OutPen}_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	10.0	02	.58	.10	1.13	1.83	954
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	10.0	.05					.934
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
	Model 1	1(52			_	2.9e5	9.4e5	
	$Score_{ti} = \pi_{0i} + \pi_{1i}(Day_{ti}) + r_{ti}$			642				2 4 . 5
	$\pi_{0i} = \beta_{00} + u_{0i}$	-1052						2.465
Task	$\pi_{1i} = \beta_{10} + u_{1i}$							
Score	Model 2A <sup>a</sup>							
	$\text{Score}_{ti} = \pi_{0i} + \pi_{1i} (\text{Day}_{ti}) + r_{ti}$	-1470	205 <sup>†</sup>	796	_273	2 8e5	7 5e5	2 5e5
	$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$	-14/0	-305	/90	-273	2.003	5 7.585	2.303
	$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$							
Coefficient	estimates in bold are significant at p	< .05						

<sup>†</sup>Signficant at p < .10

<sup>a</sup>Model includes control variables (coefficients not printed for ease of presentation)

learners tended to make more correct engagements ( $\beta_{10} = 3.82, p < .05$ ), fewer incorrect

engagements ( $\beta_{10} = -3.10$ , p < .05) and therefore score more points each performance trial ( $\beta_{11} =$ 

796, p < .05), these rates were significantly lower for the stereotype threatened female learners

 $(\beta_{11} = -1.60, p < .05 \text{ for correct engagements}; \beta_{11} = 1.59, p < .05 \text{ for incorrect engagements}; \beta_{11}$ 

= -273, p < .05 for task score). Thus, despite achieving approximately similar levels of

performance during the learning trials, stereotype threat learners were generally less effective on

the more complex performance trials (which is also consistent with previous research on stereotype threat effects and task difficulty, Nguyen & Ryan, 2008).

Lastly, an additional set of exploratory analyses were performed on the declarative knowledge tests assessed at the end of each day. As noted previously, many investigations in the research literature have used similar declarative knowledge tests to evaluate the presence of stereotype threat effects in a variety of contexts; consequently, such analyses serve as a useful point of comparison for characterizing the stereotype threat effects observed in the present study. Table 20 presents the results from the MRCM analyses for performance on the declarative knowledge tests. In general, all female learners improved their performance on the examinations slightly each day ( $\beta_{10} = .03, p < .05$ ) suggesting that individuals were improving in their basic understanding of the task with additional experience. Both the main effect of stereotype threat  $(\beta_{01} = -.04, p = .10)$  and the interactive effect of stereotype threat over time  $(\beta_{11} = -.03, p = .06)$ achieved marginal levels of significance, indicating a trend in which females in the experimental condition appeared to perform slightly worse on the tests on average and generally did not improve their test performance across days relative to control condition learners. In sum, evidence from the performance trial data and the declarative knowledge test generally appeared to support Hypotheses 13 and 14.

Table 20

Assessments	U					C				
	Parameter estimates									
Model	$\beta_{00}$	$\beta_{01}$	$\beta_{10}$	$\beta_{11}$	$\sigma^2$	τ <sub>00</sub>	$\tau_{11}$			
Model 1										
$\text{Test}_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$	73		02		017	021	002			
$\pi_{0i} = \beta_{00} + u_{0i}$	•/5		.02		.017	.021	.002			

11

.008

MRCM Parameter Estimates for Performance on the Declarative Knowledge

 $\pi_{1i} = \beta_{10} + \beta_{11}(\text{Condition}_i) + u_{1i}$ Coefficient estimates in bold are significant at p < .05

 $\pi_{1i} = \beta_{10} + u_{1i}$ Model 2

 $\text{Test}_{ti} = \pi_{0i} + \pi_{1i}(\text{Day}_{ti}) + r_{ti}$ 

 $\pi_{0i} = \beta_{00} + \beta_{01}(\text{Condition}_i) + u_{0i}$ 

<sup>a</sup>Model includes control variables (coefficients not printed for ease of presentation)

**.76**  $-.04^{\dagger}$  **.03**  $-.03^{\dagger}$  .026 .070

## DISCUSSION

Research on stereotype threat theory and its effects has a rich investigative history, spanning numerous applications, content domains, and psychological/educational disciplines (cf., Nguyen & Ryan, 2008). A primary goal of the present investigation was to extend work in this area of study to the acquisition and development of task-relevant knowledge by individuals facing conditions of stereotype threat during learning activities. Initial research (Rydell, Rydell, & Boucher, 2010; Rydell, Shiffrin, et al., 2010) suggested that the presence of negative stereotypes relevant to a content domain was capable of adversely influencing a targeted individual's ability to successfully learn and later recall/demonstrate learned task-relevant skills. The current study built upon this proof of concept in a number of ways. First, the incorporation of Kraiger et al.'s (1993) empirically supported and widely cited taxonomy of learning outcomes provided a strong theoretical foundation from which to approach and explore stereotype threat effects during learning. Second, the use of multiple metrics/measurement techniques each tapping substantively different activities and consequences relevant to the knowledge acquisition process permitted a detailed interpretation of the specific learning outcomes affected by stereotype threat. Lastly, the longitudinal design of this investigation enabled examination of the way in which expertise and the knowledge acquired by groups of threatened versus nonthreatened individuals developed over time—a crucial consideration when studying the inherently dynamic process of learning.

Table 21 provides a summary of the research hypotheses and the overall result of their accompanying analytic tests. Analyses of the characteristic "shape" and descriptive nature of participants' knowledge structures, behaviors/cognitions related to efficient and effective task strategy development, and task-specific performance outcomes each revealed unique insights

Table 21 *Hypothesis Summary* 

Hypotheses	Result
<u>Hypothesis 1</u> : The knowledge structures of females who learn under conditions of stereotype threat will be less similar to those from top performers/men than the knowledge structures of females who learn under control conditions.	Not Supported
<u>Hypothesis 2</u> : The knowledge structures of females who learn under conditions of stereotype threat will be less correlated with those from top performers/men than the knowledge structures of females who learn under control conditions.	Not Supported
<u>Hypothesis 3</u> : The knowledge structures of females who learn under conditions of stereotype threat will be less coherent than the knowledge structures of females who learn under control conditions.	Not Supported
<u>Hypothesis 4</u> : The knowledge structures of females who learn under conditions of stereotype threat will have significantly more links (i.e., be less parsimonious) than the knowledge structures of females who learn under control conditions.	Not Supported
<u>Hypothesis 5</u> : The clustering of concepts in the knowledge structures of females who learn under conditions of stereotype threat will be significantly different than that for non-threatened women.	Not Supported
<u>Hypothesis 6</u> : The similarity between the knowledge structures of females who learn under conditions of stereotype threat with those from top performers/men will improve at a slower rate compared to females who learn under control conditions.	Not Supported
<u>Hypothesis 7</u> : The correlation between the knowledge structures of females who learn under conditions of stereotype threat with those from top performers/men will improve at a slower rate compared to females who learn under control conditions.	Not Supported
<u>Hypothesis 8</u> : The coherence of the knowledge structures of females who learn under conditions of stereotype threat will improve at a slower rate compared to females who learn under control conditions.	Not Supported
Table 21 (cont'd)

Hypotheses	Result
<u>Hypothesis 9</u> : The number of links in the knowledge structures of females who learn under conditions of stereotype threat will increase at a faster rate (i.e., structures will become less parsimonious) compared to females who learn under control conditions.	Not Supported
<u>Hypothesis 10</u> : The knowledge structures of females who learn under conditions of stereotype threat will demonstrate less integration of related task concepts over time (i.e., fewer and less efficient associations between related task concepts) compared to females who learn under control conditions.	Supported
<u>Hypothesis 11</u> : Females who learn under conditions of stereotype threat condition will exhibit poorer/more basic cognitive task strategies than females who learn under control conditions.	Supported
<u>Hypothesis 12</u> : Females who learn under conditions of stereotype threat will develop less optimal procedural decision strategies for task completion than females who learn under control conditions.	Supported
<u>Hypothesis 13</u> : Females who learn under conditions of stereotype threat will demonstrate worse performance on the learned task than females who learn under control conditions.	Supported
<u>Hypothesis 14</u> : Females who learn under conditions of stereotype threat will improve their performance on the learned task at a slower rate than females who learn under control conditions.	Supported

into stereotype threat's effects on the learning process. Consequently, the discussion of the results below will begin with a brief overview of the rationale and findings from the present study which attempts to integrate the observed pattern of results into a coherent picture of stereotype threat effects during learning. Next, more specific comments/treatments for each of the primary dependent variables (knowledge structures, task strategy, and performance) are explored. Lastly, this section concludes with implications regarding the influence of stereotype

threat during the learning process, suggestions for future research within the domain, and study limitations.

#### **Summary of Key Findings**

Identifying the psychological mechanisms and outcomes associated with the learning process which help characterize an individual's transition from novice to expert within a domain/task has been of significant interest to researchers in a variety of disciplines. A general consensus from this broad stream of research is that the act of learning typically begins with the acquisition of basic declarative and operational facts/statements, advances to the generation of generalizable procedural rules or definitions that help organize task-relevant activities/cognitions, and culminates in the ongoing development of conditional principles and strategies which improve the efficiency/effectiveness with which an individual is able to access and/or apply knowledge towards the completion of task-relevant objectives (Anderson, 1982, 1996; Gagne, 1984; Kanfer & Ackerman, 1989). The learning outcomes associated with these latter two activities are of particular interest as they are often key indicators of the degree of proficiency one has developed within a task/domain comprehension; that is, the factors which appear to most directly distinguish a novice from a learned expert is the possession of well-organized, efficient knowledge schemata that facilitate problem representation and minimize the amount of effortful processing required to derive/coordinate task-relevant strategies and solutions (e.g., Anderson, 1993b; Chase & Simon, 1973; Kahneman & Frederick, 2005). Notably, the production of both of these advanced learning outcomes relies heavily on working memory capacity as it necessitates the ability to perceive and hold information in active awareness while simultaneously integrating that input with either existing knowledge or additional information from the environment (cf., Hunt, 1994). Consequently, factors which affect working memory capacity—such as stereotype

threat (Schmader et al., 2008)—should have a noticeable impact on the development of efficient/effective knowledge structures as well as related task strategies/heuristics (Kraiger et al., 1993).

The results of the present study were fully consistent with this conclusion. Females learning a complex decision-making task under conditions of stereotype threat appeared to develop knowledge structures whose schematic relations were less efficiently organized than females who did not experience such adverse conditions over the course of the three experimental sessions. The policy capturing analyses used to decompose the observed procedural decision-making strategies developed by individuals within these experimental groups provided further evidence of significant differences in learning proficiency as a result of stereotype threat. Specifically, this set of findings revealed that control condition females were generally more effective at and achieved optimal levels of proficiency more quickly when interpreting and applying learned information in order to make critical task decisions than stereotype threatened females. As expected, differences in the manner by which these concepts were learned ultimately manifested in differences in objective indicators of task performance as well. In sum, the observed pattern of evidence supports what would be anticipated if the working memory capacity of individuals was negatively influenced by stereotype threat within the learning environment—affected participants appear to have greater difficulty integrating task information into procedurally useful rules and strategically efficient heuristics, which ultimately influences their ability to operate as effectively within the performance domain.

## Stereotype Threat Effects on Knowledge Organization

The organization of information into well-constructed procedural knowledge relations is among the most critically important outcomes of the learning process (Anderson, 1993b, 1996;

Anderson et al., 2004; Kraiger et al., 1993; Rouse & Morris, 1986). A whole host of cognitive machinery influences learners' ability to transform their basic comprehension of a task space into meaningful inferences about the interdependencies among components, events, rules, actions, etc. relevant to producing task-specific outcomes/results. Among these mechanisms, working memory plays a significant role by enabling individuals to simultaneously coordinate, interpret, and integrate multiple sources of information into contextually informative knowledge capable of directing task behaviors (e.g., Cantor & Engle, 1993; Engle, 2002; Johnson-Laird, 1983; Kane & Engle, 2003). Consequently, the present study hypothesized that stereotype threat's empirically documented deleterious influence on working memory efficiency (i.e., Beilock et al., 2006; Beilock et al., 2007; Schmader, 2010; Schmader et al., 2009; Schmader & Johns, 2003) would negatively influence knowledge organization by impairing threatened individuals' ability to adequately maintain task-relevant information in an activated/accessible state needed to develop an integrated mental representation of the content.

Hypotheses 1-10 in the present study examined this general prediction for female learners under conditions of stereotype threat using a variety of measurement techniques intended to characterize a knowledge structure's representational form. Perhaps the most readily obvious observation from this set of hypotheses was the lack of significant differences between females who learned under stereotype threat conditions and control condition females using the quantitative knowledge structure metrics (i.e., similarity, correlation, coherence, number of links, and the various descriptive metrics from the exploratory graph theory analyses) versus the qualitative analyses. Although there may be many explanations for why this pattern of nonsignificant findings was observed, two interpretations seem particularly plausible. First, the specific set of knowledge concepts used in the present study (Table 7) may have limited the

degree of variability in the possible linkages among concepts. As Goldsmith et al. (1991) note, the number and content of concepts used when eliciting knowledge structures holds significant influence on the composition/interpretation of individuals' observed knowledge organization. In the present study, the use of relatively homogenous knowledge concept subsets (i.e., decisionmaking and procedural/strategy concepts) may have contributed to less variability in the linkages likely to emerge in a given knowledge structure—which is precisely what the quantitative metrics are designed to capture. Furthermore, unlike previous studies which examined knowledge structures using the TANDEM task paradigm, the majority of concepts employed in this investigation were selected so as to draw inferences about how individuals integrated taskcritical information in a manner meaningful to decision-making rather than to examine differences in target selection and/or task operations. Although analyses of the task practice behaviors did not suggest that female learners in either condition approached these task activities in significantly different manners, subtle differences in such "gameplay" activities could have differed enough to be detected by quantitative assessments of knowledge structures had more operational task concepts been used. As it stands though, the relatively more structured/patterned nature of the knowledge concepts implemented in the present study may have been too rigid to generate substantially large differences in the quantitative knowledge structure metrics across female learners.

A second possible interpretation is simply that the quantitative indices may not have been particularly informative metrics given that stereotype threat was expected to primarily influence the *effectiveness/efficiency* of participants' organization of task-relevant knowledge concepts. Many studies which assess learners' knowledge structures do so for the explicit purpose of evaluating the degree of relatedness between novice and expert mental models (e.g., Chi, Glaser,

& Farr, 1988; Day et al., 2001; Ford & Kraiger, 1995) and/or to use the metrics as predictors of some task-relevant performance outcome (e.g., Dorsey et al., 1999; Kozlowski, Gully, et al., 2001; Schuelke et al., 2009). In such cases, the underlying theoretical assumption is generally that the pattern of relations among network concepts carries some "meaning" about the way in which learners have made sense of a given domain space, though deducing that particular meaning is not of central importance or is too speculative in nature. As a result, the use of quantitative indices in these investigations serve as convenient indicators of knowledge structure quality, and conclusions about what individuals have learned and/or whether the meaning of that organization is logical/interpretable are largely ignored. The present study, on the other hand, specifically proposed that impediments to working memory elicited by stereotype threat would make it more difficult for learners facing such conditions to integrate and store task-critical information in a manner that promoted efficient and effective task decisions and strategies. The specific relations which emerged among concepts in an individual's knowledge structure and the meaning that those relations implied were therefore of critical concern and expected to be among the most sensitive to stereotype threat. In fact, differences in the characteristic shape/make-up of a network (i.e., what is measured by the quantitative knowledge structure indices) would likely only manifest in situations where a large or severely disruptive situational factor was present—a relatively rare finding in the area of stereotype threat research (Steele & Aronson, 1995; Nguyen & Ryan, 2008).

Consequently, the qualitative interpretations of the knowledge structures pursued in Hypotheses 5 and 10 arguably stood to offer the most significant insights into the effects of stereotype threat during learning. To this end, the unique choices made regarding the design elements of the TANDEM architecture in the present study relative to previous administrations

of the task (e.g., consideration of multiple subdecision outcomes to make a Final Engagement decision, creating variation in the relative probabilities between subdecision and Final Engagement outcomes, etc.) made it possible to consider a variety of probable structural patterns a priori and then examine whether such relations emerged across groups of learners. As described in the Methods and Results sections, examinations of between-group knowledge structure differences in functional versus feature relations as well as relations indicative of efficient heuristic reasoning were of particular interest. Interestingly, the average knowledge structures of stereotype threatened and non-stereotype threatened females revealed that individuals in both groups tended to exhibit functional network relations among the decisionmaking knowledge concepts consistent with the rules/parameters of the task environment outlined in Tables 4 and 8. However, when changes in the clustering patterns over time were examined, the average knowledge structure of stereotype threat learners appeared to differ noticeably from that of control condition learners. More specifically, by the end of Day 3 in the study, the knowledge structures of females in the control condition appeared to be organized in a manner consistent with a highly efficient heuristic useful for making the correct Final Engagement decision quickly and relatively accurately based on the two most diagnostic pieces of information (a target's Class and Intent). Alternatively, females in the stereotype threat condition had not extracted this heuristic; their organization of the decision-making concepts seemed more consistent with individuals attempting to memorize all possible outcome permutations rather than the most likely outcomes based on a given piece of information.

One useful way in which to characterize this difference between the observed patterns of knowledge structures can be gleaned from the works of Simon (1956, 1990) and later Gigerenzer and colleagues (e.g., Gigerenzer 1991, 1993; Gigerenzer, Todd, & the ABC Research Group,

1999; Todd & Gigerenzer, 2007) on bounded rationality and the manner by which heuristics are developed to serve the inferential needs of human decision-makers. In brief, the theory of bounded rationality proposes that because of limitations in information processing capabilities and environmental conditions which impose restrictions on the availability of information, intelligent systems (i.e., humans) make use of approximate computational/judgment processes for many decisions. Such heuristic approximations are generally satisficing in that they stipulate a decision process should end when a reasonable, though not necessarily perfect, conclusion is reached (Simon, 1957). Building upon this notion, Gigerenzer's "adaptive toolbox" indicates that individuals develop such satisficing heuristics within a given task domain over time based on their past experiences and the situational influences/demands of the environment in which they operate (cf., Gigerenzer & Selten, 2001). Consequently, given the demands of the TANDEM task environment and the performance objectives imposed on individuals, both bounded rationality and the adaptive toolbox would suggest that a critical indicator of effective learning in the present study is the acquisition of efficient decision heuristics which minimize the amount of information processing needed in order to make a reasonably accurate Final Engagement decision for a target.

Thus, although all female learners drew accurate functional relations among knowledge structure concepts, the introduction of the stereotype threat manipulation appeared to impair the ability of females in the stereotype threat condition to organize these relations into a pattern consistent with an efficient satisficing heuristic. An explanation derived from the principles of the adaptive toolbox would suggest that this discrepancy might be attributable to differences in individuals' representation of the environmental/situational demands in the stereotype threat versus control conditions. In their original conceptualization of the effect, Steele and Aronson

(1995) postulated that a key motivation for individuals facing stereotype threat is to avoid confirming the validity of the negative stereotype through their actions. In the language of the adaptive toolbox then, individuals facing stereotype threat would perceive this need as an added requirement in their operational environment while those free from such pressures would not. One way in which a person might meet this perceived environmental demand in a learning context is by attempting to perfectly memorize/acquire all information about a particular task and its components, as this would seem to ensure that the individual would not be deficient in the *amount* of knowledge they possess about the problem space. However, the rationale underlying bounded rationality and the adaptive toolbox indicates that this strategy is likely to be less effective/efficient at improving an individual's ability to apply that knowledge to domainrelevant problems. Consequently, the organization of knowledge concepts into schemata conducive to simple yet effective heuristics-a hallmark of expert cognition within a domain (e.g., Chase & Simon, 1973; Lipschitz, Levy, & Orchen, 2006)-may run counter to the environmental demands perceived by learners facing conditions of stereotype threat and therefore be less likely to emerge.

In sum, differences observed in the longitudinal comparisons of female participants' knowledge structures across conditions of stereotype threat appear to be relatively consistent with broader theories of cognition depicting the manner by which individuals develop adaptive heuristic reasoning. The reported findings generally support the notion that stereotype threat creates an added environmental pressure for learners (Steele & Aronson, 1995), which, in a learning context, appears to impede the acquisition/development of efficient heuristics conducive to task performance. The conclusions of the present study are similar to those advanced by Rydell, Shiffrin, et al. (2011), who postulated that females experiencing stereotype threat were

also less proficient at acquiring effective heuristics for completing a visuospatial task. However, a significant contribution of the current investigation was the extension of this result to the domain of cognitive-based/decision-making heuristics as well as the ability to empirically verify this supposition through the use of knowledge organization data. Furthermore, the relatively rare use of a repeated measures experimental design to examine the influence of stereotype threat over time and knowledge structure development (Ifenthaler et al., 2011; Ifenthaler & Seel, 2005) provided further insight into the dynamical nature of the effect that would have otherwise been unobserved had a cross-sectional, single observation design been implemented.

## **Stereotype Threat Effects on Cognitive Strategy Acquisition**

As implied in the rationale of the previous section, the effective organization of knowledge into meaningful procedural relations and the development/refinement of cognitive strategies are closely interrelated processes. Knowledge structures provide the mental framework for interpreting causal relations within a task domain which can then be used to systematically direct and coordinate attention and other cognitive resources towards effective task completion (Chi et al., 1981; Chi et al., 1982; Simon & Simon, 1978). In the present study, a variety of indicators were examined in an attempt to capture potential differences in aspects of cognitive strategy acquisition between female learners in the stereotype threat and control conditions. Specifically, the manner by which individuals structured learning activities related to declarative knowledge acquisition and active task practice as well as the extent to which participants engaged in self-regulatory metacognitive activity were each examined. In addition to these strategic learning behaviors, analyses were also conducted to extract and quantify differences in the decision-making heuristics implied by the composition of the average knowledge structures observed for stereotype threat versus control condition learners.

With respect to strategic learning behaviors, the overall findings from this set of analyses indicated that although female learners in both conditions were generally similar in their task practice behaviors, stereotype threat learners tended to exhibit poorer declarative knowledge acquisition behaviors by spending less time studying the most task-critical information available in the manual while simultaneously reporting lower levels of metacognitive awareness related to these learning activities. At a broader level, this pattern of results is intriguing in light of previous research that has advanced decrements to motivation/task engagement as the key mechanism accounting for stereotype threat effects, as opposed to the working memory hypothesis upon which the present study was based (e.g., Crocker et al., 1998; Grimm et al., 2009; Major et al., 1998; Marx & Stapel, 2006a, 2006b; Pronin et al., 2004; Wheeler & Petty, 2001). To the extent the former explanation holds true, noticeable discrepancies in the amount of effort and/or time spent engaged in learning-related activities by individuals facing stereotype threat would be expected; furthermore, the effects of disengagement on knowledge acquisition would likely also be greatly exaggerated in exploratory learning environments in which the individual has near complete control over what, when, and how much to learn.

In the present study, observed differences in the amount of time stereotype threatened females spent studying the task manual relative to control condition learners as well as mean differences in levels of metacognitive awareness could both be interpreted as supportive of a (de)motivational or disengagement hypothesis. However, the lack of significant differences observed in participants' task practice behaviors conflicts with this interpretation as it indicates that learners in both groups were exerting approximately the same degree of effort during the hands-on learning periods. Additionally, the apparent disengagement from task manual study only occurred during the final experimental session and was not accompanied with a decrease in

the validity of the concept relationships observed in the Day 3 knowledge structures that might be expected if stereotype threatened learners had simply abandoned their attempts to appropriately learn decision cue-outcome relationships. In fact, the number of functional relationships observed among decision-making concepts actually *increased* between Days 2 and 3 for stereotype threat learners, suggesting that, if anything, stereotype threatened females had become more knowledgeable of the rules of engagement between these time points. Thus, although decreased engagement during certain phases of declarative knowledge acquisition and diminished levels of metacognitive awareness were observed in the present study, the temporal pattern of these variables in conjunction with the demonstration of equally effortful levels of task practice behavior were not fully consistent with the motivational hypothesis for stereotype threat effects.

Although speculative in nature, at least two alternative explanations more consistent with the cognitive/working memory hypothesis for stereotype threat effects could instead account for these findings. First, the observed disengagement from task manual study may have occurred because stereotype threat participants felt that they had memorized nearly all of the critical information and thus their time was better spent learning to recognize/apply this information during the hands-on practice sessions rather than engaging in continued study. As summarized in the previous section, the saliency of the stereotype threat manipulation may have stimulated participants to completely memorize the decision-making cues so as to avoid confirming the validity of the negative stereotype. Such a reaction could have subsequently inhibited recognition that memorization was simply the first—but not only—step in learning how to most efficiently/effectively engage targets in the task. This might also partially help explain why the metacognitive awareness of this group was slightly lower as well; if individuals were employing

a portion of their self-regulatory capabilities to continually monitor whether they were confirming the negative stereotype as self-characteristic (which seems plausible given the significant between-group differences observed in the perceived stereotype threat measure), they would likely have had fewer cognitive resources to direct towards identifying/correcting changes in task/learning behaviors. Alternatively, a second possible account for the observed findings could be that the apparent decrease in study time was indicative of stereotype learners changing their strategic learning approach by more narrowly focusing their attention on a smaller subset of topics during later stages of learning (e.g., focusing on relations among only a single set of subdecision cues and outcomes, focusing only on how to monitor defensive perimeters, etc.). This possibility would suggest that stereotype threatened learners did recognize the need to improve their task effectiveness; however, their solution for doing so involved attempting to memorize a smaller range of topics rather than gleaning more broadly applicable heuristicsessentially forsaking the forest for the trees. Unfortunately, the present research is unable to empirically contrast the validity of these possible differences in strategic learning approaches adopted by stereotype threatened learners, marking a significant opportunity for future research in this topic area.

While the results of the strategic learning behavior analyses outlined above were less easily interpretable with respect to stereotype threat's influence on learning outcomes, results from the decision-making strategy analyses were relatively clearer. Of greatest interest, the policy capture findings examining whether female learners across experimental conditions had learned to effectively interpret information to make task-critical decisions nicely complemented the qualitative conclusions reached from examination of participants' averaged knowledge structures. More specifically, the observation that only control condition learners appeared to be

acquiring efficient satisficing heuristics relevant to target engagement on the basis of the knowledge organization data was also borne out in examinations of differences in the decision weights used to interpret task information when making critical decisions employed by female participants across experimental conditions. Interestingly, the results of the MRCM analyses on the development of these decision weights revealed few differences in either average values or changes in these values over time between stereotype threat and control condition learners. Comparison of these values from each group against the optimally derived decision weights, however, unequivocally revealed that control condition participants tended to be substantially better at applying knowledge about cue validity to identify the most probable decision outcome.

The rationale underlying the results of the decision weight analyses are strongly reminiscent of the theoretical processes encapsulated in Brunswik's Lens Model of human perception and decision-making (cf., Brunswik, 1952). In brief, the Lens Model provides a conceptual framework for describing the manner by which the objective diagnostic values of informational cues in an environment are replicated in the judgments of human decision-makers. Brunswik postulated that all decisions are informed by a variety of different cues specific to the criteria's environment. For example, if one were interested in examining judgments about tenure promotion decisions, a candidate's publication history, years of experience, service to the department, etc. might all be considered cues that hold some informative value in relation to the decision criteria of interest. In the abstract, each of these cues possesses some absolute degree of *ecological validity* which characterizes its overall importance, relevance, or value to the final criteria within the decision-making environment. However, when an individual attempts to use these same cues to reach a judgment about the criteria of interest, imperfections in the human reasoning process and/or extraneous situational factors can lead to variability in *cue utilization* 

that influences perceptions of the informative/diagnostic value of the various cues. Returning to the example above, individuals on a tenure review committee may tend to over- or under-value the relevance of total number of publications in their tenure decisions due to a variety of past experiences which have shaped their belief in the informative value of this cue in making tenure promotion decisions. As a result, a failure to reproduce the most ecologically valid tenure decision based on information about number of publications could occur if the consistency between observed and objective cue values differs greatly.

One significant implication of the Brunswikian Lens Model is that, because the ecological validity of a cue represents its maximal utility for decision-making under a given set of circumstances, individuals who are better at reproducing the ecological validity of various decision-relevant cues in their own judgments should be more effective decision-makers within the context of the applicable task environment. In the present study, the ecological validity of the environmental cues were given by the optimal regression weights computed from the policy capture analyses and reflect the objective likelihood of a particular Final Engagement decision given a particular subdecision outcome, while estimates of cue utilization were provided by the policy capture analyses performed on participants' observed engagement decisions. The systematic underweighting of cue values observed by female learners in the stereotype threat conditions therefore indicates that the salience of stereotype threat during learning activities significantly impaired individuals' ability to learn how to appropriately weigh information from the environment to make relevant task decisions. Furthermore, trends in the data also suggested that stereotype threat learners were less proficient at calibrating their cue weights towards the ecologically valid cue values over time in relation to control condition learners.

Overall then, the findings from the complete set of analyses pertaining to cognitive strategy development appeared most consistent with the hypothesis that stereotype threat primarily impairs cognitive processes (i.e., working memory) as opposed to motivational processes during learning activities. Particularly relevant to exploratory learning contexts, female learners facing stereotype threat seemed less proficient at effectively structuring their declarative knowledge acquisition activities, which may have also been influenced by poorer self-regulation of strategic learning behaviors. Lastly, the added situational pressures introduced by stereotype threat also appeared to demonstrably impede the development of decision-making heuristics which capitalized on the most ecologically valid cue-outcome relations in the task environment, resulting in stereotype threat learners generally exhibiting suboptimal decision strategies.

### **Stereotype Threat Effects on Task Performance**

As noted in the introductory paragraphs of this manuscript, the overwhelming majority of investigations on stereotype threat have examined effects on task performance during which it was made clear that participants' activities would be evaluated for diagnostic or comparative purposes (e.g., Nguyen & Ryan, 2008). In the present study, substantial efforts were made to distinguish learning and performance outcome variables to permit separate analysis of these components in subsequent hypothesis testing. Note that it was never a goal of this investigation to directly assess whether learning outcomes significantly impacted performance outcomes; this result has been demonstrated numerous times across numerous contexts (even within the TANDEM task environment, see Bell, 2002 and Kozlowski, Gully, et al., 2001) and is thus less substantively interesting. Nevertheless, examination of the direction and magnitude of the zero-order correlations presented in Table 9 suggests that many of the learning outcome measures (e.g., the quantitative knowledge structure metrics and cognitive strategy indicators) were related

to performance outcomes in the expected manner. Consequently, it is a virtual certainty that stereotype threat's impact on task performance measures would be mediated by outcomes associated with the quality of task-relevant learning activities.

However, the lack of significant mean or longitudinal differences in the objective performance indicators across conditions of stereotype threat during the learning/practice trials somewhat conflicts with this proposition-especially considering that such effects were present during the final performance rounds. Given that control condition females appeared to possess greater comprehension of the task domain than stereotype threatened females, one might have expected control condition females to have achieved better performance during the practice trials as well. One explanation for this pattern of effects may rest with differences in the instructional primes elicited by the exploratory learning recommendations versus those introduced during the performance trials. Consistent with the tenets of guided active learning interventions (e.g., Mayer, 2004), the practice trials were framed as opportunities to explore and experiment within the task space to identify procedures and strategies which made sense to the learner. Additionally, the provided learning recommendations offered some minimal guidance regarding how to orient one's learning activities, but did not provide strict instruction regarding how best to satisfy the task objectives or solve relevant problems within the environment. Lastly, although feedback was provided to participants regarding their performance following each learning trial, the feedback instructions notified individuals that the purpose of this information was to help identify areas where one might wish to focus greater attention during subsequent learning activities. The end result of this instructional design element is to deemphasize efforts to reach high levels of performance achievement for all learners during practice trials and instead encourages learners to engage in activities with promote task comprehension/mastery. It is

therefore common for improvements in training performance during exploratory learning interventions to be less drastic relative to other training designs (Bell, 2002).

Of additional significance, the practice trials for each day always presented the exact same targets in the exact same configuration. As a result, participants could begin to memorize, or at least develop a greater awareness of, the correct engagement decisions for targets presented during the learning trials. Given that there was some evidence that stereotype threatened learners may have been more prone to a learning style based on rote memorization rather than the development of efficient heuristics, this strategy may have permitted these individuals to achieve performance levels during the practice trials equivalent to those produced by control condition learners despite differences in overall task comprehension. However, the effectiveness of this strategy is negated when changes in task demands and/or complexity are introduced which require reliance on adaptive expertise and generalized inferences to achieve task performance (Bell & Kozlowski, 2002, 2008; Ford et al., 1998; Ivancic & Hesketh, 2000), a finding which is also consistent with the need for tasks to be sufficiently difficult in order to elicit stereotype threat effects (cf., Nguyen & Ryan, 2008). Given that such learning outcomes appeared to have been more difficult for individuals targeted by stereotype threat to develop, the observed disparity in performance during the final performance rounds for stereotype threatened females was consistent with the notion that poorer task comprehension relates to lower levels of task achievement.

A final consideration related to differences in the task performance outcomes observed in the present study concerns the significant interaction effect between time and stereotype threat reported in Table 19. The direction of this finding indicated that stereotype threat effects during learning appeared to compound the performance difficulties faced by stereotype threatened

individuals over time. Specifically, the results of these analyses revealed that individuals who did not face the threat of confirming the validity of a negative stereotype were not only better performers on average than those who were faced with the stereotype, but these individuals also demonstrated greater improvements in performance after additional task practice. This finding clearly suggests that the learning activities of stereotype threatened individuals are not simply less proficient overall; they also place such individuals at a distinct disadvantage relative to other learners that cannot be overcome simply by further exposure to the task environment.

## **Implications and Directions for Future Research**

Since its inception, stereotype threat theory has generated the greatest amount of interest in the domain of educational testing and personal selection/assessment. The contention that stereotype threat effects might contribute to the prevalence of adverse impact and/or exacerbate subgroup performance differences has stimulated significant debate within the research community regarding the extent to which the phenomenon is truly a legitimate concern in high stakes testing (e.g., Cullen et al., 2004; Cullen et al., 2006; Good et al., 2003; Sackett et al., 2004; Sackett & Ryan, 2012; Schmidt, 2002; Steele & Davies, 2003; Stricker & Ward, 2004). While the present results do not speak directly to whether stereotype threat influences performance assessments in applied settings, the findings summarized herein (as well as those reported by Rydell colleagues) demonstrate that stereotype threat can negatively influence the knowledge acquisition process under certain conditions-which most certainly has the potential to manifest as subsequent performance disparities. Interestingly, one reason frequently cited for why stereotype threat effects are unlikely to manifest in high-stakes testing environments is that such contexts tend to be heavily governed by organizational/legal regulations (e.g., Sackett et al., 2001). Training and learning environments, on the other hand, tend to be far less standardized; in

fact, a large proportion of job/task learning tends to occur either informally or through other "on the job" sources that often fall outside the purview of organizational control (Loewenstein & Speltzer, 2000). As a result, the likelihood of conditions conducive to stereotype threat emerging during learning or training activities may be greater than in traditional operational testing contexts. Future research in this area should more closely examine organizational education/training practices and policies for their potential to induce stereotype threat effects that may lead to poorer learning outcomes and, ultimately, domain performance for certain groups of individuals.

The primary findings from the current investigation indicate that stereotype threat appears to demonstrably influence the acquisition of domain expertise by inhibiting the development of efficient knowledge structures and heuristic reasoning. Though a critical component of performance in many task applications, such cognitive learning outcomes represent only one portion of the possible construct space related to training and knowledge acquisition that might be negatively influenced by stereotype threat effects. The classification of learning outcomes from Kraiger et al. (1993) summarized in Figure 3 offers a number of alternative variables whose relationships with stereotype threat could be assessed. The results of Rydell, Shiffrin, et al. (2010) provide an initial investigation of the acquisition of skill-based outcomes; however, no research yet exists depicting the influence of stereotype threat on the acquisition of affective-motivational outcomes. This particular set of learning outcomes seems a particularly easy/logical next step for future research in the area given the empirical results and interest which have emerged concerning the influence of stereotype threat on individual's motivational dispositions (e.g., Crocker et al., 1998; Grimm et al., 2009; Major et al., 1998; Marx & Stapel, 2006a, 2006b, etc.). Continued systematic research based on theoretically grounded

conceptualizations of the learning outcomes of interest would be highly beneficial to mapping out the boundary conditions and domains of greatest concern relevant to stereotype threat effects during learning.

Among the boundary conditions warranting further exploration, the structure of the learning environment seems likely to have a significant influence on the manifestation of stereotype threat effects on learning outcomes. In the present study, individuals engaged in learning within the context of an exploratory learning paradigm. While the benefits and prevalence of this particular learning environment have been described at length, a variety of other instructional delivery techniques and modalities are commonly encountered in educational and training settings (e.g., proceduralized instruction, behavioral modeling, group learning, etc.). Each of these approaches possesses their own unique sets of advantages and disadvantages, and may be more or less conducive to manifestation of stereotype threat effects. For example, the relatively unstructured nature of the exploratory learning paradigm employed in the current investigation may have played a significant role in the acquisition of poorer task heuristics on the part of stereotype threatened learners as these individuals were allowed to pursue whatever strategic learning approach seemed most appropriate. Perhaps a more integrative approach to training that attempted to impose stricter sequencing of the effective declarative, procedural, and strategic knowledge components of task performance would help to alleviate the observed deficiencies in stereotype threatened learners' knowledge organization. Future research on the influence of stereotype threat under alternative learning paradigms is needed to more clearly elucidate its potential effects on learning outcomes.

As has been emphasized in a variety of points in this manuscript, the use of a longitudinal/repeated measures design was instrumental in revealing the influence of stereotype

threat effects on the assessed learning outcomes. Despite conceptual frameworks/models which implicate the dynamic nature of the processes involved in the experience of stereotype threat (cf., Schmader et al., 2008), many studies examining the phenomenon have been conducted using cross-sectional designs at a single point in time. While such approaches may be appropriate for certain questions/contexts, the present results exemplify the importance of research examining the influence of stereotype threat utilizing more dynamic methodological designs. Of particular note, nearly all of the significant between-condition differences observed in the present study had not manifested by the end of Day 1, indicating that the effects of stereotype threat on the learning process and other experiential activities may not manifest or begin to create noticeable discrepancies until some amount of time has elapsed. Furthermore, the significant interactions observed between time and experimental condition for many of the measured learning and performance outcome variables also suggest that the adverse effects of stereotype threat may be cumulative such that poor functioning early on during a key developmental process can significantly impede an individual's ability to achieve a desired level of effectiveness. In short, future examinations of stereotype threat need strongly consider employing methodological approaches such as repeated measures designs or computational modeling techniques in order to gain greater insight into the underlying mechanisms driving differences in relevant learning or performance variables.

Of final note, continued research on how the underlying psychological mechanisms and processes which characterize stereotype threat influence relevant outcomes represents both a conceptually and practically important pursuit. For example, the fact that the presence of a negative stereotype pertaining to performance in mathematical domains triggered threat effects for female participants in the present study despite their not believing the task was strongly

related to mathematical ability lends a degree of uncertainty to exactly what conditions need be present in order to elicit stereotype threat. As shown in Figure 1, Schmader et al.'s (2008) summary of the research literature suggested that a positive propositional relation between self and ability is necessary to generate conditions conducive to stereotype threat such that individuals must believe they possess aptitude within the ability domain targeted by the stereotype. In many instances, this linkage is assessed through measures of domain identification that indicate the extent which performance in a particular area is relevant or important to one's sense of self (Smith & White, 2001). While a number of researchers have theorized and reported empirical findings indicating that domain identification is an important precondition for threat effects (e.g., Steele, 1997; Steele & Aronson, 1995), the pattern of results from the present study and those emerging from other recent research (Rydell, Shiffrin, et al., 2010; Jamieson & Harkins, 2007; see Nguyen & Ryan, 2008, for a review) suggest that this relationship may be either nonessential or, perhaps more likely, inadequately specified. Findings consistent with Grimm et al.'s (2009) postulation that stereotype threat effects can be erased by removing the regulatory mismatch between approach/avoidance performance orientations and the reward structure of the environment seem to indicate that all that may be necessary is for individuals to care about doing well/not doing poorly in order for stereotype threat effects to manifest, regardless of their identification with the domain.

More generally, decoding the significance and causal relevance of the working memory accounts advocated by Schmader et al.'s (2008) model and the motivational/disengagement accounts exemplified by Grimm et al. (2009) and other researchers is integral to the development of effective interventions capable of minimizing the deleterious effects of stereotype threat. It seems a virtual certainty that resolution of this ambiguity will involve the integration of

components from both theoretical foundations in a manner that acknowledges the reciprocal dynamic relationship that exists among an individual's cognitive, behavioral, and affective reactions. With respect to learning, the design and demands of the present study all but ensured that learned cognitive inefficiencies would result in worse performance. There may be numerous situations though in which such learned inefficiencies simply make a job or task unnecessarily more difficult, leading to greater fatigue, diminished motivation, burnout, lowered satisfaction, etc. Such a pattern might suggest that stereotype threat effects would exert a stronger influence on cognitive- and/or skill-based outcomes during initial phases of learning, which subsequently leak into more affective/motivational aspects. Irrespective of the true nature of this pattern, research focused solely on identifying differences in outcomes attributable to stereotype threat without simultaneously modeling/testing predictions which permit identification of the primary psychological mechanisms and their interactions responsible for these effects will be far less effective at advancing this research domain towards a more complete understanding of the phenomenon.

#### **Study Limitations and Generalizability**

There are a few limitations to the present study which are relevant to interpreting the validity and scope of the reported findings. With respect to the study sample, the use of a repeated measures design was integral in revealing a number of unique effects. However, the relatively small sample size in relation to most between-group examinations of stereotype threat and concerns related to attrition across the three-day study period are worthy of mention. With respect to sample size, findings from the MRCM analyses which were not supported by the data (e.g., Hypotheses 1-9) were generally not close to being statistically significant, suggesting that any differences in these variables attributable to stereotype threat would be very small and the

inclusion of additional participants would not be likely to generate a different pattern of results. Attrition rates were also fairly low (approximately 20% between Days 1 and 3) and did not appear to vary greatly across experimental condition. However, it is possible that those female participants who did choose to drop out from either the control or stereotype threat conditions may have differed in some non-trivial manner from those who completed all three days. Note that the use of the MRCM analyses was somewhat helpful here as, unlike repeated-measures ANOVA tests, they do not employ listwise deletion and therefore make use of all available data points; consequently, individuals with missing data and/or who did not complete all experimental sessions were still included in the reported analyses and contributed to the observed pattern of results. Nevertheless, collecting a larger sample size with data from all possible time points would help to improve confidence in the stability and generalizability of the present findings.

Another possible limitation is broadly related to the nature of the TANDEM experimental task and the boundary conditions which this task environment implies for drawing generalizable conclusions about the influence of stereotype threat on learning. The TANDEM task environment is one in which rapid application of learned knowledge to make accurate decisions is integral to task completion; consequently, the acquisition of efficient heuristic reasoning is heavily rewarded in this context and one's failure to do so is likely to lead to noticeable deficiencies in performance. However, there may be instances in which developing and learning to rely on such cognitive "shortcuts" may not be as strongly related to desired outcomes, and thus stereotype threat effects during learning might not lead to such problematic outcomes. For example, complex tasks which require careful, deliberate interpretation of large quantities of information in order to make decisions and/or instances in which there are many possible courses of action (e.g., selecting/promoting upper-level organizational executives, forecasting, military

command and control decisions, etc.) may necessitate that all information—rather than only the most diagnostic—is considered before making a decision. Nevertheless, while continued research in this domain is needed to help clarify the generalizability of stereotype threat's impact to learning in other contexts, it seems plausible that the present pattern of results should extend to situations in which the acquisition of heuristic-based reasoning is central to task performance.

Also of relevance to the discussion of generalizability is that the present results were generated under artificial conditions in an experimental lab setting, a common criticism levied against most stereotype threat research (Sackett & Ryan, 2012). Given that this study marks one of the first examinations of stereotype threat during learning-and the only examination of stereotype threat's influence on the development of knowledge structures and cognitive strategy acquisition over time—a critical goal of this research was to demonstrate the possible effects that might be engendered by the presence of negative stereotypes in a learning environment. The decision to continually expose participants to the stereotype threat manipulation over the course of the current experiment was made in order to maximize the likelihood of observing a significant stereotype threat effect on the assessed learning outcomes. As was mentioned previously, while there are far fewer regulations/standardizations regarding the construction of training as opposed to personnel selection/assessment systems, the presence of conditions as strongly adverse in real world applications as those employed presently is unlikely. It seems prudent then to consider the present findings as an initial proof of concept demonstrating what can happen under conditions of stereotype threat, with further research needed to understand how severe such outcomes may be under varying conditions.

Lastly, although mentioned previously, it is worth noting again that the results from the present research were obtained under an exploratory learning paradigm as opposed to a more

traditional proceduralized instructional approach. A significant body of research exists detailing the advantages and disadvantages of such active learning approaches to learning and desirable knowledge acquisition outcomes (cf., Bell, 2002; Bell & Kozlowski, 2008; Kozlowski, Toney, et al., 2001). In general, the literature on this topic indicates that such techniques are particularly useful for enabling learners to develop advanced expertise that allows them to apply learned knowledge and/or acquire new relevant knowledge in a variety of circumstances that extend beyond the training environment. However, many conventional instructional techniques are quite effective at developing foundational declarative knowledge and, to a lesser extent, procedural knowledge in learners. While the present research and that conducted by Rydell and colleagues provides some insights into the possible effects of stereotype threat on these desirable learning outcomes, future research directly comparing these instructional methodologies is needed to ascertain whether they are equally susceptible to similar situational factors.

## Conclusion

The present study was designed to provide insight into the influence of stereotype threat effects on the knowledge acquisition of targeted individuals during learning activities. Utilizing a conceptual framework of critical learning outcomes (Kraiger et al., 1993), empirical attention was specifically focused towards understanding how decrements to working memory purportedly associated with the experience of stereotype threat influenced participants' organization of knowledge and their development of strategic/heuristic reasoning. Assessment of these outcomes over the course of three experimental sessions revealed that females facing stereotype threat during exploratory learning activities had greater difficulty formulating knowledge structures indicative of efficient/adaptive heuristic reasoning and conducive to performance relative to females who did not experience such situational circumstances. Consistent with broader theories

of human cognition and the psychological experiences which purportedly accompany the experience of stereotype threat, it was postulated that such deficiencies emerged due to a perceived need on the part of stereotype threatened individuals to not appear less knowledgeable than others. Such motivations may stimulate stereotype threatened learners to focus more heavily on rote memorization of task-critical concepts as opposed to acquiring more proficient heuristics.

There can be little argument that the elicitation of stereotype threat requires a highly specific confluence of situational and intrapersonal characteristics in order to manifest. However, the present results and those of many other researchers reveal that when such conditions come to together to concoct this perfect storm, individuals facing its presence can be impacted in ways that diminish the likelihood of successfully attaining many desirable and important achievements. Overall, the theoretical framework and empirical results summarized in this investigation offer a general point of departure for examining the role of stereotype threat in learning/instructional contexts that may also contribute to a better understanding of the manifestation of stereotype threat during performance. Future research capable of expanding on this basic foundation stands to provide important insights into a theoretically intriguing and potentially practically meaningful area of psychological research.

## **FOOTNOTES**

<sup>1</sup> While certain characteristics of the environment can make the *experience* of stereotype threat more or less difficult to manage (e.g., Beilock et al., 2007; Spencer et al., 1999; Steele & Aronson, 1995; Steele & Davies, 2003), they should not alter the *presence* of stereotype threat. Consider the following analogy. One evening you are trying to fall asleep in your bed when you suddenly hear an annoying sound from somewhere in your room that will not stop. While situational characteristics such as the acoustics of the room, how tired you are, your location relative to the source of the noise, etc. all may affect how distracting that sound is to you, they do not change the fact that an irritating sound exists in your room. In much the same way, certain situational features can make the experience and influence of stereotype threat more salient, but they do not determine its existence in any meaningful way (Steele, 1997). Thus, the conceptualization adopted here—and which mirrors that of Steele and Schmader—holds that once the cognitive imbalance between self, group, and ability is struck, stereotype threat is present and its consequences can be experienced.

<sup>2</sup> There are a great many more cognitive models than those summarized here that characterize the mechanisms of working memory and information processing and which may be informative to stereotype threat research. For example, features similar to Baddeley's (1986) working memory model are found in Anderson and colleagues' adaptive control of thoughtrational (ACT-R) theory of cognition (e.g., Anderson, 1993, 1996; Anderson & Lebiere, 1998; Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004), Kieras and Meyer's executiveprocess/interactive control (EPIC) model (1997; Kieras, Meyer, Mueller, & Seymour, 1999), Newell's (1990) Soar architecture, and Just and Carpenter's (1992) 3CAPS framework (capacityconstrained collaborative activation-based production system). For the present purposes however, the characterization of these attention-regulating processes as depicted by Baddeley, Engle and their colleagues adequately describe the underlying mechanisms important to the conceptual focus of this study.

<sup>3</sup> For ease of discussion, Table 1 and the associated references in-text refer to these as Studies 1-6; Studies 1-3 refer to the experiments presented in Rydell, Rydell, & Boucher (2010) and Studies 4-6 refer to the experiments conducted in Rydell, Shiffrin, et al. (2010). However, aside from their similar research question, the papers are independent from one another.

<sup>4</sup> Even in an experiment where stereotype threat is introduced only during the presentation of the declarative knowledge to be learned and not reintroduced again during subsequent measurements of declarative learning, the cognitive imbalance between group, ability, and self that gives rise to accompanying physiological stress, hyperactive monitoring, and thought suppression responses has been triggered and is unlikely to dissipate by the time learning assessments are taken. The experience of threat has already emerged at that point and, if no other manipulations are introduced, is likely to persist through subsequent attempts to extract verifiably correct or incorrect facts, statements, and "knowledge questions" from participants with self-report measures (e.g., Beilock et al., 2007; Rydell, Rydell, & Boucher, 2010, Study 3; Rydell, Shiffrin, et al., 2010, Study 3).

<sup>5</sup> The primary modification made to this version of the automated OSPAN from the original was that participants were not provided with any feedback regarding their performance on the math items, the number of letters they correctly recalled following each memory recall, or their final span score at the end of the working memory task.

<sup>6</sup> Note that for the purposes of the current study, it was assumed that the similarity between any two concepts in an individual's knowledge structure satisfied the metric axioms of symmetry and the triangle inequality (Tversky, 1977); that is, the similarity ratings participants provided when asked "How similar is A to B?" were assumed identical to the ratings they would provide if asked "How similar is B to A?" Although previous research by Tversky (1977) has suggested that these features of relational similarity can be (and often are) violated in an individual's judgment, such differences were not important to the aims of the present study. In instances where these discrepancies are of interest, similarity ratings are obtained for every permutation of concept pair combinations ( $k^*(k-1)$  ratings). In the analysis of such data, the knowledge structures are said to be "directed" (i.e., concepts are organized according to an observed causal ordering or hierarchy, Steyvers & Tennenbaum, 2005) and the primary substantive question of interest involves interpreting the meaning of the causal chain found in the concept relations. However, in instances where one is interested in simply assessing relational organization/similarity and the logical clustering of concepts within an area—as in the current study—collecting a single rating for each pairwise combination of concepts (a total of  $k^*(k-1)/2$ ratings) and analyzing undirected knowledge structures is appropriate.

<sup>7</sup> The top 15 performers on the experimental task consisted of three females and five males from the stereotype threat condition, and three females and four males from the control condition. Although this distribution appears to indicate that the stereotype threat manipulation had relatively little impact on the performance capabilities of the very best individuals, it does reflect a reasonably large sex difference in task performance. Bayes' Theorem can be used to calculate the probability that an individual among the top 15 performers will be of a given sex:

$$p(\text{Top15} | \text{Sex}) = \frac{p(\text{Sex} | \text{Top 15}) * p(\text{Top 15})}{p(\text{Sex})}$$

Based on the Day 3 sample sizes reported in Table 2, the probability of a top 15 performer being a woman was only .05, while the probability for males was nearly four times that amount (.19).

<sup>8</sup> For the analyses comparing participant knowledge structures against those of top performers (i.e., knowledge structure similarity and correlation), individuals included in the top performer sample were excluded from the comparison sample. Consequently, these MRCM analyses are based on data from only 139 females. APPENDICES

#### APPENDIX A

#### **Online Informed Consent**

**Project Title:** Learning in a Radar Control Simulation

Investigators:

ors: James A. Grand, M.A. Ann Marie Ryan, Ph.D.

General Description and Explanation of Procedure: The purpose of this study is to examine learning and problemsolving skills that are important to performance in a variety of areas. In this study, you will be tasked with learning to operate a computer-based radar control simulation over multiple practice trials spread over three experimental sessions held on consecutive days. In the radar control simulation, you will be presented with a number of targets on your computer screen which you must assess in order to determine what action to take against each contact. You will also be asked to answer questions and provide ratings about what you learned during the simulation that will help us understand how people approach learning in the task and how those outcomes affect overall performance.

This study has two parts. First, you will be asked to complete an online questionnaire; these measures include questions about basic demographics, your SAT/ACT scores, and other characteristics related to the radar control simulation you will learn. After completing the questionnaires, you will then be asked to participate in the radar control simulation on the days you selected. Filling out the online questionnaire will begin immediately upon agreeing to participate in the study at the bottom of the page and will take less than 30 minutes [1 credit] to complete. For the second part of the study, you will go to the ADAPT Lab in Room 204 of the Psychology Building for your scheduled sessions. The first session on Day 1 will last approximately 2 hours [4 credits], while the following sessions on Days 2 and 3 will each last 1.5 hours [3 credits each].

If you complete all experimental sessions, you will earn 11 subject pool credits for participation in this study. Additionally, you will also be eligible to receive a \$60 award if you complete the entire experiment; additional information about the monetary awards will be provided during the first experimental session. Winners will be determined at the end of the study and will be contacted so they can claim their prize.

Estimated Time Required:	Now: 30 minutes for online questionnaire [1 credit] Day 1: 2 hours [4 credits] Day 2: 1.5 hours [3 credits] Day 3: 1.5 hours [3 credits]
Risks and Discomforts:	None anticipated.
Benefits:	In addition to your compensation for research, you will gain experience operating a computer-based training simulation. This can be valuable as many organizations use computer-based training and simulation programs to teach and/or assess a variety of knowledge, skills, and abilities in their employees or applicants. You will also get to see the results of a number of interesting psychological measures about yourself (e.g., working memory capacity, "knowledge maps") at the end of the study. Finally, the findings from this research are expected to improve our understanding of learning and problem-solving which can be used to improve the effectiveness of training and development tools in real-world situations.

Participation in this study is completely voluntary. By consenting, you also give permission to the experimenters to access or verify your ACT/SAT score from the University Registrar. Your refusal to participate will involve no penalty or loss of benefits to which you are otherwise entitled. You may refuse to participate in certain procedures or answer certain questions. You are free to withdraw this consent and discontinue participation in this project at any time without penalty. If you choose to withdraw from the study prior to its completion, you will receive credit for the time you have spent in the study (1 credit per 30 minutes).

If you have concerns or questions about this study, such as scientific issues, how to do any part of it, or to report an injury (i.e. physical, psychological, social, financial, or otherwise), please contact the project coordinator (James Grand, Department of Psychology, Michigan State University, East Lansing, MI 48824; grandjam@msu.edu; 334-787-2141). If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 207 Olds Hall, MSU, East Lansing, MI 48824.

All data will be stored for at least three years after the project closes. During that time, only the investigators listed on this form will have access to the data collected in this study, and any data reported for scientific purpose will be in aggregate form. The institutional review board will also have access to study data and results if requested in the case of an audit. Your confidentiality will be protected to the maximum extent allowable by law. All reasonable efforts will be taken to ensure your identity and data will be kept secure and confidential.

If you agree to participate, please indicate your consent by selecting "Yes" to the question below. Note that you will also be asked to complete an additional consent form for your participation in the in-person radar control simulation.

# **APPENDIX B**

# In-Person Informed Consent

<b>Project Title:</b>	Learning in a Radar Control Simulation
Investigators:	James A. Grand, M.A. Ann Marie Ryan, Ph.D.
General Description and Explanation of Procedure:	The purpose of this study is to examine learning and problem- solving skills that are important to performance in a variety of areas. In this study, you will be tasked with learning to operate a computer-based radar control simulation over multiple practice trials spread over three experimental sessions held on consecutive days. In the radar control simulation, you will be presented with a number of targets on your computer screen which you must assess in order to determine what action to take against each contact. You will also be asked to answer questions and provide ratings about what you learned during the simulation that will help us understand how people approach learning in the task and how those outcomes affect overall performance.
	You have finished the first portion of this study by completing the online questionnaires. In this part of the study, you will begin learning to complete the radar control simulation task. This first session will last approximately 2 hours [4 credits], while the following sessions on Days 2 and 3 will each last 1.5 hours [3 credits each].
	If you complete all experimental sessions, you will earn 11 subject pool credits for participation in this study. Additionally, <b>you will also be eligible to receive a \$60 award if you complete the entire experiment</b> . Awards will be based on the combined score you receive on the final performance trials for each day; those scoring in the top 10% of all participants in the study will receive the monetary award. Winners will be determined at the end of the study and will be contacted by the investigator so they can claim their prize.
Estimated Time Required:	Now: 2 hours [4 credits] Day 2: 1.5 hours [3 credits] Day 3: 1.5 hours [3 credits]
<b>Risks and Discomforts:</b>	None anticipated.
Benefits:	In addition to your compensation for research, you will gain experience operating a computer-based training simulation. This can be valuable as many organizations use computer-based training and simulation programs to teach and/or assess a variety of knowledge, skills, and abilities in their employees or
applicants. You will also get to see the results of a number of interesting psychological measures about yourself (e.g., working memory capacity, "knowledge maps") at the end of the study. Finally, the findings from this research are expected to improve our understanding of learning and problem-solving which can be used to improve the effectiveness of training and development tools in real-world situations.

Participation in this study is completely voluntary. By consenting, you also give permission to the experimenters to access or verify your ACT/SAT score from the University Registrar. Your refusal to participate will involve no penalty or loss of benefits to which you are otherwise entitled. You may refuse to participate in certain procedures or answer certain questions. You are free to withdraw this consent and discontinue participation in this project at any time without penalty. If you choose to withdraw from the study prior to its completion, you will receive credit for the time you have spent in the study (1 credit per 30 minutes).

If you have concerns or questions about this study, such as scientific issues, how to do any part of it, or to report an injury (i.e. physical, psychological, social, financial, or otherwise), please contact the project coordinator (James Grand, Department of Psychology, Michigan State University, East Lansing, MI 48824; grandjam@msu.edu; 334-787-2141). If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 207 Olds Hall, MSU, East Lansing, MI 48824.

All data will be stored for at least three years after the project closes. During that time, only the investigators listed on this form will have access to the data collected in this study, and any data reported for scientific purpose will be in aggregate form. The institutional review board will also have access to study data and results if requested in the case of an audit. Your confidentiality will be protected to the maximum extent allowable by law. All reasonable efforts will be taken to ensure your identity and data will be kept secure and confidential.

If you agree to participate, please indicate your consent by signing below. You are also asked for your name and PID to ensure that you receive credit for participating in the study and to verify your ACT/SAT score. Lastly, please provide your e-mail address and/or a phone number where you can be contacted if you win a prize.

Print Name:	Date:
Signature:	PID:
E-mail:	Phone:

## APPENDIX C Participant Feedback/Debriefing

Thank you for your participation in this investigation. The purpose of this study was to examine sex differences in learning in domains where a negative stereotype exists about a particular group. Some of you received instructions over the course of the experiment which indicated that women are stereotypically worse at mathematical tasks as a result of inefficient information processing skills than men. Previous research suggests that the presence of such stereotypes can cause members of the negatively stereotyped group to engage in overactive performance monitoring and thought suppression activities in their attempt to avoid proving the negative performance stereotype correct by their actions. Unfortunately, these psychological activities often make the problem worse, and can cause such individuals to do worse on related tasks. This phenomenon is known as *stereotype threat*.

In the present study, we were interested in the extent to which women who were exposed to stereotype threat approached and learned the radar control task differently than women who were not exposed to the threat and men. Our expectation was that informing women that this task measures skills related to math performance—a domain in which men are typically believed to be more proficient—would induce stereotype threat and thus negatively influence learning and performance outcomes in the task. However, please note the following:

- Men and women tend to differ very little, if at all, on most tests of mathematical ability. Despite prevalent stereotypes that men are better at math than women, there is substantial evidence which reports that, on average, these differences are often very small and/or not statistically significant. Additionally, there is no evidence to suggest that women are worse at distinguishing relevant from irrelevant information than men. For further reading on this topic, see:
  - Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, 43, 95-103
  - Feingold, A. (1992). Sex difference in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, 62, 61-84
  - Hedges, L.V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269, 41-45).
- Individuals will be compared against members of their same sex and experimental condition to determine who will win the monetary awards. We expected that women exposed to stereotype threat would be less proficient at the radar control simulation task than others. To ensure that everyone in the study has an equal opportunity to win the monetary awards then, awards will be distributed to males and females separately according to their experimental condition. More specifically, rewards will be allocated according to the following chart:

	Men	Women		
ST instructions	Top 10%	Top 10%		
No ST instructions	Top 10%	Top 10%		

Thus, one set of awards will be rewarded to the top 10% of men who received stereotype threat instructions, while a different set of awards will be given to the top 10% of women who received stereotype threat instructions; the same is true for men and women who did not receive stereotype threat instructions.

We hope that the information you provided in this study will help us determine whether unfortunate situational influences such as stereotype threat might adversely impact the manner by which people learn and make sense of information. We hope that in doing so, we can ultimately better design and inform educational and training policies to ensure that members of all demographic categories can effectively learn and perform in traditionally stereotyped domains.

If you have any questions about this study, please notify the investigator now. We will e-mail you with the results of the knowledge structures you completed for this study within two weeks of this experimental session and will contact winners of the monetary awards upon completion of the experiment. If you have any additional questions about the study or your involvement in it, contact:

#### James Grand, M.A. 348 Psychology Building Phone: (334) 787-2141 e-mail: grandjam@msu.edu

If you have questions or concerns about your rights as a research participant, please contact Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 207 Olds Hall, MSU, East Lansing, MI 48824

# APPENDIX D

Study Protocol

## **Day 1 Session**

## Setting up participant computers

- 1. Check the Experiment Information Sheet for the day (located in study folder) to see how many people are signed up for the session
  - A. The experiment information sheet contains a number of important items:
    - i. A roster of all individuals signed up to participate in the study, with columns to mark which days/sessions the person attended
    - ii. The dates over which the sessions will take place
    - iii. The condition for the session participants is located in the top right corner
      - 1. O = condition 1 (stereotype threat)
      - 2. T = condition 2 (no stereotype threat)
  - B. James will make sure this information sheet is updated with the correct roster prior to each Day 1 session
- 2. Log into the lab computers using the following info:
  - A. Username: XXXXX
  - B. Password: XXXXX
  - C. NOTE: If running in the Optima lab, make sure the domain name is set to Psychology
- 3. Place an informed consent sheet onto every logged in participant computer
- 4. On the desktop, open the folder labeled "Radar Control Simulation"
  - A. Double-click the shortcut labeled "RadarSim Start" to open up the start page for the Radar Control Simulation experiment.
  - B. Leave the start page open and go back to the Radar Control Simulation folder.
  - C. Double-click the shortcut labeled "OSPAN" to open up the Automated Operation Span working memory task.
  - D. Press Alt+Tab until you get back to the Radar Control Simulation folder and close the folder.
  - E. At this point, only the OSPAN task and Radar Control Simulation start page should be opened.
- 5. If needed, Alt+Tab back to the OSPAN task so that it is on the screen.
  - A. In the window labeled "Enter Subject Number," enter in the number written in the top right corner of the informed consent document you placed at that computer and press OK. NOTE that it is <u>very important</u> that this number be correct, so double check to make sure you've entered in the correct number before pressing OK.
  - B. In the window labeled "Enter Session Number," leave it at 1 and press OK.
  - C. At the confirmation window, press OK. You should now see the instructions for the OSPAN task on screen. Leave this open.

- 6. Repeat steps 4-6 for every participant computer you have logged in.
  - A. Once done, the participant computers are all set!

# **Preparing the Experimenter Station & PowerPoint Training**

- - A. If running in Optima Lab (Room 335)
    - i. Make sure domain is set to IOPSYCH
    - ii. Username: XXXXX
    - iii. Password: XXXXX
  - B. If running in Adapt Lab (Room 204)
    - i. Username: XXXXX
    - ii. Password: XXXXX
  - C. Once the computer is turned on, check to make sure the volume is turned on for the computer
- 8. Log into the experimenter website by doing the following:
  - A. Go to 35.8.48.6/james1/exppage.asp
  - B. Password: XXXXX
- 9. Check to make sure that the area labeled "Current Condition" is set to the correct condition number for the experiment that day
  - A. To check the condition number for the group you are running, look in the top right corner of the Experiment Information Sheet
    - i. If O (stands for "one") is circled, Current Condition should be set to 1
    - ii. If T (stands for "two") is circled, Current Condition should be set to 2
  - B. If you need to change the Current Condition, select the appropriate number from the drop-down box and click the button labeled Submit
    - i. IF YOU NOTICE THIS IS DIFFERENT THAN IT SHOULD BE, PLEASE CALL ME AND CHECK WITH ME FIRST BEFORE CHANGING!
  - C. Close out of the experimenter page when finished
- 10. Turn on the projector in the room by pressing the power button on the bottom of the projector
- 11. On the desktop, double-click on the icon labeled "Microsoft PowerPoint Viewer" to open up the PowerPoint training presentation
  - A. When the program starts, it will ask you to browse to the file you wish to run. Navigate to the Desktop, and open the folder labeled Radar Control Simulation Training
  - B. Select the appropriate PowerPoint training file based on the experimental condition:
    - i. If running participants for Condition 1 (O on the Experiment Information Sheet), select the file named "training\_condition1 (ST)"
    - ii. If running participants for Condition 2 (T on the Experiment Information Sheet), select the file named "training\_condition2 (NST)"

- C. The PowerPoint viewer will start with the selected presentation. To make the presentation full screen, click the on the small arrow cursor in the bottom right of the window and select full screen
  - i. The training video is now set to run I recommend clicking to start to make sure it works and the sound is turned on!
- 12. You're all set! The room is now ready to go for participants.

## **Running the Task**

13. Set out the Experimenter Information Sheet. As participants enter, provide the following instructions:

I have placed the experiment roster for today's study here. Please find your name on the roster and sign your name under the Day 1 column. I will use this sheet to record attendance and assign credits, so it is important that you sign in on this sheet every day you attend the study. If you do not see your name listed on the roster please let me know. [If somebody's name is not on the roster, go to the Troubleshooting section in the back of this document]

After you have signed in on the roster, you may select any computer that is logged in and has a consent form. Please take a moment to read this form as it provides important information about today's experiment. If after reading the form you still wish to participate in the study, please provide the requested information and sign your name in the spaces provided on the back of the sheet. Please do not press anything on the computer until I instruct you to do so.

- 14. Unless all participants have arrived, wait to begin the study until a few minutes after the scheduled start time. When you are ready to begin:
  - A. Collect the informed consent forms from participants check to make sure they are signed!
  - B. Place the Do Not Disturb sign on the door and shut the door.
    - i. Once the Do Not Disturb sign is up and the door is closed, the experiment has started. No one else is allowed in the room, so if someone arrives late, tough beans.
- 15. To begin the study, read the following to participants:

Thank you for attending the Radar Control Simulation study today. Today is the first of three sessions for this study; this session will last approximately 2 hours, while the remaining sessions will last approximately 1.5 hours. At the end of today's session, I will provide you with a reminder slip that has the date and time for the next session. Note that the credits for this experiment will be updated on the HPR site after the third session.

Before we begin, please place all your cell phones on silent and put them away for the remainder of today's session.

Before starting the Radar Control Simulation, you will first complete a brief task in which you will be asked to memorize a sting of letters while performing math problems. Please follow the instructions exactly as they are presented to you on the screen. When you complete the task, raise your hand to let me know you are finished and then remain sitting quietly until everyone is finished.

16. Once everybody has finished taking the OSPAN task, read the following instructions:

*Please press the Space Bar to exit the memory task.* [Wait to make sure everyone closes out of the task]. *You should now be at the log-in page for the Radar Control Simulation task. Please enter "start" as the start code and then press submit. On the next page, please enter in the requested information. If you receive an error, check to make sure you entered in your PID correctly; if you still receive an error, please let me know.* [If somebody continues to receive an error when trying to log-in, go to the Troubleshooting section in the back of this document]

17. After everybody has successfully logged in, read the following instructions:

I will now begin a brief video that will introduce you to the study as well as how to operate the Radar Control Simulation. Please pay close attention to the video. If you need to move in order to see the screen better, please feel free to do so now.

18. Begin the PowerPoint presentation by clicking on the screen (you should turn off a couple of the lights in the room to make it easier to see). The presentation should proceed automatically through the slides until it reaches the end (lasts about 10 minutes). Once it is complete, read the following instructions:

*Before we start the Radar Control Simulation for today, are there any questions?* [If you get any questions about how to play the game, let them know that everything they need to know to play the game can be found in the manual and they will need to look it up]

You may now press Next on your screen. We will now begin the familiarization trial for today where you will be able to briefly look at the task manual and radar screen before beginning the first learning trial for the day. NOTE that the presentation of the manual and radar screen during the familiarization trial is VERY short. This time is simply meant to allow you to see what these screens look like and familiarize yourself with how to use some of their functions. I will now put the first password onto the screen.

19. Advance the PowerPoint presentation to the first password. Check to make sure that everybody is able to enter the manual without any problems. Once the timer for the manual has expired and everybody is on the next green screen, read the following instructions:

You will now have a chance to familiarize yourself with the radar screen and how to use its menus. Please make note of the following:

- A. To hook a contact on the radar, you will use the LEFT-mouse button. To operate your sensors and radar menus, you will use the RIGHT-mouse button.
- B. At the start of every trial, you MUST RIGHT-CLICK on Start Exercise from the OPER menu to begin the trial.

#### *I will now provide the password needed to advance to the radar screen.*

20. Advance the PowerPoint presentation to the next password. At this point you should walk around the room and make sure that everybody was able to enter the game and clicked on "Start Exercise" from the OPER menu. To do so, glance at the timer on each person's screen – if it still reads "Paused 1:00" they have not started the simulation. Show them how to start the simulation by RIGHT-CLICKING on the OPER menu and then RIGHT-CLICKING on Start\_Exercise. Remind the participant that they will need to do this every time they begin a new trial. Once everybody has completed the familiarization trial and viewed the practice feedback, read the following instructions:

From here on, I will project the passwords needed to advance the task up on the projector screen. When you see the password on the screen, you may enter it into the password field and press Submit. I will put the password on the screen once everybody is ready to proceed to the next step.

Once you have completed all the learning trials and the single performance trial for today, you will be asked to complete a series of measures. These measures are important to the task, and therefore it is important to us that you to attend to them as fully and conscientiously as you do the Radar Control Simulation.

Before we begin, please put on the headphones located at your computer. Once you have put the headphones on, press the button labeled "Test Sound" to confirm that your headphones are working and that you can hear audio. Please raise your hand if you are unable to hear any sound from your headphones.

[Check to make sure everyone can hear the audio from the computer. If somebody's computer is not playing audio, make sure that the volume is not muted and is turned up enough to hear. If somebody would like to adjust the volume on the headphones, they may do so by using the controls located on the headphone cord.]

# If there are no questions before we begin, I will now put the first password on the projector screen.

- 21. Advance the PowerPoint presentation to the next password. From here on, you should simply monitor participants as they complete the task and provide the passwords when needed. NOTE that you will only present the next password once all participants are ready to advance (i.e., you see the green screen on everybody's computer).
  - A. Some things to keep track of:
    - i. There is no talking allowed or use of cell phones. If anybody is doing either of these please ask them to leave.

- ii. Jot down a note of any unusual activity that you see (i.e., somebody's not paying attention to the task, talking, cell phone use) or any computer/task malfunctions (i.e., you had to restart someone's game, etc.)
- 22. While people are completing the task, make sure you fill out the reminder slips and set them out for people to take with them as they leave.

# **Closing out the experiment**

- 23. Once the last person has finished the experiment and has left the room, you will need to go to all the computers that were used by a participant that day and complete two tasks:
  - A. Enter in "finish1" as the End Day code at the screen currently on the computer
    - i. This should advance the screen to the Welcome Back screen (this is the screen where people will begin on Day 2)
    - ii. Once on the Welcome Back screen, exit out of the window
  - B. Open up the Radar Control Simulation folder on the desktop and double-click the **Copy OSPAN data** program
    - i. This will open up a command window that will ask you to press any key to continue...go ahead and press any key to continue
    - ii. The program will copy over the TWO working memory/OSPAN files from the participant's computer to the server
    - iii. When it's done, press any key to continue and it will close on its own
- 24. Close the training/password PowerPoint Presentation and log out of the experimenter computer
- 25. Make sure all the consent forms and Experiment Information Sheet are put away
- 26. Turn off the lights and projector in the room (you may leave the participant computers on and logged in) and make sure the door is closed when you leave

# **Day 2 Session**

## Setting up participant computers

- 1. Check the Experiment Information Sheet for the day (located in study folder) to see how many people are signed up for the session
  - A. The experiment information sheet contains a number of important items:
    - i. A roster of all individuals signed up to participate in the study, with columns to mark which days/sessions the person attended
    - ii. The dates over which the sessions will take place
    - iii. The condition for the session participants is located in the top right corner
      - 1. O = condition 1 (stereotype threat)
        - 2. T = condition 2 (no stereotype threat)
  - B. James will make sure this information sheet is updated with the correct roster prior to each Day 1 session

- 2. If needed, log into the lab computers.
- 3. On the desktop, open the folder labeled "Radar Control Simulation"
  - A. Double-click the shortcut labeled "RadarSim Restart" to open up the welcome back log-in page for the Radar Control Simulation experiment.
  - B. Close the Radar Control Simulation folder.
  - C. At this point, only the Radar Control Simulation Restart page should be opened.
- 4. Repeat steps 3 for every participant computer you have logged in.
  - A. Once done, the participant computers are all set!

# **Preparing the Experimenter Station & PowerPoint Presentation**

- 5. Log into the experimenter station.
- 6. Turn on the projector in the room by pressing the power button on the bottom of the projector
- 7. On the desktop, open the folder labeled Radar Control Simulation Training and open up the appropriate PowerPoint training file (it takes a second to open, it's a big file...)
  - A. Select the appropriate PowerPoint training file based on the experimental condition:
    - i. If running participants for Condition 1 (O on the Experiment Information Sheet), select the file named "training condition1 (ST)"
    - ii. If running participants for Condition 2 (T on the Experiment Information Sheet), select the file named "training condition2 (NST)"
  - B. Navigate the presentation to begin at Slide 53
  - C. Start the presentation in full screen mode and make sure it is projecting on the screen
- 8. You're all set! The room is now ready to go for participants.

# **Running the Task**

- 9. Set out the Experimenter Information Sheet. As participants enter, ask them to sign in under the appropriate day, take a seat at one of the logged-in computers, and enter their PID on the log-in screen (NOTE: Participants do not have to sit at the same computer as they did on previous days)
- 10. Unless all participants have arrived, wait to begin the study until a few minutes after the scheduled start time. When you are ready to begin:
  - A. Collect the informed consent forms from participants check to make sure they are signed!
  - B. Place the Do Not Disturb sign on the door and shut the door.
    - i. Once the Do Not Disturb sign is up and the door is closed, the experiment has started. No one else is allowed in the room, so if someone arrives late, tough beans.

11. To begin the study, read the following to participants:

Welcome back to the Radar Control Simulation study. Today's session will last approximately 1.5 hours. Unlike yesterday's session, there will be no training video nor will you complete the memorization task.

Before we begin, please place all your cell phones on silent and put them away for the remainder of today's session.

Similar to yesterday's session, we will begin with a short familiarization period for you to re-acclimate yourself with the radar manual and screen. I will present you with the passwords to enter these screens when everyone is ready to proceed

Are there any questions before we get started? [Answer any questions]

At this point, you may press the Next button on your computer and I will provide you with the first password.

12. Advance the PowerPoint presentation to the first password. EVERYBODY WILL HAVE 30 SECONDS TO LOOK AT THE MANUAL. ONCE THEY ARE DONE WITH THE MANUAL, YOU WILL PRESENT THE NEXT PASSWORD TO BEGIN A SHORT TRIAL WITH THE RADAR TASK. Once participants have completed the familiarization trial and feedback page, read the following instructions:

The remainder of today's session will be identical to yesterday's session; I will project the passwords needed to advance the task up on the projector screen when everybody is ready to advance. When you see the password on the screen, you may enter it into the password field and press Submit.

Once you have completed all the learning trials and the single performance trial for today, you will again complete a series of measures. These measures are important to the task, and therefore it is important to us that you to attend to them as fully and conscientiously as you do the Radar Control Simulation.

Before we begin, please put on the headphones located at your computer. Once you have put the headphones on, press the button labeled "Test Sound" to confirm that your headphones are working and that you can hear audio. Please raise your hand if you are unable to hear any sound from your headphones.

*If there are no questions before we begin, I will now put the password on the projector screen.* 

13. Advance the PowerPoint presentation to the next password. From here on, you should simply monitor participants as they complete the task and provide the passwords when needed. NOTE that you will only present the next password once all participants are ready to advance (i.e., you see the green screen on everybody's computer).

- C. Some things to keep track of:
  - i. There is no talking allowed or use of cell phones. If anybody is doing either of these please ask them to leave.
  - ii. Jot down a note of any unusual activity that you see (i.e., somebody's not paying attention to the task, talking, cell phone use) or any computer/task malfunctions (i.e., you had to restart someone's game, etc.)
- 14. While people are completing the task, make sure you fill out the reminder slips and set them out for people to take with them as they leave.

## **Closing out the experiment**

- 15. Once the last person has finished and left the room, you will need to go to all the computers that were used by a participant that day and enter in "**finish2**" as the End Day code
  - A. This should advance the screen to the Welcome Back screen (this is the screen where people will begin on Day 3)
  - B. Once on the Welcome Back screen, exit out of the window
- 16. Close the training/password PowerPoint log out of the experimenter computer
- 17. Put away the Experiment Information Sheet
- 18. Turn off the lights and projector in the room (you may leave the participant computers on and logged in) and make sure the door is closed when you leave

# **Day 3 Session**

## Setting up participant computers

- 1. Check the Experiment Information Sheet for the day (located in study folder) to see how many people are signed up for the session
  - A. The experiment information sheet contains a number of important items:
    - i. A roster of all individuals signed up to participate in the study, with columns to mark which days/sessions the person attended
    - ii. The dates over which the sessions will take place
    - iii. The condition for the session participants is located in the top right corner
      - 1. O = condition 1 (stereotype threat)
      - 2. T = condition 2 (no stereotype threat)
  - B. James will make sure this information sheet is updated with the correct roster prior to each Day 1 session
- 2. If needed, log into the lab computers.
- 3. On the desktop, open the folder labeled "Radar Control Simulation"
  - A. Double-click the shortcut labeled "RadarSim Restart" to open up the welcome back log-in page for the Radar Control Simulation experiment.

- B. Close the Radar Control Simulation folder.
- C. At this point, only the Radar Control Simulation Restart page should be opened.
- Repeat steps 3 for every participant computer you have logged in.
   A. Once done, the participant computers are all set!

# **Preparing the Experimenter Station & PowerPoint Presentation**

- 5. Log into the experimenter station.
- 6. Turn on the projector in the room by pressing the power button on the bottom of the projector
- 7. On the desktop, open the folder labeled Radar Control Simulation Training and open up the appropriate PowerPoint training file (it takes a second to open, it's a big file...)
  - A. Select the appropriate PowerPoint training file based on the experimental condition:
    - i. If running participants for Condition 1 (O on the Experiment Information Sheet), select the file named "training condition1 (ST)"
    - ii. If running participants for Condition 2 (T on the Experiment Information Sheet), select the file named "training condition2 (NST)"
  - B. Navigate the presentation to begin at Slide 87
  - C. Start the presentation in full screen mode and make sure it is projecting on the screen
- 8. You're all set! The room is now ready to go for participants.

# **Running the Task**

- 9. Set out the Experimenter Information Sheet. As participants enter, ask them to sign in under the appropriate day, take a seat at one of the logged-in computers, and enter their PID on the log-in screen (NOTE: Participants do not have to sit at the same computer as they did on previous days)
- 10. Unless all participants have arrived, wait to begin the study until a few minutes after the scheduled start time. When you are ready to begin:
  - C. Collect the informed consent forms from participants check to make sure they are signed!
  - D. Place the Do Not Disturb sign on the door and shut the door.
    - iii. Once the Do Not Disturb sign is up and the door is closed, the experiment has started. No one else is allowed in the room, so if someone arrives late, tough beans.
- 11. To begin the study, read the following to participants:

Welcome back to the final session for the Radar Control Simulation study. Today's session will last approximately 1.5 hours and the procedure will be identical to yesterday's session.

Before we begin, please place all your cell phones on silent and put them away for the remainder of today's session.

Similar to yesterday's session, we will begin with a short familiarization period for you to re-acclimate yourself with the radar manual and screen. I will present you with the passwords to enter these screens when everyone is ready to proceed

Are there any questions before we get started? [Answer any questions]

At this point, you may press the Next button on your computer and I will provide you with the first password.

12. Advance the PowerPoint presentation to the first password. Advance the PowerPoint presentation to the first password. EVERYBODY WILL HAVE 30 SECONDS TO LOOK AT THE MANUAL. ONCE THEY ARE DONE WITH THE MANUAL, YOU WILL PRESENT THE NEXT PASSWORD TO BEGIN A SHORT TRIAL WITH THE RADAR TASK. Once participants have completed the familiarization trial and feedback page, read the following instructions:

The remainder of today's session will be identical to yesterday's session; I will project the passwords needed to advance the task up on the projector screen when everybody is ready to advance. When you see the password on the screen, you may enter it into the password field and press Submit.

Once you have completed all the learning trials and the single performance trial for today, you will again complete a series of measures. These measures are important to the task, and therefore it is important to us that you to attend to them as fully and conscientiously as you do the Radar Control Simulation.

Before we begin, please put on the headphones located at your computer. Once you have put the headphones on, press the button labeled "Test Sound" to confirm that your headphones are working and that you can hear audio. Please raise your hand if you are unable to hear any sound from your headphones.

# If there are no questions before we begin, I will now put the password on the projector screen.

- 13. Advance the PowerPoint presentation to the next password. From here on, you should simply monitor participants as they complete the task and provide the passwords when needed. NOTE that you will only present the next password once all participants are ready to advance (i.e., you see the green screen on everybody's computer).
  - E. Some things to keep track of:
    - iv. There is no talking allowed or use of cell phones. If anybody is doing either of these please ask them to leave.

- v. Jot down a note of any unusual activity that you see (i.e., somebody's not paying attention to the task, talking, cell phone use) or any computer/task malfunctions (i.e., you had to restart someone's game, etc.)
- 14. While people are completing the task, make sure you fill out the reminder slips and set them out for people to take with them as they leave.

# **Closing out the experiment**

- 15. Close the training/password PowerPoint Presentation and log out of the experimenter computer
- 16. Make sure the Experiment Information Sheet is put away
- 17. Turn off the lights and projector in the room (you may leave the participant computers on and logged in) and make sure the door is closed when you leave.

# Troubleshooting

I don't foresee there being any significant technical issues with the task...however, below I've listed a few scenarios that may come up. Note that you can always call me if you are having difficulty with something and I will do my best to troubleshoot remotely (or on site, if I'm around)

## 1. Somebody's not on the Participant Roster!

- a. This means that the individual did not complete the sign-up on the HPR site and, by extension, probably did not complete the required online questionnaire (so you'll have the problem below as well).
- b. If you have enough computers, let them write in their name on the roster and participate.
- c. If there is enough time before the session starts (i.e., at least 5 minutes), start up a web browser and go to http://35.8.48.6/james1. Have them complete the consent and questionnaire there.
- d. If there's not enough time before the session starts, see #2 below

## 2. Somebody's trying to log-in with their PID and it's not working!

- a. This could mean a few different things, but most likely it's either A) they entered a different/the wrong PID when they completed the online questionnaire or B) the participant did not complete the online questionnaire portion of the study
- b. In either case, here's what you should do:
  - i. Ask the person to give you their PID
  - ii. Continue on with the PowerPoint training
  - iii. Call me as soon as you can there are some things I will need to do in the database so that they can continue on with the task

- c. In the case of scenario B above, I will make a temporary fix that will allow them to participate in the session. However, they will need to take the online questionnaire before they leave the lab
  - i. At the end of the session, have them go to http://35.8.48.6/james1 and complete the consent and questionnaire
- 3. Something does not seem to be working correctly in the task! (e.g., not displaying correctly, error message, etc.)
  - a. Try to restart the simulation for the person by exiting the window and clicking on the RadarSim Restart shortcut in the desktop folder
  - b. If this doesn't fix the problem or the error persists, call James

# APPENDIX E

## Practice Trial Feedback

Page 1

Participant Feedback
INSTRUCTIONS

You will now have an opportunity to review your activity from this past practice period. You should use the information provided on the following screens to guide your study and practice. Remember that once you leave a screen, you cannot go back to review it again. You will have 1 minute to review the feedback pages.

Advance to the next screen to begin reviewing your feedback.

CODDIG	Page 2
SCORING	
Your total score on this past trial: XXXX points	
TARGET ENGAGEMENT RESULTS	
Number of non-marker targets engaged: XX out of 21 targets	
You correctly engaged XX targets, which earned you XX points.	
You incorrectly engaged XX targets, which cost you XX points.	
PERIMETER INTRUSIONS	
You allowed XX targets to cross the inner perimeter, which cost you XX points.	
You allowed XX targets to cross the outer perimeter, which cost you XX points.	
Together, you lost XX points due to targets crossing your defensive perimeters.	
Advance to the next screen to continue reviewing your feedback.	

Page 3

## CONTACT DECISIONS

Number of contacts for which you made a correct Type (Air, Surface, Sub) decision: XX out of XX total contacts engaged

Number of contacts for which you made a correct Class (Civilian, Military) decision: XX out of XX total contacts engaged

Number of contacts for which you made a correct Intent (Peaceful, Hostile) decision: XX out of XX total contacts engaged

Number of contacts for which you made a correct Final Engagement (Clear, Warn, Mark) decision: XX out of XX total contacts engaged

Remember, you must make each of the four decisions above correctly for EACH target in order to receive points! Even making one of the above decisions incorrectly will cause you to lose points.

Advance to the next screen to continue reviewing your feedback.

Page 4 OTHER INFORMATION Average time spent per target: XX seconds Number of pop-up targets you engaged: XX Number of high priority targets you engaged (correctly or incorrectly) that threatened the inner perimeter: X out of 4 Number of high priority targets you engaged (correctly or incorrectly) that threatened the outer perimeter: X out of 7 You did not hook a marker target on this trial. [You hooked and used a marker target correctly on this trial.] [You hooked a marker target, but engaged it. Engaging marker targets does not earn you points in the trial.] You did not use the Zoom feature on this trial.

Page 5

[You used the Zoom feature on this trial.]

FEEDBACK COMPLETE This concludes the trial feedback.

Please click Next to exit the feedback program and wait for further instructions.

## **APPENDIX F**

# Exploratory Learning Recommendations (adapted from Bell, 2002)

An effective method for learning the skills required in the Radar Control Simulation is to explore the task and develop your understanding of it. As you practice the scenarios, explore the task to understand what is occurring in the scenario and to discover the best strategies to deal with the situation. Also, experiment with different strategies and methods in your attempts to learn the skills needed to perform effectively in the simulation.

# The following is a list of questions that may prove useful in guiding your learning within the task:

## **Gathering and Interpreting Information**

- Have you learned how to efficiently gather the required number of information pieces from your sensors needed to process a target?
- Have you learned how to interpret the cues needed to make accurate decisions about a target's Type?
- Have you learned how to interpret the cues needed to make accurate decisions about a target's Class?
- Have you learned how to interpret the cues needed to make accurate decisions about a target's Intent?
- Have you learned how to use information about a target's Type, Class, and Intent to make accurate decisions about whether to Clear, Warn, or Mark the target?
- Have you figured out how to use the in-game feedback to determine whether you correctly engaged a target?

## **Monitoring Defensive Perimeters**

- Have you discovered how to use the Zoom function to help you look for targets near the inner and outer perimeters?
- Have you learned how to identify and use the Marker targets in order to locate and monitor the outer defensive perimeter?
- Have you developed a strategy for monitoring your inner and outer perimeters for pop-up targets?

## **Prioritizing Targets and Maximizing your Score**

- Have you learned to distinguish high priority targets from low priority targets?
- Have you come up with a strategy for prioritizing which target(s)s to engage in order to maximize the number of points you earn and minimize the number of points you lose?

## **APPENDIX G**

Stereotype Threat Condition Instructions (adapted from Beilock et al., 2007; Rydell, Shiffrin, et al., 2010)

#### Full Instructions

As mentioned during the first experimental session, there has been some controversy about whether there are true sex differences in math ability. A good deal of research indicates that males consistently score higher than females on standardized tests of math ability; however there are many cases where no such sex differences emerge. The research you are participating in today is aimed at better understanding why these differences exist.

One reason why women appear to do more poorly on some math tasks (like word problems on the SAT or ACT) is that, relative to men, women are worse at quickly and correctly distinguishing relevant information needed to solve a problem from irrelevant/distracting information that is also provided. This skill is directly assessed in the radar control task you will be using in this study.

By learning to operate the radar control simulation, we will be able to evaluate how you develop these skills and why women tend to have more difficulty processing such information than men. It is important that you give a strong effort during the experiment to help us in our analysis of why men and women differ in this math-related skill.

#### Shortened Instructions

As was mentioned at the outset, one reason why women perform more poorly than men on some math tasks like word problems on the SAT or ACT is that **women tend to have more difficultly correctly choosing information needed to solve a problem from irrelevant/distracting information that is also provided**. The research you are participating in today is aimed at examining your ability to learn this skill so that we may better understand why women have more difficulty processing such information than men. It is important that you continue to give a strong effort during the experiment to help us in our analysis of this topic.

#### Performance Trial Instructions

You have now finished all of your practice trials for today and will complete one final trial. This trial will look and function similarly to the previous scenarios you have practiced, but it will be more challenging. For this trial, the following changes have been made:

- 1. The trial length has been increased to eight minutes
- 2. The total number of targets has been increased substantially
- 3. Any target that crosses your defensive perimeters will now cause you to lose **150 points**

Remember that your performance on this trial will be taken into consideration when selecting the winners of the cash prize, so you should try to score as many points as possible.

## **APPENDIX H**

Control Condition Instructions (adapted from Beilock et al., 2007; Rydell, Shiffrin, et al., 2010)

#### Full instructions

As mentioned during the first experimental session, you may have noticed during your daily experiences that some people seem to be very good at "picking up on" how to perform new tasks they have never seen before. However, surprisingly little is known about the mental processes underlying this skill. This research is aimed at better understanding how individuals approach novel tasks and complete them.

One reason why certain people do better on such tasks is that they may be better at **quickly and correctly distinguishing relevant information needed to solve a problem from irrelevant/distracting information that is also provided**. This skill is involved in completing the radar control task you will be using in this study.

By observing how people learn to operate the radar control simulation, we hope to see how people develop these skills and why some people are differ on them. It is important that you give a strong effort during the experiment to help us in our analysis of why people differ in this skill.

#### Shortened Instructions

As was mentioned at the outset, people's ability to correctly distinguish important information needed to solve a problem from irrelevant/distracting information that is also provided may contribute to performance on many novel tasks. The research you are participating in today is aimed at examining this topic, the manner by which individuals learn this skill, and the extent to which it influences effectively operating the radar control simulation. It is important to us that you continue to give a strong effort during the experiment to help us in our analysis of this topic.

#### Performance Trial Instructions

You have now finished all of your practice trials for today and will complete one final trial. This trial will look and function similarly to the previous scenarios you have practiced, but it will be more challenging. For this trial, the following changes have been made:

- 1. The trial length has been increased to **eight** minutes
- 2. The total number of targets has been **increased** substantially
- 3. Any target that crosses your defensive perimeters will now cause you to lose 150 points

The purpose of this trial is to give you an opportunity to test your newly developed skills in a more challenging context.

#### **APPENDIX I**

#### Online Individual Difference Measures

#### **Demographics**

The questions below ask you to provide some basic information about yourself. Please answer these questions by selecting or typing the appropriate response.

- 1. What is your sex?
  - a. Female
  - b. Male
- 2. What is your age?

a. \_\_\_\_

- 3. Is English your first language (i.e., the language you consider your "native" language and the one in which you are most comfortable conversing in)?
  - a. No
  - b. Yes
- 4. Are you right- or left-handed?
  - a. Right-handed
  - b. Left-handed
- 5. How often do you play any sort of video game (on computer, console, phone, etc.)?
  - a. Never (e.g., less than once a month)
  - b. Rarely (e.g., once a week)
  - c. Sometimes (e.g., 2-3 times a week)
  - d. Frequently (e.g., 4-5 times a week)
  - e. Always (e.g., daily)

#### Cognitive Ability (SAT/ACT scores)

In the spaces below, please provide your highest SAT and/or ACT score. Note that this score will ONLY be used for research purposes and will be kept confidential. If you do not remember your score, please put a zero in the space for SAT score.

SAT score: \_\_\_\_\_

ACT score:

#### Math Domain Identification (Smith & White, 2001)

Using the following scale, please indicate the number that best describes how much you agree with each of the statements below.

12345Strongly<br/>disagreeModerately<br/>DisagreeNeither agree<br/>or disagreeModerately<br/>AgreeStrongly agree

- 1. Mathematics is one of my best subjects.
- 2. I have always done well in Math.
- 3. I get good grades in Math.
- 4. I do badly on tests of Mathematics (R).

Please indicate the number that best describes you for each of the statements below using the following scale:

1	2	3	4	5
Not at all		Somewhat		Very much

- 5. How much do you enjoy math-related subjects?
- 6. How likely would you be to take a job in a math-related field?
- 7. How important is Math to the sense of who you are?
- 8. How important is it to you to be good at Math?
- 9. Compared to other students, how good are you at math?
  - a. Very poor
  - b. Poor
  - c. About the same
  - d. Better than average
  - e. Excellent

## **APPENDIX J**

#### **Experimental Session Measures**

#### Metacognitive Activity (Bell, 2002)

For each of the items below, rate the extent to which you were thinking about these issues during the practice and performance trials today.

1	2	3	4	5
Never	Rarely	Sometimes	Frequently	Constantly

- 1. While practicing the simulation, I monitored how well I was learning its requirements.
- 2. I thought carefully about my performance on the previous trial before selecting what to study and practice.
- 3. As I performed in the practice trials, I evaluated how well I was learning the skills of the simulation.
- 4. When my methods were not successful, I experimented with different procedures for performing the task.
- 5. I considered the skills that needed the most work when choosing what to study and practice.
- 6. I tried to monitor closely the areas where I needed the most study and practice.
- 7. I noticed where I made the most mistakes during practice and focused on improving those areas.
- 8. I carefully determined what to study and practice in order to improve on weaknesses identified in previous trials.
- 9. I used my performance on the previous trial to revise how I would approach the task on the next trial.
- 10. I thought about new strategies for improving my performance.
- 11. I thought ahead to what I would do next to improve my performance.
- 12. I told myself things to encourage me to try harder.

#### Perceived Stereotype Threat (adapted from Ployhart, Ziegert, & McFarland, 2003)

The following series of questions ask about your perceptions and feelings regarding the experiment and radar control simulation task you just completed. Please answer each question honestly and openly to the best of your ability using the provided scale.

1	2	3	4	5
Strongly	Disagree	Neither agree	Agree	Strongly agree
disagree		or disagree		

- 1. People likely believe that I would perform poorly on this task because of my gender.
- 2. This task may have been easier for people of my gender. (R)

- 3. The experimenter expected me to do poorly on this task because of my gender.
- 4. I was not worried that anyone would draw conclusions about my abilities on this task based on my gender. (R)
- 5. Tasks like the one I just completed have been used to discriminate against people from my gender.
- 6. During the experiment, I wanted to show that people of my gender could perform well on the radar control simulation task.
- 7. A negative opinion exists about how members of my gender should perform on this type of task.

## Manipulation Check

Please indicate the extent to which you agree with each of the following statements using the provided scale.

1	2	3	4	5
Strongly disagree	Disagree	Neither agree	Agree	Strongly agree

- 1. The radar control task assesses skills related to mathematical ability.
- 2. People with high mathematical ability would likely do well on the radar control task.

## Declarative Knowledge (adapted from Bell, 2002)

The following is a test that asks questions about facts you may have learned about the radar control simulation. Please answer each question to the best of your ability.

## [Day 1 Test]

- 1. If a target's Response is Authorized, what is its likely Intent?
  - a. Military
  - b. Hostile
  - c. Civilian
  - d. Peaceful\*
- 2. Which of the following characteristics indicates that a target is a Submarine?
  - a. Speed =  $20 \text{ knots}^*$
  - b. Communication Time = 55 seconds
  - c. Direction of Origin = Orange Bay
  - d. Countermeasures = Jamming

- 3. A Maneuvering Pattern of Code Delta indicates the target is which of the following?
  - a. Air
  - b. Military\*
  - c. Surface
  - d. Civilian
- 4. A Green Beach Direction of Origin indicates the target is which of the following?
  - a. Unknown
  - b. Submarine
  - c. Peaceful\*
  - d. Military
- 5. If a target's Altitude/Depth is 10 feet, what is the Type of the target?
  - a. Air\*
  - b. Surface
  - c. Submarine
  - d. Unknown
- 6. If a target's Signal Strength is Indistinct, what Class does this suggest for the target?
  - a. Surface
  - b. Civilian
  - c. Military
  - d. Unknown\*
- 7. If a target's characteristics are Communication Time = 20 seconds and Speed = 50 knots, which of the following actions should you take?
  - a. Choose Intent as Peaceful
  - b. Choose Type as Surface
  - c. Get another piece of information
  - d. Choose Type as Air\*
- 8. A Communication Time of 52 seconds indicates that the target is likely:
  - a. Air
  - b. Surface\*
  - c. Submarine
  - d. Unknown
- 9. If a target's characteristics are Countermeasures = None and Maneuvering Pattern = Code Foxtrot, which of the following actions should you take?
  - a. Choose Class as Military
  - b. Choose Intent as Peaceful
  - c. Choose Class as Civilian\*
  - d. Choose Intent as Unknown

- 10. Which of the following targets should you make the final engagement decision to Clear?
  - a. Air, Military, Peaceful\*
  - b. Submarine, Civilian, Peaceful
  - c. Air, Civilian, Peaceful
  - d. Surface, Civilian, Hostile
- 11. If a target's Speed is 40 knots, what does this suggest about the target?
  - a. The target is Surface
  - b. The target is Civilian
  - c. The target is Air\*
  - d. The target is Military

## [Day 2 Test]

- 1. If a target's Identification Tag is Prince, what is its likely Intent?
  - a. Military
  - b. Hostile
  - c. Peaceful\*
  - d. Civilian
- 2. Which of the following characteristics indicates that a contact is a Surface target?
  - a. Speed = 20 knots
  - b. Direction of Origin = Blue Lagoon
  - c. Countermeasures = Jamming
  - d. Communication Time = 55 seconds\*
- 3. A Response that is Inaudible indicates the target is which of the following?
  - a. Military
  - b. Civilian
  - c. Unknown\*
  - d. Hostile
- 4. Jamming Countermeasures suggest that the target is which of the following?
  - a. Surface
  - b. Military\*
  - c. Submarine
  - d. Civilian
- 5. If a target's Speed is 31 knots, what is the Type of the target?
  - a. Surface\*
  - b. Air
  - c. Submarine
  - d. Unknown

- 6. Which of the following targets should you make the final engagement decision to Warn?
  - a. Air, Military, Hostile
  - b. Surface, Civilian, Peaceful
  - c. Air, Military, Peaceful
  - d. Submarine, Civilian, Peaceful\*
- 7. If a target's characteristics are Countermeasures = Inactive and Signal Strength = Weak, which of the following actions should you take?
  - a. Choose Class as Civilian
  - b. Get another piece of information\*
  - c. Choose Class as Military
  - d. Choose Intent as Peaceful
- 8. A Communication Time of 110 seconds indicates that the target is likely:
  - a. Air
  - b. Surface
  - c. Submarine\*
  - d. Unknown
- 9. If a target's characteristics are Speed = 35 knots and Altitude/Depth = 15 feet, which of the following actions should you take?
  - a. Choose Type as Surface
  - b. Choose Type as Air\*
  - c. Choose Class as Military
  - d. Choose Intent as Peaceful
- 10. If a target's Response = Invalid, this suggests that the target falls into which category?
  - a. Intent is Hostile\*
  - b. Class is Civilian
  - c. Type is Air
  - d. Class is Unknown
- 11. If a target's Direction of Origin is Orange Bay, what does this suggest about the target?
  - a. The target is Military
  - b. The target is a Submarine
  - c. The target is Hostile\*
  - d. The target is Peaceful

#### [Day 3 Test]

- 1. If a target's Maneuvering Pattern is Code Foxtrot, what is its likely Class?
  - a. Civilian\*
  - b. Military
  - c. Hostile
  - d. Peaceful

- 2. Which of the following characteristics indicates that a contact is an Air target?
  - a. Speed = 30 knots
  - b. Communication Time = 35 seconds\*
  - c. Maneuvering Pattern = Code Delta
  - d. Signal Strength = Weak
- 3. Which of the following targets should you make the final engagement decision to Mark?
  - a. Surface, Civilian, Hostile
  - b. Air, Civilian, Hostile
  - c. Submarine, Military, Peaceful
  - d. Submarine, Military, Hostile\*
- 4. Which of the following characteristics indicates that a contact is Hostile?
  - a. Maneuvering Pattern = Code Delta
  - b. Communication Time = 115 seconds
  - c. Response = Invalid\*
  - d. Countermeasures = Jamming
- 5. If a target's Speed = 35 knots, what is the Type of the target?
  - a. Air\*
  - b. Surface
  - c. Submarine
  - d. Unknown
- 6. If a target's Identification Tag is Tango, what Intent does this suggest for the target?
  - a. Civilian
  - b. Peaceful
  - c. Unknown
  - d. Hostile\*
- 7. If a target's characteristics are Signal Strength = Weak and Maneuvering Pattern = Code Delta, which of the following actions should you take?
  - a. Choose Intent as Hostile
  - b. Choose Class as Civilian
  - c. Choose Class as Military\*
  - d. Choose Intent as Peaceful
- 8. A Communication Time of 40 seconds indicates that the target is likely:
  - a. Air\*
  - b. Surface
  - c. Submarine
  - d. Unknown

- 9. If a target's characteristics are Direction of Origin = Orange Bay and Identification = Golf, which of the following actions should you take?
  - a. Choose Class as Military
  - b. Choose Type as Air
  - c. Choose Intent as Hostile
  - d. Get another piece of information\*
- 10. If a target's Direction of Origin is Blue Lagoon, this suggests that the target falls into which category?
  - a. Intent is Unknown\*
  - b. Class is Military
  - c. Intent is Peaceful
  - d. Class is Unknown
- 11. If a target's Speed = 0 knots, what does this suggest about the target?
  - a. The target is Civilian
  - b. The target is Air
  - c. The target is Surface
  - d. The target is a Submarine\*

#### **APPENDIX K**

Knowledge Structure Assessment Instructions (cf., Goldsmith et al., 1991)

#### [Page 1 Instructions]

Your task in this next exercise will involve judging the relatedness of pairs of concepts central to completing the Radar Control Simulation. In making these types of judgments, there are several ways to think about the items being judged.

For instance, two concepts might be related because they **share common features**, **frequently occur together**, **help you perform a task in the simulation**, or **serve a similar goal in the simulation**. While this kind of detailed analysis is possible, our concern is to obtain your initial impression of "overall relatedness" among the presented concepts. Therefore, please base your ratings on your first impression of relatedness.

The list of concepts important to operating the Radar Control Simulation are listed below; take a moment to look at the list and locate one or two highly related and unrelated pairs to give yourself an idea of what it means for two things to be highly versus not highly related. For example, why might **Identify contact Class as Civilian** and **Make decision to Warn contact** be related concepts?

Concepts					
<ol> <li>Identify contact type as Air</li> <li>Identify contact type as Surface</li> <li>Identify contact type as Submarine</li> <li>Identify contact class as Civilian</li> <li>Identify contact class as Military</li> <li>Identify contact intent as Peaceful</li> <li>Identify contact intent as Hostile</li> <li>Make decision to Clear contact</li> </ol>	<ul> <li>9) Make decision to Warn contact</li> <li>10) Make decision to Mark contact</li> <li>11) Gain/Lose Points</li> <li>12) Zoom out/Zoom In</li> <li>13) Monitor inner perimeter</li> <li>14) Monitor outer perimeter</li> <li>15) Find/engage pop-up targets</li> <li>16) Prioritize targets (engage targets</li> <li>likely to cross a perimeter first)</li> </ul>				
b) Make decision to creat contact	likely to cross a perimeter first)				

## [Page 2 Instructions]

In this exercise, a pair of concepts from the previous list will be presented on the screen along with a 9-point "relatedness" scale as shown in the example below. Using this scale, you are to indicate your judgment of relatedness for each pair by selecting the appropriate number on the scale. You can think of these numbers as points along a relatedness scale, with higher numbers representing greater relatedness. Thus, if you feel that the two concepts shown to you are not related at all, you would select "1" for that concept pair; if you feel the concepts are highly related, you would select "9" for the concept pair.

#### For example:

How related are... Dog and Pet?

Not at	0	0	0	0	0	0	0	0	0	Highly
all related	1	2	3	4	5	6	7	8	9	related

## [Page 3 Instructions]

Before beginning the rating task, here are a few important things to note when making your ratings:

- Try to use the full range of the rating scale to make your ratings.
  - Values near the ends of the scale (e.g., 1-3, 7-9) should be used if you are very certain about how two concepts are related to each other
  - Values from the middle of the scale (e.g., 4-6) should be used to reflect either medium relatedness or uncertainty about how the concepts relate
- Go with your "gut" when making ratings.
  - o It's best to make quick/intuitive judgments rather than deliberate on each pair
  - A good goal to shoot for is to spend no more than 5 seconds on each rating
- Do your best to provide an honest and accurate portrayal of how you believe these concepts are related to each other.
  - At the end of the experiment, you will receive a copy of your knowledge maps that you will be able to compare against others in the experiment to see how you match up—so the more honest you are, the better (and more interesting) your results will be!

On Days 2 and 3, participants also saw the following instruction:

- Base your relatedness ratings on what you've learned about the task up to this point.
  - Your ideas about how concepts in the task relate to one another may have changed since last time due to the extra practice you've had with the task
  - Make your judgments based on your current understanding of the task rather than trying to reproduce the ratings you provided previously

REFERENCES

#### REFERENCES

- Ackerman, P.L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence*, 10, 109-139.
- Ackerman, P.L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, 102, 3-27.
- Ackerman, P.L., Beier, M.E., & Boyle, M.O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, 131, 30-60.
- Ackerman, P.L., Bowen, K.R., Beier, M.E., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, 93, 797-825.
- Adams, J.W., & Hitch G.J. (1997). Working memory and children's mental addition. *Journal of Experimental Child Psychology*, 67, 21-38.
- Aiman-Smith, L., Scullen, S.E., & Barr, S.H. (2002). Conducting studies of decision making in organizational contexts: A tutorial for policy-capturing and other regression-based techniques. *Organizational Research Methods*, 5, 388-414.
- Anderson, J.R. (1982). Acquisition of a cognitive skill. Psychological Review, 89, 369-406.

Anderson, J.R. (1993a). Rules of the mind. Hillsdale, NJ: Erlbaum.

- Anderson, J. A., (1993b). Problem-solving and learning. American Psychologist, 48, 35-44.
- Anderson, J.R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 355-365.
- Anderson, J.R., & Lebiere, C. (1998). The atomic components of thought. Mahwah, NJ: Erlbaum.
- Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111, 1036-1060.
- Anderson, J.R., Reder, L.M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*, 221–256.
- 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.
- Ausubel, D.P. (1963). Cognitive structure and the facilitation of meaningful verbal learning. *Journal of Teacher Education, 14*, 217–221.

Baddeley, A.D. (1986). Working memory. Oxford, England: Oxford University Press.

Baddeley, A.D. (1992). Working memory. Science, 255. 556-559.

- Baddeley, A.D. (1997). *Human memory: Theory and practice*. East Sussex, England: Psychology Press.
- Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.
- Baddeley, A.D. (2001). Is working memory still working? American Psychologist, 56, 849-864.
- Baddeley, A.D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829-839,
- Baddely, A.D., & Hitch, G. (1974). Working memory. In G.A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.
- Baldwin, T.T., Ford, J.K, & Blume, B.D. (2009). Transfer of training 1988-2008: An updated review and agenda for future research. In G.P. Hodgkinson & J.K Ford (Eds.), *International review of industrial and organizational psychology* (Vol. 24, pp. 41-70).
- Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*, 83–100.
- Barrouillet, P., & Lépine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *Journal of Experimental Child Psychology*, *91*, 183–204.
- Bates, D., Maechler, M., Bolker, B. (2011). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4
- Baumeister. R.F., & Vohs, K.D. (2004). Handbook of self-regulation: Research, theory, and applications. New York: Guilford.
- Beier, M.E., & Ackerman, P.L. (2005). Working memory and intelligence: Different constructs. Reply to Oberauer et al. (2005) and Kane et al. (2005). *Psychological Bulletin*, 131, 72-75.
- Beilock, S.L., & Carr, T.H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science*, *16*, 101-105.
- Beilock, S.L., Gunderson, E.A., Ramirez, G., & Levine, S.C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 1860-1863.

- 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, 256-176.
- Bell, B.S. (2002). An examination of the instructional, motivational, and emotional elements of error training. (Doctoral Dissertation). Michigan State University, East Lansing, MI.
- Bell, B.S., Kanar, A.M., & Kozlowski, S.W.J. (2008). Current issues and future directions in simulation-based training in North America. *The International Journal of Human Resource Management*, 19 (8), 1416-1434.
- Bell, B.S., & Kozlowski, S.W.J. (2002). Adaptive guidance: Enhancing self-regulation, knowledge, and performance in technology-based training. *Personnel Psychology*, 55, 267-306.
- Bell, B.S., & Kozlowski, S.W.J. (2007). Advances in technology-based training. In S. Werner (Ed.), *Managing Human Resources in North America* (pp. 27-42). New York: Routledge.
- Bell, B.S., & Kozlowski, S.W.J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology*, 93, 296-316.
- Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. Journal of Experimental Social Psychology, 41, 174–181.
- Bereiter, C., & Scardamalia, M. (1985). Cognitive coping strategies and the problem of "inert" knowledge. IN S. Chipman, J.W. Segal, & R. Glaser (Eds.), *Thinking and learning skills: Vol. 2. Research and open questions* (pp. 65-80). Hillsdale, NJ: Erlbaum.
- Blascovich, J., Spencer, S.J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, *12*, 225–229.
- 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.
- Brodish, A.B., & Devine, P.G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology*, 45, 180-185.
- Brown, R.P., & Day, E.A. (2006). The difference isn't black and white: Stereotype threat and the race gap on Raven's Advanced Progressive matrices. *Journal of Applied Psychology*, *91*, 979-985.
- Brown, R.P., & Lee, M.N. (2005). Stigma consciousness and the race gap in college academic achievement. *Self & Identity*, *4*, 149–157.
- 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.
- Brunswik, E. (1952). The conceptual framework of psychology. International Encyclopedia of Unified Science, (Vol. 1, No. 10, pp. 655-768). Chicago, IL: The University of Chicago Press.
- Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models. Newbury Park, CA: Sage.
- Budd, D., Whitney, P., & Turley, K.J. (1995). Individual differences in working memory strategies for reading expository text. *Memory & Cognition, 23*, 735–748.
- Cable, D., & Judge, T. (1994). Pay preferences and job search decisions: A person-organization fit perspective. *Personnel Psychology*, 47, 317-348.
- Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? *Psychological Science*, *16*, 572-578.
- Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (1993). A theory of performance. In N. Schmitt & W.C. Borman (Eds.), *Personnel selection in organizations*. (pp. 35-70). San Francisco: Jossey-Bass Publishers.
- Cantor J., & Engle, R.W. (1993). Working-memory capacity as long-term memory activation: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1101–1114.
- Carpenter P.A., Just M.A., & Shell P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*, 404-431.
- Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell R.B. (1943). The measurement of adult intelligence. Psychological Bulletin, 40, 153–193.

Chandler, M. (1999). Secrets of the SAT [PBS Frontline]. Boston, MA: WGBH Studios.

Chase, W.G. & Simon, H.A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

- Chi, M.T.H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.
- Chi, M.T.H., Glaser, R., & Farr, M.J. (1988). The nature of expertise. Hillsdale, NJ: Erlbaum.
- Chi, M.T.H., Glaser, R., & Rees, E, (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence*. (pp. 7-75). Hillsdale, NJ: Erlbaum.
- Cloud, J. (2009). How stereotypes defeat the stereotyped. *TIME*. Retrieved from http://www.time.com/time/health/article/0,8599,1897009,00.html.
- Colby, S., Lee, S., Lewinger, J.P., & Bull, S. (2010). *pmlr: Penalized multinomial logistic regression*. R package version 1.0. http://CRAN.R-project.org/package=pmlr.
- Cole, B., Matheson, K., & Anisman, H. (2007). The moderating role of ethnic identity and social support on relations between well-being and academic performance. *Journal of Applied Social Psychology*, 37, 592-615.
- Conway A.R.A., Cowan N., Bunting M.F., Therriault D.J., & Minkoff S.R.B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–184.
- Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769-786.
- Crocker, J., Major, B., & Steele, C. (1998). Social stigma. In D. Gilbert, S. Fiske, and G. Lindzey (Eds.), *The handbook of social psychology*, (Vol. 2, 4th ed., pp. 504-553). Boston: McGraw-Hill.
- Croizet, J., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The Intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24,* 588-594.
- Croizet, J.-C., Després, G., Gauzins, M.-E., Huguet, P., Leyens, J.-P., & Méot, A. (2004). Stereotype threat undermines performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30*, 721–731.
- Cullen, M.J., Hardison, C.M., & Sackett, P.R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology*, *89*, 220-230.
- Cullen, M.J., Waters, S.D., & Sackett, P.R. (2006). Testing stereotype threat theory predictions for math-identified and non-math-identified students by gender. *Human Performance*, 19, 421-440.

- Daily, L.Z., Lovett, M.C., & Reder, L.M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Sciences*, *25*, 315–353.
- Daneman, M., & Merikle, P.M. (1996). Working memory and language comprehension: A metaanalysis. *Psychonomic Bulletin & Review, 3*, 422-433.
- Danaher, K., & Crandall, C.S. (2008). Stereotype threat in applied settings re-examined. *Journal* of Applied Social Psychology, 38, 1639-1655.
- Day, E., Winfred, A., & Gettman, D. (2001). Knowledge structures and the acquisition of a complex skill, *Journal of Applied Psychology*, 86, 1022-1033.
- Dearholt, D.W., & Schvaneveldt, R.W. (1990). Properties of pathfinder networks. In R.W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 1-30). Norwood, NJ: Ablex Publishing Corp.
- Deary, I.J, Strand S., Smith, P., & Fernandes, C. (2007) Intelligence and educational achievement. *Intelligence 35*, 13–21.
- de Freitas, S., & Neumann, T. (2009). The use of "exploratory learning" for supporting immersive learning in virtual environments. *Computers & Education, 52*, 343-352.
- Debowski, S., Wood, R.E., & Bandura, A. (2001). Impact of guided exploration and enactive exploration on self-regulatory mechanisms and information acquisition through electronic search. *Journal of Applied Psychology*, *86*, 1129–1141.
- DeShon, R.P., Kozlowski, S.W.J., Schmidt, A.M., Milner, K.R., & Weichmann, D. (2004). A multiple-goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology*, *89*, 1035-1056.
- Dorsey, D.W., Campbell, G.E., Foster, L.L., & Miles, D.E. (1999). Assessing knowledge structures: Relations with experience and posttraining performance. *Human Performance*, *12*, 31-57.
- Dunlosky, J., & Kane, M.J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology*, 60, 1227-1245.
- Engle, R.W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19-23.
- Engle, R.W., & Kane, M.J. (2004). Executive attention, working memory capacity, and a twofactor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 145–199). New York: Academic Press.

- Engle, R.W., Tuholski, S.W., Laughlin, J.E., & Conway, A.R.A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.
- Esposito, C. (1990). A graph-theoretic approach to concept clustering. In R.W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 89-100). Norwood, NJ: Ablex Publishing Corp.
- Espy, K.A., McDiarmid, M.M., Cwik, M.F., Stalets, M.M., Hamby, A., & Senn, T.F. (2004). The contribution of executive functions to emergent mathematics skills in preschool children. *Developmental Neuropsychology*, 26, 465–486.
- Faria, A.J. (1998). Business simulation games: Current usage levels—an update. *Simulation & Gaming, 29, 295-308.*
- Faria, A.J., & Nulsen, R. (1996). Business simulation games: Current usage levels a ten year update. *Developments in Business Simulation & Experiential Exercises, 23,* 22-28.
- Feldman Barrett, L., Tugade, M.M., & Engle, R.W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin*, 130, 553-573.
- Fernandez-Duque, D., Baird, J. A., & Posner, M. I. (2000). Executive attention and metacognitive regulation. *Consciousness and Cognition*, *9*, 288–307.
- Festinger, L. (1957). A theory of cognitive dissonance. Stanford, CA: Stanford University Press.
- Forbes, C.E., Schmader, T., & Allen, J.J.B. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of minorities under threat. *Social Cognition and Affective Neuroscience*, *3*, 253-261.
- Ford, J.K., & Kraiger, K. (1995). The application of cognitive constructs and principles to the instruction systems model of training: Implications for needs assessment, design and transfer. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and* organizational psychology (pp. 1–48). Chichester, United Kingdom: Wiley.
- Ford, J.K., Smith, E.M., Weissbein, D.A., Gully, S.M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology*, 83, 218-233.
- Frantz, C.M., Cuddy, A.J.C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, 30, 1611-1624.
- Frese, M., Albrecht, K., Altmann, A., Lang, J., Papstein, P.V., Peyerl, R., et al. (1988). The effects of an active development of the mental model in the training process:

Experimental results in a word processing system. *Behaviour and Information Technology*, 7, 295–304.

- Frey, M.C., & Detterman, D.K. (2004). Scholastic assessment or g? The relationship between the Scholastic Assessment Test and general cognitive ability. *Psychological Science*, 15, 373-378.
- Gagne, R.M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist, 39*, 377-385.
- Gawronski, B., & Bodenhausen, G.V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692–731.
- Geary, D.C., Hoard, M.K., Byrd-Craven, J., & DeSoto, M.C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology*, 88, 121-151.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*, 254–267.
- Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 284–313), London: Routledge.
- Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.
- Gigerenzer, G., Todd, P.M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gilhooly, K.J., Logie, R.H., Wetherick, N.E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition*, 21, 115-124.
- Glaser, R. (1990). The reemergence of learning theory within instructional research. *American Psychologist, 45, 29-39.*
- Goldsmith, T.E., & Davenport, D.M. (1990). Assessing structural similarity of graphs. In R.W. Schvaneveldt (Ed.), *Pathfinder associative networks: Studies in knowledge organization* (pp. 75-87). Norwood, NJ: Ablex.
- Goldsmith, T.E., Johnson, P.J., Acton, W.H. (1991). Assessing structural knowledge. *Journal of Educational Psychology*, 83, 88-96.

- Goldstein, I.L., & Ford, J.K. (2002). *Training in organizations: Needs assessment, development, and evaluation* (4<sup>th</sup> ed.). Belmont, CA: Wadsworth.
- Gonzales, P.M., Blanton, H., & Williams, K.J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin, 28*, 659-670.
- Good, C., Aronson, J., & Harder, J.A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology*, 29, 17-28.
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24, 645-662.
- Gottfredson, L.S. (1997) Why g matters: The complexity of everyday life. *Intelligence 24*, 79–132.
- Grand, J.A., & Kozlowski, S.W.J. (in press). Seven basic principles for adaptability training in synthetic learning environments. In C. Best, G. Galanis, J. Kerry, & R. Sottilare (Eds.), *Fundamental issues in defence training and simulation*. Aldershot, UK: Ashgate.
- Grand, J.A., Ryan, A.M., Schmitt, N., & Hmurovic, J. (2011). How far does stereotype threat reach? The potential detriment of face validity in cognitive ability testing. *Human Performance, 24*, 1-28.
- Grier, R.A., Warm, J.S., Dember, W.N., Matthews, G., Galinsky, T.L., Szalma, J.L., & Parasuraman, R. (2003). The vigilance decrement reflects limitations in effortful attention, not mindlessness. *Human Factors*, 45, 349–359.
- Grimm, L.R., Markman, A.B., Maddox, W.T., & Baldwin, G.C. (2009). Stereotype threat reinterpreted as a regulatory mismatch. *Journal of Personality and Social Psychology*, 96, 288-304.
- Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: L. Erlbaum Associates.
- Halpern, D.F., Benbow, C.P., Geary, D.C., Gur, R.C., Hyde, J.S., & Gernsbacher, M.A. (2007).
  The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8, 1-51.
- Harkins, S. G. (2006). Mere effort as the mediator of the evaluation–performance relationship. *Journal of Personality and Social Psychology*, 91, 436–455.

- Harrison, L.A., Stevens, C.M., Monty, A.N., & Coakley, C.A. (2006). The consequences of stereotype threat on the academic performance of white and non-white lower income college students. *Social Psychology of Education*, *9*, 341-357.
- Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.
- Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, *21*, 2409-2419.
- Hernstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York, NY: Free Press.
- Hess, T.M., Auman, C., Colcombe, S.J., & Rahhal, T.A. (2003). The impact of stereotype threat on age differences in memory performance. *Journal of Gerontology: Psychological Sciences, 58*, 3–11.
- Higgins, E.T. (1987). Self-discrepancy theory: A theory relating self and affect. *Psychological Review*, *94*, 319–340.
- Hunt, E. (1994). Problem-solving. In R.J. Sternberg (Ed.), *Thinking and problem-solving*. (pp. 215-232). San Diego, CA: Academic Press.
- Hyde, J.S., Fennema, E., & Lamon, S.J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107,139-153.
- Ifenthaler, D., Masduki, I., & Seel, N.M. (2011). The mystery of cognitive structure and how we can detect it: Tracking the development of cognitive structures over time. *Instructional Science*, *39*, 41-61.
- Ifenthaler, D., & Seel, N.M. (2005). The measurement of change: Learning-dependent progression of mental models. *Technology, Instruction, Cognition and Learning, 2*, 317–336.
- Imbo, I., & Vandierendonck, A. (2007). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology*, 96, 284-309.
- Interlink. (2011). FAQs. Retrieved July 12, 2011 from http://interlinkinc.net/FAQ.html
- Inzana, C.M., Driskell, J.E., Salas, E., &Johnston, J.H. (1996). Effects of preparatory information on enhancing performance under stress. *Journal of Applied Psychology*, *81*, 429-435.
- Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*, 365-371.

- Inzlicht, M., McKay, L., & Aronson, J. (2006). Stigma as ego depletion: How being the target of prejudice affects self-control. *Psychological Science*, *17*, 262–269.
- Iran-Nejad, A. (1990). Active and dynamic self-regulation of learning processes. *Review of Educational Research, 60*, 573-602.
- Ivancic, K., & Hesketh, B. (2000). Learning from error in a driving simulation: Effects on driving skill and self-confidence. *Ergonomics*, 43, 1966–1984.
- Jaeggi, S.M., Buschkuehl, M., Jonides, J., & Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 6829-6833.
- Jamieson, J.P., & Harkins, S.G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, 93, 544–564.
- Jensen, A.R. (1998). The g factor: The science of mental ability. Westport, CT: Praeger.
- Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: The influence of emotion regulation. *Journal of Experimental Psychology: General, 137*, 691-705.

Johnson-Laird, P. (1983). Mental models. Cambridge, MA: Harvard University Press.

- Jonassen, D.H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hilsdale, NJ: Lawrence Erlbaum.
- Josephs, R.A., Newman, M.L., Brown, R.P., & Beer, J.M. (2003). Status, testosterone, and human intellectual performance: Stereotype threat as status concern. *Psychological Science*, *14*, 158–163.
- Just, M.A., & Carpenter, P.N. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149.
- Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K.J. Holyoak & R.G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*. (pp. 267-293). New York: Cambridge University Press.
- Kamouri, A.L., Kamouri, J., & Smith, K.H. (1986). Training by exploration: Facilitating the transfer of procedural knowledge through analogical reasoning. *International Journal of Man-Machine Studies*, *24*, 171-192.
- Kane, M.J., Bleckley, M.K., Conway, A.R.A., & Engle, R.W. (2001). A controlled-attention view of WM capacity. *Journal of Experimental Psychology: General*, 130, 169–183.

- Kane, M.J., & Engle, R.W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, 132, 47–70.
- Kane, M.J., Hambrick, D.Z., & Conway, A.R.A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131, 66-71.
- Kane, M.J., Hambrick, D.Z., Tuholski, S.W., Wilhem, O., Payne, T.W., & Engle, R.W. (2004). The generality of working memory: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189– 217.
- Kane, M.J., Conway, A.R.A, Hambrick, D.Z., & Engle, R.W. (2007). Variation in working memory as variation in executive attention and control. In A.R.A. Conway, C. Jarrold, M.J. Kane, A. Miyake, & J.N. Towse (Eds.), *Variation in working memory* (pp. 21-48). Oxford, England: Oxford University Press.
- Kanfer, R., & Ackerman, P.L. (1989). Training the human information processor. In I.L. Goldstein (Ed.), *Training and development in organizations* (pp. 121-182). San Francisco: Jossey-Bass.
- Karren, R.J., & Barringer, M.W. (2002). A review of the policy-capturing methodology in organizational research: Guidelines for research and practice. *Organizational Research Methods*, *5*, 337-361.
- Keifer, A.K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, *43*, 825-832.
- Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, 47, 193–198.
- Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' math performance. *British Journal of Educational Psychology*, *77*, 323-338.
- Kieras, D., & Meyer, D.E. (1997). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction*, *12*, 391–438.
- Kieras, D.E., Meyer, D.E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor performance. In P. Shah & A. Miyake (Eds.), *Models of working memory: Mechanisms of*

*active maintenance and executive control* (pp. 183–223). Cambridge, England: Cambridge University Press.

- Koch, S.C., Müller, S.M., & Sieverding, M. (2008). Women and computers. Effects of stereotype threat on attribution of failure. *Computers & Education*, *51*, 1795-1803.
- Koenig, A.M., & Eagly, A.H. (2005). Stereotype threat in men on a test of social sensitivity. *Sex Roles, 52*, 489-496.
- Koenig, K.A., Frey, M.C., & Detterman, D.K. (2008). ACT and general cognitive ability. *Intelligence*, *36*, 153-160.
- König, C.J., Bühner, M., & Mürling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance*, *18*, 243-266.
- Kosslyn, S.M., & Rosenberg, R.S. (2004). *Psychology: The brain, the person, the world* (2<sup>nd</sup> ed.). Boston, MA: Pearson.
- Koubek, R.J., Clarkston, T.P., & Calvez, V. (1994). The training of knowledge structures for manufacturing tasks: An empirical study. *Ergonomics*, 37, 765–780.
- Kozlowski, S.W.J., Gully, S.M., Brown, K.G., Salas, E., Smith, E.M., & Nason, E.R. (2001). Effects of training goals and goal orientation traits on multidimensional training outcomes and performance adaptability. *Organizational Behavior and Human Decision Processes*, 85, 1-31.
- Kozlowski, S.W.J., Toney, R.J., Mullins, M.E., Weissbein, D.A., Brown, K.G., & Bell, B.S. (2001). Developing adaptability: A theory for the design of integrated-embedded training systems. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research* (Vol. 1, pp. 59-123). Amsterdam: JAI/Elsevier Science.
- Kraiger, K., Ford, J.K, & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*, 311-328.
- Kraiger, K., Salas, E., & Cannon-Bowers, J.A. (1995). Measuring knowledge organization as a method for assessing learning during training. *Human Factors*, 37, 804-816.
- Kray, L.J., Galinksy, A.D., & Thompson, L. (2002). Reversing the gender gap in negotiations: An exploration of stereotype regeneration. *Organizational Behavior and Human Decision Processes*, 87, 386-409.
- Kray, L.J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology*, 80, 942–958.

- Kristof-Brown, A.L., Jansen, K.J., & Colbert, A.E. (2002). A policy-capturing study of the simultaneous effects of fit with jobs, groups, and organizations. *Journal of Applied Psychology*, *87*, 985-993.
- Kyllonen, P.C. (1996). Is working memory capacity Spearman's g? In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49-75). Hillsdale, NJ: Erlbaum.
- Kyllonen P.C., & Christal R.E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence*, *14*, 389-433.
- Lievens, F., Reeve, C.L., & Heggestad, E.D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology*, 92, 1672-1682.
- Lesko, A.C., & Corpus, J.H. (2006). Discounting the difficult: How high math identified women respond to stereotype threat. *Sex Roles, 54*, 113-125.
- Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, *71*, 1092-1107.
- Leyens, J.-P., Désert, M., Croizet, J.-C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin, 26*, 1189-1199.
- Lipshitz, R., Levy, D.L., & Orchen, K. (2006). Is this problem to be solved? A cognitive schema of effective problem-solving. *Thinking and Reasoning*, *12*, 413-430.
- Loewenstein, M.A., & Speltzer, J.R. (2000). Formal and informal training: Evidence from the NLSY. *Research in Labor Economics*, 18, 403-438.
- Loman, N.L., & Mayer, R.E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology*, 75, 402-412.
- Maas, C.J., & Hox, J.J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*, 86-92.
- Major, B., Spencer, S.J., Schmader, T., Wolfe, C.T., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin, 24*, 34-50.
- Marshall, H. (1996). Recent and emerging theoretical frameworks for research on classroom teaching: Contributions and limitations [Special Issue]. *Educational Psychologist*, 31(3/4).

- Marshall, N., & Glock, M.D. (1979). Comprehension of connected discourse: A study into the relationships between the structure of text and information recalled. *Reading Research Quarterly*. 14, 10-56.
- Marx, D.M. & Goff, P.A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology*, *44*, 645-657.
- Marx, D.M., & Stapel, D.A. (2006a). 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.
- Marx, D.M., & Stapel, D.A. (2006b). It's all in the timing: Measuring emotional reactions to stereotype threat before and after taking a test. *European Journal of Social Psychology*, *36*, 687–698.
- Marx, D.M., Stapel, D.A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparison under threat. *Journal of Personality and Social Psychology*, *88*, 432-446.
- Mayer, R.E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist*, *59*, 14-19.
- McDaniel, M.A., & Schlager, M.S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction*, 7, 129–159.
- McNamara, D.S., & Scott, J.L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*, 10-17.
- Medin, D.L., Ross, N.O., Atran, S. Cox, D., Coley, J., Proffitt, J.B., & Blok, S. (2006). Folkbiology of freshwater fish. *Cognition*, 99, 237-273.
- Mendoza-Denton, R., Purdie, V., Downey, G., & Davis, A. (2002). Sensitivity to status-based rejection: Implications for African-American students' college experience. *Journal of Personality and Social Psychology*, 83, 896–918.
- Messick, S. (1984). Abilities and knowledge in educational achievement testing: The assessment of dynamic cognitive structures. In B.S. Plake (Ed.), *Social and technical issues in testing: Implications for test construction and usage* (pp. 156-172). Hillsdale, NJ: Erlbaum.
- Meyer, B.J.F., Brandt, D.M., & Bluth, G.J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, *16*, 72-103.
- Meyer, B.J.F., & Rice, E. (1982). The interaction of reader strategies and the organization of text. *Text*, *2*, 155-192.

- McGlone, M.S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. Journal of Applied Developmental Psychology, 27, 486-493.
- McKay, P.F., Doverspike, D., Bowen-Hilton, D., & Martin, Q. D. (2002). Stereotype threat effects on the Raven Advanced Progressive Matrices scores of African Americans. *Journal of Applied Social Psychology*, *32*, 767–787.
- McKay, P.F., Doverspike, D., Bowen-Hilton, D, & McKay, Q.D. (2003). The effects of demographic variables and stereotype threat on black/white differences in cognitive ability test performance. *Journal of Business and Psychology*, *18*, 1-14.
- Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 63, 81-97.
- Muraven, M., & Baumeister, R.F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*, 247–259.
- Nagy, P. (1984). Cognitive structure and the spatial metaphor. In P. Nagy (Ed.), *The representation of cognitive structure* (p. 1-11). Toronto, Canada: Ontario Institute for Studiesin Education.
- Neisser, U., Boodoo, G., Bouchard, T.J., Boykin, A.W., Brody, N., Ceci, S.J.,...Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77-101.
- Newell, A. (1990). Unified theories of cognition. Cambridge, MA: Harvard University Press.
- Nguyen, H-H., O'Neal, A., & Ryan, A.M. (2003). Relating test-taking attitudes and skills and stereotype threat effects to the racial gap in cognitive ability test performance. *Human Performance*, *16*, 261-293.
- Nguyen, H-.H., & Ryan, A.M. (2008). Does stereotype threat test affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314-1334.
- Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, *83*, 44–59.
- O'Brien, L.T., & Crandall, C.S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29*, 782–789.
- Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H-M. (2005). Working memory and intelligence—their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 61-65.
- Phillips, D.C. (1998). How, why, what, when, and where: Perspectives on constructivism in psychology and education. *Issues in Education*, *3*, 151–194.

- Ployhart, R.E., Ziegert, J.C., & McFarland, L.A. (2003). Understanding racial differences on cognitive ability tests in selection contexts: An integration of stereotype threat and applicant reactions research. *Human Performance*, 16, 231-259.
- Prawat, R.S. (1989). Promoting access to knowledge, strategy, and disposition in students: A research synthesis. *Review of Educational Research*, *59*, 1-41.
- Pressing, J. (1999). The referential dynamics of cognition and action. *Psychological Review*, 106, 714-747.
- Pressley, M., Snyder, B.S., Levin, J.R., Murray, H.G., & Ghatala, E.S. (1987). Perceived Readiness for Examination Performance (PREP): Produced by initial reading of text and text containing adjunct questions. *Reading Research Quarterly*, 22, 219-236.
- Pronin, E., Steele, C., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology*, 40, 152-168.
- Quinn, D.M., Kahng, S.K., & Crocker, J. (2004). Discreditable: Stigma effects of revealing a mental illness history on test performance. *Personality and Social Psychology Bulletin*, 30, 803-815.
- Quinn, D.M., & Spencer, S.J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, 57, 55–71.
- R Development Core Team (2012). *R: A language and environment for statistical computing*. R Statistical Computing, Vienna, Austria. www.R-project.org.
- Radvansky, G.A., & Zacks, R.T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 940-953.
- Reder, L.M., & Anderson, J.R. (1980). A comparison of texts and their summaries: Memorial consequences. *Journal of Verbal Learning & Verbal Behavior, 19*, 121-134.
- Rieman, J. (1996). A field study of exploratory learning strategies. ACM Transactions on Computer-Human Interaction, 3, 189-218.
- Rivers, C. (2007). Shock jocks wield dangerous 'stereotype threat.' *WomensEnews*. Retrieved from http://www.womensenews.org/story/media-stories/070416/shock-jocks-wield-dangerous-stereotype-threat.
- Rohde, T.E, & Thompson L.A. (2007) Predicting academic achievement with cognitive ability. *Intelligence 35*, 83–92.

- Rosen, V.M., & Engle, R.W. (1997). The role of working memory capacity in retrieval. *Journal* of Experimental Psychology: General, 126, 211–227.
- Rosen, V.M., & Engle, R.W. (1998). Working memory capacity and suppression. Journal of *Memory and Language, 39*, 418–436.
- Rotundo, M., & Sackett, P.R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, *87*, 66-80.
- Rouse, W.B., & Morris, N.M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100,* 349-363.
- Rowe, A.L., Cooke, N.J., Hall, E.P., & Halgren, T.L. (1996). Toward an on-line knowledge assessment methodology: Building on the relationship between knowing and doing. *Journal of Experimental Psychology: Applied*, 2, 31–47.
- Royer, J.M., Tronsky, L.N., Chan, Y., Jackson, S.J., & Marchant, H. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology*, 24, 181-266.
- Rumelhart, D.E., & Norman, D.A. (1978). Accretion, tuning and restructuring: Three models of learning. In R. L. Klatzky & J. W. Cotton (Eds.), *Semantic factors in cognition* (pp. 37– 53). Hillsdale, NJ: Lawrence Erlbaum.
- Rydell, R.J., Rydell, M.T., & Boucher, K.L. (2010). The effect of negative performance stereotypes on learning. *Journal of Personality and Social Psychology*, 99, 883-896.
- Rydell, R.J., Shiffrin, R.M., Boucher, K.L., Van Loo, K., & Rydell, M.T. (2010). Stereotype threat prevents perceptual learning. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 14042-14047.
- Sackett, P.R., Hardison, C.M., & Cullen, M.J. (2004). On interpreting stereotype threat as accounting for African American–White differences on cognitive tests. *American Psychologist, 59*, 7–13.
- Sackett, P.R., & Ryan, A.M. (2012). Concerns about generalizing stereotype threat research findings to operational high stakes testing. In M. Inzlicht & T. Schmader (Eds.), *Stereotype threat* (pp. 249-263). Oxford Press.
- Sackett, P.R., Schmitt, N., Ellingson, J.E., & Kabin, M.B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, 56, 302–318.

- Scherbaum, C.A., & Ferreter, J.M. (2009). Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods*, *12*, 347-367.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's Math performance. *Journal of Experimental Social Psychology*, *38*, 194-201.
- Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science, 19*, 14-18.
- Schmader, T., Forbes, C.E., Zhang, S., & Berry Mendes, W. (2009). A metacognitive perspective on the cognitive deficits experience in intellectually threatening environments. *Personality and Social Psychology Bulletin, 35*, 584-596.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440–452.
- Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. Sex Roles, 50, 835-850.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336-356.
- Schmidt, F.L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*, 187-210.
- Schoenfeld, A.H., & Herrmann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 484–494.
- Schuelke, M.J., Day, E.A., McEntire, L.E., Boatman, P.R., Boatman, J.E., Kowollik, V., & Wang, X. (2009). Relating indices of knowledge structure coherence and accuracy to skill-based performance: Is there utility in using a combination of indices? *Journal of Applied Psychology*, 94, 1076-1085.
- Schunn C.D., & Reder L.M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General, 130*, 59-76.
- Schwartz, D.L., & Bransford, J.D. (1998). A time for telling. *Cognition and Instruction*, 16, 475-522.
- Seel, N. M. (1999). Educational diagnosis of mental models: Assessment problems and technology-based solutions. *Journal of Structural Learning and Intelligent Systems*, 14, 153–185.

- Seibt, B., & Förster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology*, *87*, 38–56.
- Shavelson, R.J. (1972). Some aspects of the correspondence between content structure and cognitive structure in Physics education. *Journal of Educational Psychology*, 63, 225– 234.
- Shavelson, R.J. (1974). Methods for examining representations of a subject-matter structure in student memory. *Journal of Research in Science Teaching*, *11*, 231–249.
- Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317-1323.
- Shiffrin, R.M., & Lightfoot, N. (1997). Perceptual learning of alphanumeric-like characters. *Psychology of Learning and Motivation, 36*, 45-81.
- Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Shih, M., Pittinsky, T.L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*, 80-83.
- Shih, M., Pittinsky, T.L., & Trahan, A. (2006). Domain-specific effects of stereotypes on performance. *Self and Identity*, *5*, 1-14.
- Shimamura, A.P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition*, *9*, 313-323.
- Simon, D., & Simon, H.A (1978). Individual differences in solving physics problems. In R. Siegler (Ed.), *Children's .thinking: What develops?* (pp. 324-348). Hillsdale, NJ: Erlbaum..
- Simon, H.A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129-138.
- Simon, H. A. (1957). Model of man: Social and rational. New York: Wiley.
- Simon, H.A. (1974). How big is a chunk? Science, 183, 482-488.
- Simon, H.A. (1990). Invariants of human behavior. Annual Review of Psychology, 41, 1-19.
- Smith, J.L., & White, P.H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement*, 61, 1040-1057.

- Smith, J. (2004). Understanding the process of stereotype threat: A review of meditational variables and new performance goal directions. *Educational Psychology Review*, *16*, 177-206.
- Smith, J. (2006). The interplay among stereotypes, performance-avoidance goals, and women's math performance expectations. *Sex Roles*, *54*, 287-296.
- Smith, J.L., Sansone, C., & White, P.H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology*, *99*, 99-114.
- Smith, M.E., McEvoy, L.K., & Gevins, A. (1999). Neurophysiological indices of strategy development and skill acquisition. *Cognitive Brain Research*, *7*, 389-404.
- Smith-Jentsch, K.A., Mathieu, J.E., & Kraiger, K. (2005). Investigating linear and interactive effects of shared mental models on safety and efficiency in a field setting. *Journal of Applied Psychology*, *90*, 523–535.
- Snijders, T.A.B. & R.J. Bosker (1993). Standard errors and sample sizes in two-level research. *Journal of Educational Statistics*, 18, 237-260.
- 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.
- Stangor, C. (2000). (Ed.). Stereotypes and prejudice. Philadelphia, PA: Psychology Press.
- Stangor, C., & Lange, J. (1994). Mental representations of social groups: Advances in conceptualizing stereotypes and stereotyping. Advances in Experimental Social Psychology, 26, 357-416.
- Steffe, L.P., & Gale, J. (Eds.). (1995). Constructivism in education. Mahwah, NJ: Erlbaum.
- Steele, C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613-629.
- Steele, C.M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.
- Steele, C.M., & Aronson, J. (1998). Stereotype threat and the test performance of academically successful African Americans. In C. Jencks & M. Phillips (Eds.), *The Black–White test* score gap (pp. 401–427). Washington, DC: Brookings.
- Steele, C.M., & Davies, P.G. (2003). Stereotype threat and employment testing: A commentary. *Human Performance, 16*, 311-326.

- Steele, C.M., Spencer, S.J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), Advances in experimental social psychology (pp. 379–440). San Diego, CA: Academic Press.
- Sternberg, R.J., Conway, B.E., Ketron, J.L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology*, 41, 37-55.
- Sternberg, R.J., & Grigorenko, E.L. (2004). Intelligence and culture: How culture shapes what intelligence means, and the implications for a science of well-being. *Philosophical Transactions of the Royal Society B*, 359, 1427-1434.
- Steyvers, M., & Tennenbaum, J.B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29, 41-78.
- Stone, J. (2002). Battling doubt by avoiding practice: The effect of stereotype threat on self-handicapping in white athletes. *Personality and Social Psychology Bulletin, 28*, 1667-1678.
- Stone, J., Lynch, C.I., Sjomeling, M., & Darley, J.M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, 77, 1213-1227.
- Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology*, 44, 445-452.
- Stricker, L.J., & Ward, W.C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology*, 34, 665-693.
- Stricker, L.J., & Ward, W.C. (2008). Stereotype threat in applied settings re-examined: A reply. Journal of Applied Social Psychology, 38, 1656-1663.
- Summers, G.J. (2004). Today's business simulation industry. *Simulation & Gaming, 35,* 208-241.
- Sweller, J., Mawer, R.F., & Ward, M.R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General*, 112, 639-661.
- Taber, K.S. (2000). Multiple frameworks? Evidence of manifold conceptions in individual cognitive structure. *International Journal of Science Education & Training*, *22*, 399–417.
- te Nijenhuis J., van Vianen A.E.M., & van der Flier H. (2007) Score gains on *g*-loaded tests: No *g*. *Intelligence 35*, 283–300.

- Todd, P.M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, 16, 167-171.
- Turner, M.L., & Engle, R.W. (1989). Is working memory task dependent? *Journal of Memory and Language, 28*, 127–154.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.
- Unsworth, N., Heitz, R.P., Schrock, J.C., & Engle, R.W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505.
- van Merriënboer, J.J.G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17, 147-177.
- von Hippel, W., von Hippel, C., Conway, L., Preacher, K. J., Schooler, J. W., & Radvansky, G. A. (2005). Coping with stereotype threat: Denial as an impression management strategy. *Journal of Personality and Social Psychology*, 89, 22-35.
- Wagner, R.K., & Sternberg, R.J. (1984). Alternative conceptions of intelligence and their implications for education. *Review of Educational Research*, *54*, 179-223.
- Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles*, *41*, 219-240.
- Walton, G.M., & Cohen, G.L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467.
- Watts, D.J. (1999). *Small worlds: The dynamics of networks between order and randomness*. Princeton, NJ: Princeton University Press.
- Watts, D.J., & Strogatz, S.H. (1998). Collective dynamics of "small-world" networks. *Nature*, 393, 440–442.
- Weaver, J.L., Bowers, C.A., Salas, E., & Cannon-Bowers, J.A. (1995). Networked simulations: New paradigms for team performance research. *Behavioral Research Methods*, *Instruments*, & Computers, 27, 12–24.
- Webber, S.S., Chen, G., Payne, S.C., Marsh, S.M., & Zacarro, S.J. (2000). Enhancing team mental model measurement with performance appraisal practices. *Organizational Research Methods*, *3*, 307-322.
- Weiss, H.M. (1990). Learning theory and industrial psychology. In M.D. Dunnnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology*. Palo Alto, CA: Consulting Psychologists Press.

- Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology*, 51, 59–91.
- Wexley, K.N., & Latham, G.P. (1991). *Developing and training human resources in organizations*. New York: Harper Collins.
- Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology*, *44*, 792-799.
- Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognition and Affective Neuroscience*, *2*, 12–19.
- Yeung, N.C.J., & von Hippel, C. (2008). Stereotype threat increases the likelihood that female drivers in a simulator run over jaywalkers. *Accident Analysis & Prevention, 40,* 667-674.
- Yopyk, D.J.A., & Prentice, D.A. (2005). Am I an athlete or a student? Identity salience and stereotype threat in student-athletes. *Basic and Applied Social Psychology*, *27*, 329-336.
- Zentall, S.S. (1990). Fact-retrieval automatization and math problem solving by learning disabled, attention disordered, and normal adolescents. *Journal of Educational Psychology*, *82*, 856–865.