

THE GOAL OF DISRUPTING DISTRACTION: INVESTIGATING *HOW*
STEREOTYPE THREAT DISRUPTS WORKING MEMORY
VIA THE DUAL PROCESS OF CONTROL

by

Amber Cherie Ferris

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

of

Master of Science

in

Psychological Science

MONTANA STATE UNIVERSITY
Bozeman, Montana

May, 2012

©COPYRIGHT

by

Amber Cherie Ferris

2012

All Rights Reserved

APPROVAL

of a thesis submitted by

Amber Cherie Ferris

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency and is ready for submission to The Graduate School.

Dr. Colleen F. Moore

Approved for the Department of Psychology

Dr. Jessi L. Smith

Approved for The Graduate School

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Amber Cherie Ferris

May, 2012

TABLE OF CONTENTS

| | |
|--|----|
| 1. INTRODUCTION | 1 |
| Stereotype Threat Defined | 2 |
| Stereotype Threat: The Process | 4 |
| Stereotype Threat Effects on Cognitive Resources | 10 |
| The Stroop Task | 15 |
| Goal Maintenance | 15 |
| Response Competition | 16 |
| Stereotype Threat and the Stroop Task | 18 |
| Project Overview | 22 |
| 2. STUDY 1 METHODS | 27 |
| Participants | 27 |
| Procedure and Design | 27 |
| Materials | 28 |
| Working Memory Capacity Measure | 29 |
| The Stroop Task | 29 |
| Independent Variables | 30 |
| Stereotype Threat Manipulation | 30 |
| Control Condition | 31 |
| List Proportion Congruency Manipulation | 31 |
| Dependent Variables | 32 |
| Overall Stroop Effects | 32 |
| Domain Identification | 32 |
| 3. STUDY 1 RESULTS | 33 |
| The Distraction Hypothesis | 34 |
| The Response Competition Hypothesis | 36 |
| 4. STUDY 1 DISCUSSION | 38 |
| Item Specific Proportion Congruency | 42 |
| Reactive and Proactive Control Strategies | 48 |
| Study 2 Project Overview | 51 |
| 5. STUDY 2 METHODS | 59 |
| Participants | 59 |
| Procedure and Design | 59 |

TABLE OF CONTENTS - CONTINUED

| | |
|--|--------|
| Materials | 60 |
| Working Memory Capacity Measure..... | 61 |
| The Counting Stroop Task | 61 |
| Independent Variables | 62 |
| Stereotype Threat Manipulation | 62 |
| Control Condition | 62 |
| Item Specific Proportion Congruency Manipulation..... | 63 |
| Primary Dependent Variables | 63 |
| Overall Stroop Effects..... | 63 |
| Mostly Congruent Item and Mostly Incongruent Item Stroop Effects | 64 |
| Item Specific Proportion Congruency Effects | 64 |
| Secondary Dependent Variables | 64 |
| Domain Identification | 65 |
| Task Rating Survey..... | 65 |
| Perceived Competence..... | 66 |
| Task Interest..... | 66 |
| Future Task Interest | 66 |
| Value | 66 |
| Task Involvement..... | 66 |
| Future Interest in Math..... | 66 |
| Achievement Goal Questionnaire Revised | 67 |
| Mastery-Approach Goal..... | 67 |
| Mastery-Avoidance Goal | 67 |
| Performance-Approach Goal | 67 |
| Performance-Avoidance Goal..... | 67 |
| Exam Concerns Survey..... | 68 |
| Stereotype Concerns | 68 |
| Individual Impression Concerns | 68 |
| Group Impression Concerns | 68 |
| 6. STUDY 2 RESULTS..... | 69 |
| Preliminary Analysis..... | 69 |
| Primary Dependent Variables | 76 |
| Overall Stroop Effects..... | 76 |
| Mostly Congruent Item and Mostly Incongruent Item Stroop Effects | 77 |
| Item Specific Proportion Congruency Effects | 79 |
| Secondary Dependent Variables | 79 |
| Task Rating Survey..... | 83 |
| Perceived Competence..... | 83 |
| Task Interest..... | 83 |
| Future Task Interest | 83 |

TABLE OF CONTENTS - CONTINUED

| | |
|--|-----|
| Value | 84 |
| Task Involvement..... | 84 |
| Future Interest in Math..... | 84 |
| Achievement Goal Questionnaire Revised | 84 |
| Mastery-Approach Goal..... | 85 |
| Mastery-Avoidance Goal | 85 |
| Performance-Approach Goal | 85 |
| Performance-Avoidance Goal..... | 85 |
| Exam Concerns Survey | 86 |
| Stereotype Concerns | 86 |
| Individual Impression Concerns | 86 |
| Group Impression Concerns | 87 |
| 7. STUDY 2 DISCUSSION..... | 88 |
| 8. GENERAL DISCUSSION | 94 |
| Limitations and Future Directions | 96 |
| Implications..... | 99 |
| REFERENCES CITED..... | 107 |

LIST OF TABLES

| Table | Page |
|--|------|
| 1. Correlations between WMC, primary dependent variables and secondary dependent variables..... | 73 |
| 2. Mean, standard deviation, t- values, and significance levels for WMC and Stroop Effects within the Mostly Congruent List Condition and Mostly Incongruent List..... | 74 |
| 3. Mean, standard deviation, t- values, and significance levels for WMC and Stroop Effects within the Stereotype Threat Condition and Control Condition..... | 74 |
| 4. Mean, standard deviation, F values, and significance levels for WMC and Stroop Effects within the Four Experimental Groups..... | 75 |
| 5. Mean, standard deviation, t- values, and significance levels for all Secondary Dependent Variables separated into their subscales and DIM within the Mostly Congruent List Condition and Mostly Incongruent List Condition..... | 80 |
| 6. Mean, standard deviation, t- values, and significance levels for all Secondary Dependent Variables separated into their subscales and DIM within the Stereotype Threat Condition and Control Condition..... | 81 |
| 7. Mean, standard deviation, F- values, and significance levels for all Secondary Dependent Variables separated into their subscales and DIM within all four experimental groups. | 82 |

LIST OF FIGURES

| Figure | Page |
|--|------|
| 1. Predicted Values of Stroop Effects in Error Rates by Working Memory Capacity (WMC) for Stereotype Threat and Control participants within Mostly Congruent (left) or Mostly Incongruent (right) lists..... | 36 |
| 2. Average Stroop Effect Error Rates for the Mostly Congruent list and Mostly Incongruent list. | 76 |
| 3. Average Error Rates on Mostly Congruent Items for the Mostly Congruent list and Mostly Incongruent list. | 78 |

ABSTRACT

Few researchers would argue with the notion that stereotype threat inflicts its deleterious effects by depleting working memory resources. Even fewer can explain the true nature of stereotype threat's disruptive nature on working memory. Based upon the dual process framework of control (Kane & Engle, 2003) we examined whether or not stereotype threat produces its negative effects by creating distraction that debilitates individual's ability to maintain the goal of the task and one's ability to resolve competition between conflicting responses. In Study 1, 187 men engaged in a working memory capacity measure (Ospan) and then engaged in the color Stroop task under mostly congruent list conditions or mostly incongruent list conditions. The Stroop task was described as a test of verbal ability (stereotype threat condition) or not (control). The results demonstrated that stereotype threat disrupts working memory by interfering with an individual's ability to maintain the task goals increasing Stroop error effects, especially for those with lower working memory capacity. Study 2 was designed to replicate Study 1 using a different stereotyped population and to investigate what cognitive control strategy operates while under stereotype threat. Specifically, we examined if stereotype threat increases the use of reactive control strategies whereby individuals do not actively maintain the task goal, make an error that indicates conflict, which in turn reactivates the goal for a short while. In Study 2, 144 women were given the automated version of the Reading Span task to assess working memory capacity, then completed the counting Stroop task under mostly congruent list conditions or mostly incongruent list conditions with item specific congruency manipulated within both lists. The Stroop task was described as a quantitative ability task (stereotype threat condition) or not (control). Overall errors and reaction time were again assessed. In addition, errors and reaction times for mostly congruent items, mostly incongruent items, and item-specific effects were assessed. The results did not provide significant evidence as to what type of control strategy operates under stereotype threat. Implications of how stereotype threat disrupts working memory and how interventions can address these effects are discussed.

INTRODUCTION

‘Although the idea of stereotype threat interfering with performance by creating a disruptive mental workload is largely admitted, evidence of the disruptive nature is nevertheless still lacking.’

Croizet, Despres, Guazines, Huguet, Leyens, & Meot (2004), p. 722

Through the simple act of having African Americans indicate their race by marking a box prior to taking a standardized test, Steele and Aronson (1995) famously showed that a stereotype threat is created that negatively influences performance on the test. Since Steele and colleagues’ initial series of studies, scholars have replicated and generalized these findings to other groups, a multitude of domains, and discovered an array of consequential outcomes (Brown & Day, 2006; Levy, 1996; Seibet & Forster, 2004; Smith & White, 2002). In addition to stereotype threat’s negative effects on African Americans, it has been shown to hinder performance of various social groups in a plethora of domains ranging from women’s math performance (Smith & White, 2002; Spencer, Steele, & Quinn, 1999), men’s performance on tests of verbal ability (Kiefer & Shih, 2006), and even men’s athletic ability (i.e., African American men when the sport is said to measure athletic intelligence; i.e., Caucasian men when the sport is said to measure athletic ability) (Stone, Lynch, Sjomeling, & Darley, 1998).

Given that previous research has established the robustness of stereotype threat, scholars have turned toward understanding the mechanisms that determine and predict how a stigmatized individual will perform in a given domain in various “stereotype threat” situations (Beilock & Carr, 2005; Brodish & Devine, 2009; Croizet et al., 2004; Inzlicht & Kang, 2010; Jamieson & Harkins, 2007; Johns, Inzlicht, & Schmader, 2008;

Osbourne, 2001; Schmader & Johns, 2003; Smith, 2004; for review see Shapiro & Neuberg, 2007). Although much research has attempted to empirically demonstrate the mechanisms underlying stereotype threat effects, the results are still mixed. In large part, much of the stereotype threat mechanism research has been theoretical and/or subject to measurement shortcomings (see Smith, 2004 for a review).

The goal of the current project was to use a well-established measure of cognitive processes to investigate just how stereotype threat affects working memory as well as to provide evidence that the effects are due to both situational factors and cognitive resources. This project expected to contribute to the stereotype threat literature by moving beyond just stating that stereotype type threat affects performance by depleting cognitive resources and provide evidence as to how the working memory system is disrupted.

Stereotype Threat Defined

In over 20 years of research on stereotype threat, the field of social psychology has refined the definition of stereotype threat, shed light on many important boundary conditions for when stereotype threat effects are most likely to occur, and delved into documenting and eliminating possible mechanisms for how stereotype threat inflicts damage. Social identity threat, from which stereotype threat derives, states that individuals associate themselves with a collective group whereby features of the group begin to be integrated into their self-concept forming part of their identity (Steele, Spencer, & Aronson, 2002). Social identity threat occurs when individuals feel that the

collective that they identify with has been negatively judged, creating worry for the group as a whole. In this case, individuals can feel a threat to their social identity that is not associated with any known stereotype. For example, if a woman highly identifies with her role as a psychology student and a feature of her identity in psychology is brought into question (i.e. psychology is not a science) this may create a threat resulting a need to distance herself from the group or the situation in order to reduce the psychological discomfort associated with the threat.

Within the umbrella of social identity threat there are multiple subsets of threat that also affect individual's behavior and identity because individuals can identify with multiple groups. One of those possible groups may include a stereotyped group. Even if a stereotype is not endorsed or acted upon, they are known. When a person identifies with a known stereotyped group they are susceptible to stereotype threat. Stereotypes such as "women are bad at math," produce negative performance effects when individuals are experiencing stereotype threat (Smith & White, 2002; Spencer, Steele, & Quinn, 1999). However, these stereotype threat effects are not because of a fear of the group being negatively evaluated, but instead of the fear that they as an individual will be negatively judged based on the stereotype. Fortunately, research has established conditions that predict when individuals are more likely to experience a threat to their social identity.

Researchers have targeted several boundary conditions that make negative performance effects due to stereotype threat more likely to occur. In addition, these conditions also predict the extent to which these negative stereotype threat effects are experienced between individuals and situations. One such boundary condition is the

extent to which a person identifies with the domain in question (Smith, 2006; Smith & Johnson, 2006). For example, research shows women who are highly identified with Math, Science and Engineering (MSE) are more likely to detect situational gender disparities when experiencing stereotype threat while watching a video of a “conference setting” (Murphy, Steele, & Gross, 2007). This was evidenced in the female participants’ ability to recall more details of the MSE related items of the “conference” video when the female gender was under represented (Murphy, et al., 2007). Similarly, women who are highly identified with being a woman are more likely to recognize and remember words that are related to gender prejudice, suggesting individuals that are highly identified with a group are much more sensitive, divert attention to, and are more aware of possible threats to their identity when in a threatened state (Davies, Spencer, Quinn, & Gerhardstein, 2002; Kaiser, Vick, & Major, 2006). Other boundary conditions include that the effects are more pronounced when the test is difficult (Smith & White, 2002), a person cares about success in the domain (Smith & Johnson, 2006; Smith & White, 2001), is strongly identified with their (stereotyped) group (Schmader, 2002), experience greater levels of stigma consciousness (Brown & Pinel, 2003), and have a lower working memory capacity (Kaiser & Hagiwara, 2011).

Stereotype Threat: The Process

Although over 300 studies have been published on stereotype threat, what the exact mechanisms and how these mechanisms inflict their damage still eludes researchers (for review see Shapiro & Neuberg, 2007; Smith, 2004). Initially, researchers assumed

anxiety about confirming the stereotype was the primary underlying mechanism for stereotype threat (Steele, 1997), but years of empirical studies have shown mixed and inconclusive results (for review see Smith, 2004). At times anxiety causes a level of arousal that helps performance and at other times a level that reduces performance, depending on the predominate response of the task (Harkins, 2006; Jamieson & Harkins, 2007; Osbourne, 2001). The “mere effort” explanation states that when an individual is faced with the possibility of evaluation, they may experience a boost in performance due to anxiety, if the anxiety is at an optimal level (Harkins, 2006). However, Osborne (2001) found evidence to the contrary; increased anxiety decreased performance of stereotyped groups (i.e., race and gender) on a task that was evaluative of their abilities. One possible reason for this discrepancy within the literature could be due to the complexity at which stereotype threats operate. For example, many studies (Beilock & Carr, 2005; Inzlicht & Kang, 2010; Jamieson & Harkins, 2007; Osbourne, 2001) have examined only a single mechanism which could mediate results. However, there could be multiple mechanisms which could contribute to the detriments of stereotype threat (Schmader, Johns, & Forbes, 2008; Smith, 2004).

One such model that examines and describes how multiple mechanisms could be working as determinants of stereotype threat effects was first proposed by Smith (2004). This approach was termed the “Stereotyped Task Engagement Process” (STEP) model. STEP, as proposed by Smith (2004), framed a motivational account of negative performance effects postulating that the phenomenological experience of stereotype threat occurs at multiple levels and through multiple pathways triggered by achievement

goals. The STEP model postulates that in the domain of achievement settings competence and performance on a task is based on the goals the individual has adopted. The STEP model has since provided evidence that individuals experiencing stereotype threat tend to adopt performance-avoidance (PAV) goals (individual's wanting to avoid failing or performing poorly compared to the performance of others) (Brodish & Devine, 2009), which sets in motion a series of factors such as self-handicapping (Keller, 2002), test anxiety (Aronson et al., 1999), low risk taking (Carr & Steele, 2010), reduced task interest, lower task involvement, task value, and lower perceived competence (Smith, Sansone, & White, 2007). Providing support for the STEP account of multiple mechanisms Brodish and Devine (2009), found that PAV goals and worry about one's performance mediated the negative performance effects of stereotype threat on women's math performance.

Although STEP is a motivational account of stereotype threat effects and pertains mostly to explicit manipulations of stereotype threat, researchers have more recently begun to investigate cognitive accounts of stereotype threat effects (Beilock & Carr, 2005; Inzlicht & Kang, 2010; Schmader & Johns, 2003; Stone & McWhinnie, 2008). The "Integrated Process Model" or *working memory depletion* account, has been one such approach and has received abundant attention in the stereotype threat literature (Schmader, Forbes, Zhang, & Mendes, 2009; Schmader, Johns, & Forbes, 2008). The Integrated Process Model of working memory depletion posits that there may be three separate, but interrelated mechanisms, which all create negative effects on performance of stereotyped tasks. In addition, depending on how the stereotype threat is manipulated

and thus triggered, may affect cognitive resources differently (Stone & McWhinnie, 2008). The first process mechanism is due to physiological arousal that the situation creates causing people under stereotype threat to expend all their cognitive resources on reducing anxiety rather than devoting it to the task. The second process mechanism that was suggested by the Integrated Processes Model, states that the person under stereotype threat actively monitors his or her performance. This constant monitoring of performance also referred to as error monitoring causes a decrease in performance. Having adopted an inappropriate strategy for the task, individuals try to avoid making any errors, which causes them to drastically slow down. Ironically, this causes them to commit more errors due to being unable to complete the task. The final process mechanism proposed was that stereotype threat causes the individual to engage in self-regulation. Self-regulation results in the individual trying to control their affective responses, which in turn directs effort toward suppressing negative thoughts and emotions rather than the task.

It is likely though that both the motivational and cognitive accounts for the effects of stereotype threat are accurate, and that the manner in which the threat arises is based upon the type of manipulation used, which determines which process is triggered. For instance, the “dual process” model theorizes that the blatant versus subtle nature of the stereotype threat manipulation may result in a STEP process versus a working memory (WM) process (Stone & McWhinnie, 2008). When the environment contains blatant stereotype threats, such as when individuals are explicitly told that the group they identify with is known to perform poorly on the task, individuals do not have to determine if there is a possibility of a stereotype threat. Rather, when a blatant stereotype manipulation is

used, which has shown a greater effect size in racial stereotype threat studies (for meta analyses see Nguyen & Ryan, 2008) participants become focused on monitoring their performance, such that they adopt strategies which may reduce their success. When a blatant form of stereotype threat is “in the air” individuals expend energy on regulating physiological responses and adopt a prevention focus. As the STEP model postulates (Smith, 2004), this adoption of a PAV goal occurs in order for the stereotyped individual to avoid mistakes and the appearance of failure so as not to confirm the stereotype presented. However, this prevention focus however, can cause an individual to adopt ineffective strategies disrupting their ability to perform optimally on the particular task.

Conversely, the Dual Process Model states that the nature of subtle manipulations (i.e., indicating gender or race before a test) create a situation in which the individual tries to determine if there is a possibility of threat, causing a cognitive load which can have relatively more negative effects on cognitive resources than blatant manipulations (for review see Nguyen & Ryan, 2008; Stone & McWhinnie, 2008). In addition, subtle manipulations have greater effect sizes in gender stereotype threat studies (for meta analyses see Nguyen & Ryan, 2008), tend to be used more often, and are more likely encountered in day-to-day life. The cognitive load that is experienced when participants are exposed to a subtle threat may be due to entering a state of hypervigilance that is indicative of error monitoring, at which time they slow down to ensure they are not making errors (Kaiser et al., 2006).

In support for this dual process model of stereotype threat, Stone and McWhinnie (2008) found evidence suggesting that, when stereotype threat is manipulated with subtle

cues individuals expend cognitive resources on determining that there is a possibility of threat and are thus unable to attend to the task at hand because they are distracted. These results suggest that individuals exposed to subtle stereotype threats may also become distracted. It is theorized that this distraction is due to the individual's attempt to determine if a stereotype threat is operating and if performance on the task will be judged based on the stereotype, resulting in the inability to maintain the goal of the task (Kaiser et al., 2006; Murphy, et al., 2007; Nygugen & Ryan, 2008). For example, research has provided evidence that when participants experiencing stereotype threat are subliminally presented with a series of words pertaining to their identity, to an illness or injury, and to household objects, participants are more likely to falsely report having "seen" words related to their social identity (Kaiser et al., 2006). This research suggests that even when stereotype threat is subtly triggered through subliminal presentation individuals are attending to the environment and trying to determine if there is a possibility of bias. This in turn causes them to expend their cognitive resources on determining if stereotype threat is present. Distraction created by the presence of stereotype threat was also supported by Murphy et al., (2007), whose results suggested women in situations which utilize Math, Science, and Engineering may be especially sensitive to the situational cues which make their gender salient when in group settings in these fields.

The literature on motivational and cognitive accounts both converge on the notion that multiple mechanisms are at play in stereotype threat effects. Focusing on the cognitive account of multiple mechanisms, the current project was designed to test just *how* working memory is affected by stereotype threat and how it may determine what

process is engaged or hindered when subtle manipulations are used. Specifically, it was designed to test whether or not stereotype threat creates distraction from the task goal or slowing of responses to decrease errors.

Stereotype Threat Effects on Cognitive Resources

The vast research on stereotype threat has concentrated on studying performance effects on three different categories of tasks: cognitive ability tasks (math, verbal, memory), sensorimotor tasks (tasks that involve fluid movement, automatic behavior and controlled behavior), and social interaction tasks (maintaining a social interaction with someone when a stereotype has been presented) (for review see Schmader, Johns, & Forbes, 2008). These three tasks have one common factor: they all require effortful processing, controlled attention, and active self-regulation (Schmader, Johns, & Forbes, 2008). In order to engage in these processes individuals utilize their working memory (WM) system. The WM system contains four different structures: a phonological loop, a visiospatial sketchpad, an episodic buffer, and an executive control function, (Baddeley, 1998). It has been suggested that the executive control structure is the primary structure tested in regards to tasks that require effortful processing, controlled attention, and self-regulation. In addition, when stereotype threat effects are researched in the framework of cognitive explanations or when cognitive tools are used, it is the aforementioned tasks that are utilized, all of which require the effortful processing and controlled attention that the WM system contains. The cognitive ability tasks utilized in stereotype threat research, such as Raven's Advance Matrices, (Croizet, et al., 2004; Regner, et al., 2010), dual tasks

requiring attention to two different stimuli (Beilock, Jellison, Rydell, McConnell, & Carr, 2006), and Stroop tasks (Carr & Steele, 2010; Inzlicht & Kang, 2010), require the use of WM and intelligence.

A majority of these conflict tasks, which require participants to maintain a task goal while attempting to suppress a habitual automatic response. Thus, these conflict tasks require active maintenance of the task goal and competition resolution when the goal is actively maintained. Such tasks also greatly correlate with trait levels of working memory capacity (WMC) (Kane & Engle, 2003), due to the executive control function in this system, which holds the information currently being used at the time as well as manipulating old and new information in order to carry out the task. Conflict tasks in the cognitive literature include the Antissacade task (a person has to focus on the middle of the screen, wait for a cue to be presented on one side, and divert attention to the opposite side in order to view the stimuli), Go-No Go tasks (changing instructions during the task), and resolving Garden Path Sentences (sentences that converge on multiple meanings, where individuals generally predict the most likely). Performance in each of these tasks correlate with WMC.

Due to the recent attention the working memory depletion account of stereotype threat effects has garnered, scholars have begun to investigate how outside situational factors affect the individual's ability to effectively utilize their WM. Cognitive research has shown that trait levels of WMC affect performance (Kane & Engle, 2003), such that those lower in WMC are more prone to increased intrusion errors due to failures in goal maintenance and increased reaction times due to response competition resolution.

Stereotype threat research has also shown that trait levels of WMC have a predictive value in determining the performance level of individuals under stereotype threat (Beilock, 2008; Beilock & Carr, 2005; Beilock et al., 2006; Regner et al., 2010, Schmader & Johns, 2003). For example, a series of studies conducted by Schmader and Johns (2003) found a striking relationship between WMC and stereotype threat effects that supported the working memory depletion account.

The first study measured WMC with the operation-span (Ospan) and the reading-span (Rspan) tasks in order to determine if there were differences between pre- and post - WMC levels when women were threatened with a math-related gender stereotype (Schmader & Johns, 2003). This study found that, although both the control group (men) and the experimental group (women) participants reported feeling concern that they would be judged based on their gender, only women showed decreases in WMC after performing the stereotyped task. These results were also replicated and generalized in the second study using a different population and stereotype; Latino/a's and intelligence. These two studies provided support that WM is depleted when an individual is experiencing stereotype threat and that WMC plays a vital role in negative performance effects. In addition, the third and final study provided evidence that WMC mediated women's poor performance on a challenging math test when experiencing stereotype threat (Schmader & Johns, 2003). Thus, these three studies found evidence that stereotype threat depletes cognitive resources, resources needed to perform at optimal levels, and that it was WMC that mediated these effects.

However, just knowing that cognitive resources are depleted and that this depletion mediates negative performance effects does not explain what WM process is more adversely affected by this depletion: the ability to maintain the goal or response conflict resolution. What WM process is negatively affected when stereotype threat is triggered? The previously outlined series of studies provided evidence that WMC mediates negative stereotype threat performance by depleting cognitive resources and creating a mental workload. Stereotype threat negatively affects WM, but how and what process is being affected; a person's ability to maintain the task goal or successfully resolving conflict between responses? And what role, if any, do trait levels of WMC play in this process? The current series of studies aimed to move past just stating that cognitive resources are depleted and examine just what WM process stereotype threat triggers that may contribute to stereotype threat effects.

Debate has stirred in the stereotype threat literature regarding whether or not trait levels of WMC affect performance when under stereotype threat and what levels are greater predictors of performance; high or low WMC. The "choking under pressure" literature suggests that when experiencing stereotype threat individuals with higher levels of WMC have greater decreases in performance on math questions, specifically difficult math questions that would require the most working memory resources, as opposed to those with lower levels of WMC (Beilock & Carr, 2005; Beilock et al., 2006; Beilock, Rydell, & McConnell, 2007). It was suggested that performance anxiety due to stereotype threat has a greater influence on individuals who would generally be more likely to display superior performance on tasks which place greater demands on WMC because of

the depletion of cognitive resources. Similarly it has been shown that experts, for example collegiate golfers, when under stereotype threat, show greater decreases in putting performance compared to novices (Beilock et al., 2006). Decreases in performance on a task that the participants are highly skilled at is attributed to “choking under pressure,” where it is stated that processes that are usually automatic or well learned become degraded because of a focus on the stereotype threat.

In contrast, other research suggests that stereotype threat effects are more pronounced among those lower in WMC (Kaiser & Hagiwara, 2011; Regner et al., 2010). Specifically, research has presented evidence that while under gender stereotype threat, women engineering students experience greater difficulties on the Raven Advance Matrices than women who were not under stereotype threat and a control group of men (Regner et al., 2010). However, these results depended on trait levels of WMC, whereby only women lower in WMC as assessed with Reading Span, experienced negative performance effects under stereotype threat. These results suggest that individuals with higher working memory may be immune to stereotype threat effects. As these two reviews illustrate, there are mixed results in the literature as to what the exact role trait WMC plays in stereotype threat effects on performance. Due to these conflicting hypotheses the current research was designed to investigate the role of trait levels of WMC on stereotype threat effects as well.

The Stroop Task

In the Stroop task a participant views a series of trials consisting of color words and is instructed to quickly state the ink color as opposed to the word itself. These trials are composed of either congruent items, viewed in the same ink color as the color word, or incongruent items, viewed in an ink color different from the word (i.e., the word red viewed in blue letters where the correct response is “blue”). This task requires the participant to maintain the goal of stating the ink color while simultaneously suppressing the habitual response to automatically read the word, which places a high demand on WMC (Kane & Engle, 2003). The Stroop task allows researchers to distinguish between participant’s inability to actively maintain the task goal (state the color of the word not the word itself) and their ability to successfully resolve response competition (the conflict between the controlled response of stating the ink and the habitual word reading response), by varying the proportion of congruent and incongruent trials within the list. By varying the trial proportion congruency, goal maintenance and response resolution can be dissociated by measuring changes in error rate and reaction times (RT) to congruent and incongruent conditions within each congruency proportion (Kane & Engle, 2003). Stroop effects are calculated as differences between incongruent RTs or error rates minus congruent RTs or error rates.

Goal Maintenance

Failures in goal maintenance (sometimes called “goal neglect”) occur when the participant fails to maintain the goal of stating the ink color rather than reading the word.

This occurs more frequently in lists made of mostly congruent trials. For example, in a list that is mostly made up of congruent trials, a participant may begin to neglect the goal and default to word reading. However, when a rare incongruent trial (“red” in blue ink) is encountered that participant will commit an error because the goal was not actively maintained (saying red when the response should have been blue).

Failures in goal maintenance are exhibited by interference errors (stating the word on incongruent trials) and facilitated reaction times (faster responses on congruent trials presumably due to reading the word). Interference errors are computed by taking the average error score on incongruent trials and subtracting the average error score on neutral trials (trials in which a random word such as dog are presented). Facilitation reaction time is computed by taking the average reaction time on neutral trials and subtracting the average reaction time on congruent trials. Stereotype threat effects on goal maintenance in the Stroop task would test if, while under stereotype threat, people have greater difficulty maintaining the goal due to distraction and therefore experience more errors (interference errors) and faster congruent responses (facilitation reaction time).

Response Competition

Response competition occurs when a participant whom is accurately maintaining the task goal experiences a slowing in reaction time due to difficulty suppressing the habitual response of reading the word. For example, in the Stroop task this would be evidenced in slower reaction times when a person tries to suppress word reading (saying red when the word is in an incongruent color). This would only occur if the participant

actively maintains the task goal, because the habitual response of reading the word and the goal appropriate ink color response are in “competition.”

Response competition is found in interference reaction time (individuals slow down because of the conflict between the habitual response to read the word and the goal to state the color) in lists that are composed of mostly incongruent trials. Interference reaction time is computed by taking the average reaction time for incongruent trials and subtracting the average reaction time on neutral trials (Incongruent – Neutral). For example, an individual would have a slower reaction time responding “red” when the word is blue versus responding “red” to a word that is neutral. Participants that actively maintain the task goal but have difficulty suppressing the word reading response will then have a larger average reaction time. Thus, stereotype threat effects may manifest as response competition deficits if individuals under stereotype threat respond slower, because of monitoring their performance, than individuals not under stereotype threat because they are monitoring their performance and experience difficulties resolving conflict.

In order to establish with greater certainty how and what WM process is impaired by stereotype threat, it is useful to employ well validated measures of cognitive performance such as the Stroop task (Stroop, 1935). The Stroop task correlates with trait levels of WMC and measures conflict in the form of how easily an individual can maintain the goal of the task as well as suppress their habitual responses (Hutchison, 2011; Kane & Engle, 2003). Thus, the Stroop task is an ideal tool to determine if individuals under stereotype threat experience decreases in their ability to maintain task

goals due to distraction (*distraction hypothesis*), if they slow down to monitor their performance (*response competition hypothesis*), or both (*dual processes hypothesis*). It also can be employed in multiple forms (traditional color Stroop task or counting Stroop task) and described in a way as a “test” that is diagnostic of a particular ability to assess these mechanisms while under stereotype threat. This is important in regards to the stereotype threat literature because past research has generally employed the use of surveys given after the stereotype-relevant task has been administered possibly disrupting the stereotype threat process. By describing the Stroop task as the actual stereotype-relevant task it allows the current project to investigate stereotype threat while it is occurring rather than before or after a “threatened” task (Inzlicht & Kang, 2010).

Individual differences in working memory capacity are also made evident in the Stroop task (Kane & Engle, 2003). Research has established that the higher an individual WMC, the less susceptible they are to failures in goal maintenance and response competition (Kane & Engle, 2003). Essentially, a person’s WMC plays a positive predictive value in determining Stroop task performance. Due to the abundance of evidence that Stroop task performance is moderated by working memory and that stereotype threat affects both Stroop task performance and working memory, the current study was designed to answer how.

Stereotype Threat and the Stroop Task

Past stereotype threat research using the Stroop task has shown that individuals under stereotype threat experience greater Stroop effects in reaction time (Carr & Steele,

2010; Inzlicht & Kang, 2010). Specifically, after completing a difficult math task highlighting gender differences favoring men, women experienced greater Stroop effects in reaction time when exposed to the stereotype than those who were not (Inzlicht & Kang, 2010). In addition, utilizing an electroencephalograph (EEG) to measure event related potentials (ERPs) defined as being indicative of the loss of executive control, it was shown that the women who had been in the gender stereotype threat condition inefficiently monitored their math performance. However, this study did not delineate between whether participants were experiencing greater response competition or goal maintenance by investigating both errors and reaction time. The dual processes also could not be distinguished in Inzlicht and Kang's (2010) study due to the fact that they had only two stimuli (red and green) and only investigated RTs, which does not allow the dissociation between the dual processes. In addition, this study measured Stroop task performance after the stereotype relevant task had been completed not while participants believed they were performing a stereotype relevant task. Consequently, due to these limitations Inzlicht and Kang's (2010) study was unable to provide evidence as to what mechanism of working memory or how working memory is actually affected while under stereotype threat. Instead only assumptions about what might possibly be occurring while under stereotype threat can be speculated.

Another study by Carr and Steele (2010), which utilized the Stroop task as a mediating mechanism to measure working memory depletion to examine risk aversion, found similar results. It was found that women in a stereotype threat condition were more risk averse and produced greater Stroop effects in reaction time. However, this study also

could not delineate between these dual processes due to the fact that they only used a 50% proportion congruency and only investigated reaction times, and again they only used the Stroop task as a post-test measure (Carr & Steele, 2010). Therefore, it is unknown if stereotype threat interferes with individuals ability to maintain task goals in the face of distraction or overcome conflict between two competing responses. In addition, because of the nature of these studies it is also unclear whether or not stereotype threat interferes with one or both of the dual processes, or if it is just an artifact of stereotype threat spilling over into other domains (Inzlicht & Kang, 2010).

Even though neither studies (Carr & Steele, 2010; Inzlicht & Kang, 2010) did not examined the dual processes of control, they do support that performance on the Stroop task is affected by stereotype threat and that these effects can spill over into other tasks. In addition, the study by Regner et al., (2010) suggests WMC does predict performance on a stereotype relevant task. However, measuring trait levels of WMC and describing the Stroop task as the stereotype-relevant task has not been employed even though it has been well established in the cognitive literature that WMC plays a predictive role in Stroop task performance (Kane & Engle, 2003).

The goal of the current project, specifically Study 1, was to examine the dual processes of control through varying trial proportion congruency of Stroop lists using congruent, incongruent, and neutral trials. In addition, Study 1 was also designed to establish how trait levels of WMC predict performance on the Stroop task while a person is under stereotype threat. The aim of the study was to investigate how the dual processes

of control may offer insight into how depleted cognitive resources play a role under stereotype threat.

One other advantage of using the Stroop task is that it can also be used to investigate the type of performance strategy the participants employ. The Stroop task allows for an investigation of how stereotype threat creates a decreased ability to allocate attentional control to the current task and which performance strategies they engage in that contributes to (un)successful performance, depending on their trait levels of WMC. Study 1 manipulated the list-wide proportion congruency (i.e., several different items are congruent or incongruent), in order to establish if stereotype threat interferes with the ability to maintain the task goal, resolve response competition, or both. In contrast, Study 2 manipulated item specific congruency (ISPC), the proportion with which specific items are congruent or incongruent in mostly congruent/incongruent lists, in order to investigate whether participants use reactive and/or proactive control strategies to tackle the task during performance (Braver, Gray, & Burgess, 2007). Unlike list-wide proportion congruency, which distinguishes between sustained goal maintenance and response competition resolution, ISPC involves specific critical items being either mostly congruent or mostly incongruent and should solely influence reactive control strategies. ISPC effects are computed as difference scores between mostly congruent items and mostly incongruent items in error rates and reaction times. These effects are evidenced as faster responding to mostly congruent items and greater intrusion errors on mostly incongruent items. By manipulating the ISPC one can investigate the type of control

strategy individuals utilize when under stereotype threat in order to complete the task and if individual differences in WMC predict which strategy is used.

Project Overview

In two studies, the current project was designed to test how WM is actually affected when experiencing stereotype threat and to move beyond just stating that it is depleted or disrupted by testing which process stereotype threat impairs. Study 1 set out to establish whether WM is affected through failures in goal maintenance (*distraction hypothesis*), decreases in the ability to resolve conflict (*response competition hypothesis*), or both (*dual process hypothesis*) by varying list-wide proportion congruency in the Stroop task (Study 1). In addition, Study 2 was designed to investigate what cognitive control strategy operates while individuals are experiencing stereotype threat by determining if they utilize proactive or reactive control strategies by varying the ISPC.

Specifically, Study 1 was designed to investigate three hypotheses by using the a male population and the stereotype that men perform poorly on tasks of verbal ability (Kiefer & Shih, 2006). The first hypothesis was that stereotype threat creates distraction, because participants are trying to determine if there is a possibility of stereotype threat, which increases intrusion errors due to the inability to actively maintain the task goal (*distraction hypothesis*). The second hypothesis was that stereotype threat would interfere with a participant's ability to resolve conflict between two competing responses causing slower reaction times (*response competition hypothesis*). The last hypothesis was that stereotype threat interferes with both processes (*dual process hypothesis*). This was

accomplished by manipulating the proportion congruency of the list and if they were under stereotype threat or not.

First, it was predicted that while under stereotype threat participants would experience greater difficulties maintaining the goal of the task in the mostly congruent list due to distraction (*distraction hypothesis*), as observed through increased intrusion errors. Second, it was predicted that while under stereotype threat participants would experience greater response competition in mostly incongruent lists due to their inability to resolve conflict and being hypervigilant to the task, leading them to slow down so that they can monitor their behavior (*response competition hypothesis*). The final prediction for Study 1 was that individual differences in WMC would contribute to these effects such that participants lower in WMC would exhibit both greater response competition and decreased goal maintenance due to their inability to devote a higher level of attentional control to the Stroop task relative to those who have the cognitive capacity to avoid such detriments.

The goal of Study 2 was to replicate these findings with a different group (women) and a different stereotype (math) as well as extend findings to uncover the type of strategy use that occurs under stereotype threat (Spencer et al., 1999). Specifically, Study 2 examined if women under stereotype threat engage in more reactive control strategies, which results in more errors due to relying on stimulus response contingencies. In line with Hutchison's (2011) prior study on Stroop effects and item-specific proportion congruency, it was predicted that when under stereotype threat women would utilize reactive control strategies which would be exhibited by increased intrusion errors overall

due to relying on learned associations based on the irrelevant task dimension. In addition, women experiencing stereotype threat were predicted to show a decreased reaction time to congruent stimuli, especially on mostly congruent lists and an increased reaction time to incongruent items. It was also expected that this pattern would repeat throughout the trials. Therefore, within mostly congruent lists it was expected that participants that were under stereotype threat would respond significantly faster to congruent stimuli than incongruent stimuli exhibiting a reactive control strategy. However, due to relying on a reactive control strategy, participants experiencing stereotype threat would also have significantly more intrusion errors.

Similar to Study 1, WMC was examined as a possible moderator of Stroop effects when participants were experiencing stereotype threat. It was predicted that women low in WMC should engage in more reactive control because they lack the attentional control necessary for proactive control. Proactive control is associated with better performance in mostly congruent lists than reactive control because of the necessity to maintain the goal internally to avoid habitual word responding. Since proactive control, if engaged, involves active goal maintenance over task trials that enhance early selection of task appropriate stimuli it tends to be engaged by those higher in WMC because it requires an abundance of attentional control. This top-down control mechanism decreases interference from irrelevant task dimensions which reduces the likelihood of participants learning an association between the word stimuli and correct response. In contrast reactive control, a bottom-up mechanism, increases the likelihood of learning associations between the irrelevant task dimensions and the correct responses because it

is a late correction mechanism. This late correction mechanism that only activates the goal after an error occurs increases learned responses. Thus, reactive control does not require an abundance of cognitive resources nor attentional control in order to operate, making it more likely to operate and be used by individuals that lack the cognitive resources to employ proactive control. Therefore participants with lower levels of WMC are more likely to use reactive control, which makes them more prone to effects of stimulus-response learning between the irrelevant stimuli and correct responses (Hutchison, 2011). Accordingly, it was predicted that individuals under stereotype threat with low WMC would display a more reactive control strategy marked by an increase in errors and a decrease in reaction time due to relying on the stimulus-response contingencies. In sum, it was expected that participants experiencing stereotype threat would use stimulus-response learning based on the irrelevant dimensions of the task stimuli that would increase error rate and decrease reaction time. It was expected that this tendency to utilize reactive control strategies would be more pronounced in individuals that were lower in WMC and who are already prone to ISPC and Stroop effects. Study 2 examined these predictions by determining overall Stroop effects in both reaction time and errors (difference score between mostly congruent and mostly incongruent lists) and also reaction times and errors to mostly congruent items, mostly incongruent items and ISPC effects (difference in Stroop effects between mostly congruent items and mostly incongruent items).

In short, the current studies were designed to use a validated cognitive task (the Stroop task) to investigate if individuals experience increases in response competition,

decreased goal maintenance, or both processes of control while under stereotype threat. In addition to investigating these two processes of control, Study 2 was designed to establish whether stereotype threat increases the likelihood of utilizing a reactive control strategy versus a proactive control strategy. While participants were under the impression that they were performing a stereotype relevant task, this project investigated if there are actual differences in these processes while the stereotype threat is operating (as opposed to after the stereotype task is complete as done in past research), as well as whether or not stereotype threat effects are dependent on trait levels of WMC. The findings from these two studies were expected to contribute to the theory of stereotype threat by determining what WM process is affected and moving past just stating that the negative performance effects are due to WM depletion. In addition, these findings hope to finally uncover just how trait levels of WMC contribute to negative performance effects. Understanding *how* working memory is affected by negative stereotypes, not just when and for who are affected, can allow educators to step in and design specific intervention programs to combat gender and racial disparities.

STUDY 1 METHOD

Participants

One hundred and eighty-seven male (21.2 years old, 88.5% Caucasian) participants from Montana State University participated for partial course credit towards an introductory psychology course requirement.

Procedure and Design

The study was a 2 (Stereotype Threat: Stereotype Threat, no Stereotype Threat) X 2 (List Congruency: Mostly Incongruent [MI], Mostly Congruent [MC]) X Continuous WMC mixed design. Prior to arrival, the research assistant prepared the computer to administer the tasks. Participants arrived at the lab individually and were informed that they were participating in a study that was being performed in conjunction with the Office of Admissions and that we were investigating a new portion of an entrance exam that was being developed for incoming freshmen. After reading the informed consent and agreeing to participate in the study, the participants were informed that they were going to be taking part in a series of tasks. The first task was described as a measure of personality traits, specifically “integrative orientation.” However the first task was actually the working memory capacity measure: Automated Operation Span (Ospan) (Unsworth, Heitz, Schrock, & Engle, 2005). After completing the measure of WMC, the participants were then informed that they would be completing second task described as a section of the exam that was being devised with the Office of Admissions where they

were then given the stereotype threat manipulation (see description below) or not (control).

The research assistant then readied the computer to perform the Stroop task. Once the computer was set up for the Stroop task the research assistant told the participant about the exam and read the instructions from the computer screen explaining to them that they must state the ink color not the word itself and to respond as quickly and accurately as possible. Participants were then given two examples, for which the experimenter read off the correct response, and the task began. After completing the Stroop task participants were then told they would complete a third and final task that would allow them to give their opinion of the exam. Participants then completed a survey that measured their domain identification with English and demographics. After the participants were finished they were debriefed, given a debriefing exit survey and thanked for their time.

Materials

All measures of WMC and the Stroop task were recorded using E-Prime Software (Shneider, Eschman, & Zuccolotto, 2002). The instructions for the WMC measure were displayed in black on a white background. Instructions for the Stroop task were displayed in white on a black background which remained black throughout the Stroop task trials of the colored words.

Working Memory Capacity Measure

Working Memory Capacity was a quasi-independent variable and was measured before the stereotype threat manipulation began. Unsworth et al., (2005) has demonstrated that this version of Ospan highly correlates with other measures of WMC and has good internal consistency ($\alpha = .78$) and test-retest reliability (.83). The Ospan task requires participants to view a series of 3-7 math equations, answer if the product of the equation is correct, given a letter, and later asked to retrieve the letter from memory when prompted (i.e., Is $2+4 \times 5 = 30$ (“NO”) H). After each response, they are presented with a letter for 800 ms to hold in memory for a later test. After 3 to 7 problems, participants are presented with a 3 x 4 matrix of letters and asked to click on the presented letters in the order in which they were shown. The Ospan score is the sum of all the letters from the set from which all letters were produced in the correct order with scores ranging from 0-75. In order to score on each 3-7 trials the participants had to correctly remember the series of letters that were given when asked to recall. Math error totals were also recorded in order to trim out participants that were not performing the math in order to remember the letters.

The Stroop Task

The Stroop was constructed using colored words (red, green, yellow and blue) and neutral words (bad, deep, poor, and legal) and the task required the participant to state the color of the word and not the word itself (i.e. the word “poor” in red ink with the correct response being “red”) (Hutchison, 2011; Kane & Engle, 2003). Participants were given two examples and 16 practice trials before the experimental manipulation portion of the

task began. The stimuli were presented and response times were recorded using E-Prime software with a 300 PST response box in a fixed random order. Target words were presented in 18-pt Courier New font at the center of the screen for 3,000 ms or until a response was given and separated by a 1,500 ms inter-trial interval. A research assistant recorded the correct (color) responses for each trial using an answer sheet to ensure the correct response coding was recorded. Participants responses were recorded as (1) correct response, (2 & 3) response error (blend (i.e., *re-green*) or incorrect color response), or (0) microphone error (microphone failed to pick up the vocal response).

Independent Variables

Stereotype Threat Manipulation

Stereotype threat was manipulated using subtle manipulation methods modeled after past research. A female experimenter administered the experiment to all male participants. This was done as a subtle manipulation as well as to control for differences that could result from having a male experimenter administer the test (i.e., stereotype lift from having a role model present). The stereotype threat condition was told that the exam was testing gender differences in verbal ability and performance in college students (Kiefer & Shih, 2006). Next participants were given a sheet for “data processing” that was titled the “Massachusetts Verbal Skills Battery” with the participant’s gender indicated and filled out prior to the participant entering the study (Siebt & Forster, 2004). While examining the “data processing” sheet the research assistant readied the computer to perform the Stroop task. Once the computer was set up for the Stroop task the research

assistant retrieved the data sheet and told the participant about the exam. The exam was described as measuring the “Verbal Skills of Men and Women” and titled the Verbal Ability Test (Siebt & Forster, 2004). Before the Stroop task began the participant was asked to indicate their gender by pressing “M” or “F,” which then advanced the screen to the Stroop task instructions (Siebt & Forster, 2004).

Control Condition

The control condition was told the study was testing processing ability and performance in college students (Kiefer & Shih, 2006). Additionally the sheet for “data processing” did not have gender listed and was titled the “Massachusetts Processing Skills Battery” (Siebt & Forster, 2004). The test was indicated as measuring “Processing Skills,” and titled Processing Ability Test (Siebt & Forster, 2004). In addition participants in the control condition did not have to indicate their gender before the Stroop task advanced to the instructions.

List Proportion Congruency Manipulation

List Congruency was manipulated through proportion congruency of the color words in the color naming Stroop task. Lists were made up of a total of 200 trials. Specifically, these trials consisted of 24 congruent trials, 24 incongruent trials, 32 neutral trials, and 120 fillers. Fillers were color words in either their congruent color in the mostly congruent lists or an incongruent color in the mostly incongruent lists (Hutchison, 2011). Mostly congruent lists then consisted of 144 congruent trials, 24 incongruent trials, and 32 neutral trials. The mostly incongruent lists consisted of 144 incongruent

trials, 24 congruent trials, and 32 neutral trials. All participants were given 16 random trials for practice before the experimental trials began.

Dependent Variables

Overall Stroop Effects

The primary dependent variables were overall Stroop effects in RT and errors. Stroop reaction time effects were determined by taking the average reaction time on incongruent trials per individual and subtracting the average reaction time of congruent trials (Incongruent-Congruent). Stroop error effects were computed by taking the average error score on the incongruent trials and subtracting the average congruent error score (Incongruent Error – Congruent Error).

Domain Identification

Domain identification was measured using the Domain Identification Measure (DIM) which measures to what extent an individual identifies with the domain under investigation (Smith & White, 2001). This variable was included as an exploratory variable that has been shown to be a boundary condition that predicts if and to what extent stereotype threat is experienced (Smith & White, 2002). This survey consists of 9 items measured on with a 5-point Likert scale 1 being “Strongly Disagree” and 5 being “Strongly Agree” for the first 4 items (i.e., “English is one of my best subjects”), 1 being “Not at all” and 5 being “Very much” for the following 4 items (i.e., How much do you enjoy English-related subjects”) and 1 being “Very poor” and 5 being “Excellent” for the last (i.e., Compared to other students, how good are you at verbal exams”).

STUDY 1 RESULTS

When analyzing the data, only correct responses were considered for the RT analyses. Fillers were not included in this analysis. Individual means and standard deviations were computed for all types of trials for each participant. Outliers were removed from the analysis using the Van Selst and Jolicoeur (1994) non-recursive outlier removal procedure which adjusts the criteria for outlier removal based on sample size. This procedure removed 2.5 % of correct RTs.

Ospan scores in the stereotype threat conditions were 39.6 ($SE = 2.5$) for those participants who received the mostly congruent list and 38.8 ($SE = 2.5$) for those who received the mostly incongruent list. Within the control groups Ospan scores were 38.1 ($SE = 2.1$) for those in the mostly congruent list condition and 40.9 ($SE = 2.7$) for participants in the mostly incongruent list condition. None of the groups significantly differed in Ospan scores (all t -values were <1).

Basic regression models were used to examine Stroop effects (i.e., incongruent – congruent conditions) in both errors and RTs. These were created in order to analyze the data with coding as follow: main effects of stereotype condition (coded as stereotype-threat = 1; no stereotype threat = -1), the main effect of list congruency (coded as mostly congruent = 1; mostly incongruent = -1), main effect of Ospan WMC (centered and measured continuously); the three two-way interaction terms and the one three-way interaction term. Participants with microphone errors greater than 25% (3 participants) and participants with incomplete data (2 participants) were excluded from analyses, resulting in 182 participants. RTs and error rates to neutral items did not vary as a

function of threat regardless of list (all t -values < 1.1) so neutral items are not discussed further. Domain Identification was included as a possible exploratory covariate. This was primarily due to its predictive value as a boundary condition of stereotype threat (Smith et al., 2007). However DIM entered into the regression model did not contribute to the overall model and thus was excluded from analysis.

The Distraction Hypothesis

Stroop effects in errors were regressed on the model. The overall model was significant $R^2 = .65$, $F(7, 174) = 18.41$, $p < .001$, revealing the typical effects of list congruency and WMC such that participants' Stroop effects were greater in the mostly congruent list than the mostly incongruent list ($\beta = -.54$, $t = 9.33$, $p < .01$), and participants that were lower in WMC had larger Stroop effects than participants that were higher in WMC ($\beta = -.23$, $t = -3.87$, $p < .01$). In addition, these two main effects were qualified by their two-way interaction ($\beta = -.22$, $t = -3.81$, $p < .01$), in which WMC differences were more pronounced in the mostly congruent list, replicating Hutchison (2011). There was also a main effect of stereotype threat ($\beta = .12$, $t = 2.11$, $p < .05$) that revealed that men in the stereotype threat condition had a larger Stroop effect than the control condition. This main effect of stereotype threat was qualified by an interaction with list condition ($\beta = .13$, $t = 2.29$, $p < .05$). Examining lists separately showed that stereotype threat increased Stroop errors among men in the mostly congruent list ($\beta = .24$, $t = 2.55$, $p < .05$), but not among men in the mostly incongruent list ($\beta = -.02$, $t = -0.17$, $p <$

.05). The regression also revealed a marginal 2-way interaction between threat condition and WMC ($\beta = -.11, t = -1.93, p = .054$).

Most importantly these main effects and two-way interactions were qualified by a significant 3-way interaction between our two independent variables (i.e., stereotype threat and list) and our quasi-independent variable (i.e., WMC) ($\beta = -.12, t = -1.99, p < .05$). Figure 1 depicts the predicted values for this 3-way interaction. In order to deconstruct this three-way interaction we first tested the simple stereotype threat x WMC interaction separately for each list. If significant, we then determined the effects of stereotype threat separately for those participants that were relatively high or low in WMC (using 1 *SD* above and below 39.50). For the mostly congruent list, the 2-way stereotype threat x WMC interaction was significant ($\beta = -.20, t = -2.14, p < .05$). Following Aiken and West (1991) guidelines for comparing simple slopes, a comparison of the simple slopes in the mostly congruent list at one *SD below* the mean WMC revealed a significant slope for the stereotype threat condition, ($\beta = -.30, t = -2.15, p < .05$) with Stroop effects increasing under stereotype threat. However, the simple slope analysis at one *SD above* the mean WMC did not reveal any significant differences ($p > .05$, see Figure 1). This suggests that within the mostly congruent list when participants would need to internally maintain the tasks goal to perform optimally, those participants that are lower in WMC made more Stroop errors if they were experiencing stereotype threat. The 2-way stereotype threat x WMC interaction was not significant however for those in the mostly incongruent list ($\beta = .01, t = 0.59, p > .05$) which suggests there was no effect of stereotype threat on Stroop performance for these participants (see Figure 1).

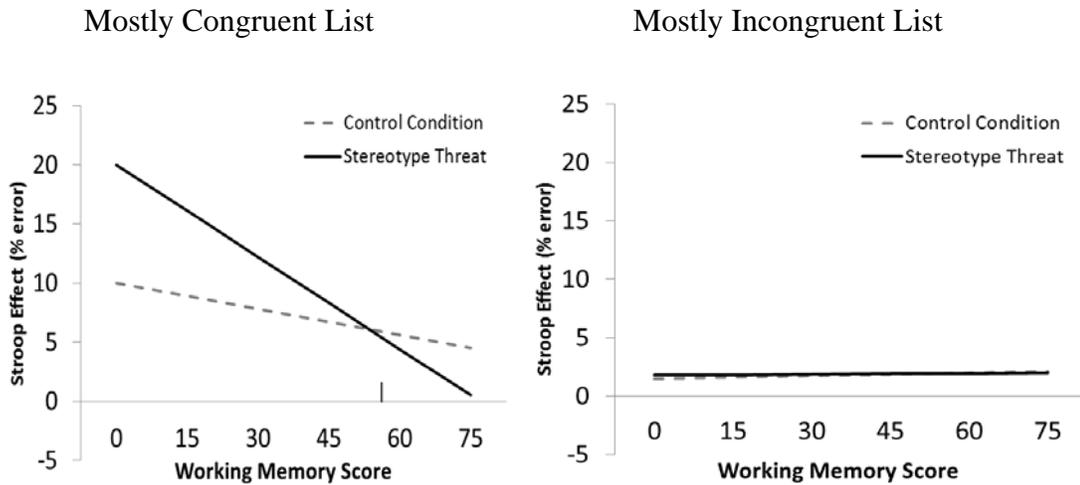


Figure 1. Predicted Values of Stroop Effects in Error Rates by Working Memory Capacity (WMC) for Stereotype Threat and Control participants within Mostly Congruent (left) or Mostly Incongruent (right) lists.

These results suggest that stereotype threat interferes with goal maintenance particularly among those lower WMC individuals, as was predicted by the distraction hypothesis, which is theorized to be due to a propensity towards increased distraction (McVay & Kane, 2009). Those individuals displaying a lower WMC showed increased Stroop errors when experiencing stereotype threat if they were performing the mostly congruent list. However, within the mostly incongruent list which provides external goal maintenance, there was no difference in Stroop performance in errors for those experiencing stereotype threat regardless of their trait levels of WMC.

The Response Competition Hypothesis

In order to test our response competition hypothesis, Stroop RT effects were regressed on the model, which was again significant $R^2 = .53$, $F(7, 174) = 28.95$, $p <$

.001. Unlike the distraction hypothesis regression, the only significant effects were that of the main effect of WMC ($\beta = -.11$, $t = -2.01$, $p < .05$) and a main effect of list ($\beta = .72$, $t = 13.949$, $p < .01$). Replicating Hutchison (2011), Stroop RT effects were larger for those lower in WMC and Stroop effects were larger in the mostly congruent list. The main effect of stereotype threat, the two-way interactions (all t -values < 1), and the three-way interaction did not approach statistical significance ($p = .29$).

STUDY 1 DISCUSSION

Study 1 tested three hypotheses. First, the distraction hypothesis suggested that stereotype threat would create a distraction (due to determining bias) that interferes with the ability to maintain the task goal. This should be especially true for those individuals that are lower in WMC. Distraction would be evidenced in an increase in intrusion errors on mostly congruent lists. In order to successfully perform on the mostly congruent list, a person must maintain focus in the face of distraction. If stereotype threat-related thoughts create distraction, individuals would have less attention to devote to the task, thus committing more errors. Second, the response competition hypothesis predicted that stereotype threat impairs participants' ability to resolve the conflict between two competing responses. The conflicting responses would be that of the correct response of the goal (i.e. name the ink color not the word) versus that of the habitual response to simply read the word. This would result in larger response latencies to the incongruent trials and presumably occurs only if the participant is maintaining the task goal. Finally, if the dual processes hypothesis were to be confirmed we would have expected to see that stereotype threat would impair both goal maintenance and response competition. Specifically, if both processes were negatively affected by stereotype threat those participants that were lower in WMC would show increased intrusion errors on mostly congruent lists and longer response latencies on mostly incongruent lists.

The main findings from Study 1 provided support for the distraction hypothesis. As predicted, participants under stereotype threat experienced a significant increase in intrusion errors on the mostly congruent list, a task that requires internal goal

maintenance for successful performance. In addition, only the participants that were lower in WMC showed this effect. These results suggest that stereotype threat interferes with internal goal maintenance by causing distraction. In contrast, if the nature of the task (mostly incongruent list) promoted external maintenance of the goal participants experiencing stereotype threat were able to perform similarly to those in the control group that were lower in WMC. Study 1 supports the hypothesis that stereotype threat impairs performance by disrupting an individual's ability to actively maintain the goal of the task at hand. Thus, stereotype threat appears to lead to a loss of task appropriate goals resulting in the intrusion of incorrect responses. This is especially true for those individuals who have lower WMC. However, individuals higher in WMC appear to be unaffected and are able to retain the ability to maintain the task goal even when faced with distraction.

The mean error rates from Study 1 revealed that the overall regression model was significant. Specifically, the main effect of list revealed that participants experienced significant increases in intrusion error rates in the mostly congruent list as compared to those in the mostly incongruent list. The main effect of WMC replicated past research on the Stroop effect, whereby participants' lower in WMC experienced greater errors under mostly congruent conditions (Kane & Engle, 2003; Hutchison, 2011). The main effect of stereotype threat showed that participants in the stereotype threat condition produced significantly more intrusion errors than those in the control condition. The significant two-way interaction between list and WMC showed that participants that were lower in WMC and in the mostly congruent list produced significantly more errors than

participants that were in the mostly incongruent list and higher in WMC. The significant two-way interaction between stereotype threat and list revealed that those participants in stereotype threat condition experienced a significant increase in intrusion errors compared to those in the control condition. This was not the case for the mostly incongruent list however. The two-way interaction between stereotype threat and WMC, although only marginally significant was qualified by the significant three-way interaction between stereotype threat, list, and WMC. This interaction showed that those participants in the stereotype threat condition that were lower in WMC produced significantly more intrusion errors on the mostly congruent lists. These results lend support to the belief that situational cues related to stereotype threat impair a person's ability to suppress stereotype-related thoughts. These thoughts create such distraction that the goal of the task at hand cannot be maintained, thus preventing individuals experiencing stereotype threat from performing successfully. This was especially true for those possessing a limited WMC, lending further support to the distraction hypothesis.

The inability of individuals experiencing stereotype threat and lower in WMC to maintain the task goal sheds insight onto what cognitive mechanism is actually affected while stereotype threat is operating and how working memory is affected. This provides support for the *working memory depletion account* suggesting that the stereotype threat disrupts working memory by negatively impacting a person's ability to maintain task goals in the face of distraction. The findings also suggest that distraction can occur for other populations, rather than just those that are stigmatized. Indeed, the results suggest for those populations that already are prone to distraction or disruptions in attention (i.e.

individuals with ADHD) and at times when cognitive resources are not at optimal levels (i.e. stress, variations in circadian rhythm), individuals may be more likely to neglect the goal of the task. However, the results do not allow inferences as to what type of cognitive control strategy operates while under stereotype threat that determines task performance. Specifically, because of how the lists were constructed to manipulate list-wide proportion congruency ISPC was not controlled. Manipulations of list-wide proportion congruency allow for investigations of what type of top-down control process is working (i.e. suppression of word reading pathway and active goal maintenance) at the general task level, but does not allow for an investigation of what type of automatic bottom-up process is working on a trial-to-trial basis throughout the task (i.e. suppression of word reading at a specific item level or reactivation of the goal after conflict occurs).

Since ISPC was not controlled, we cannot determine if the results of Study 1 where due to a general task demand unit activating a relevant or irrelevant response pathway through top-down control processes. Thus, it is possible that the results may have been due to ISPC effects in disguise, whereby automatic processes trigger word suppression at the item level or participants' learned responses to repeated pairings of specific colors and words. By controlling for ISPC, strategy use can be investigated to determine if stereotype threat disrupts top-down controlled processes, such as early-selection strategies that work in service of the goal by suppressing irrelevant task dimensions. Alternatively, controlling for ISPC can also help to establish if stereotype threat promotes the use of automatic processes that operate by activating the task goal once conflict is detected, due to relying on irrelevant task dimensions and learned

responses. Stated another way, controlling for ISPC, we can investigate if while experiencing stereotype threat participants are trying to perform well but the threat of being judged based on the stereotype disrupts their controlled processes which increases the reliance on automatic processes.

By knowing if stereotype threat interferes with controlled processes or promotes the use of automatic processes, we can determine if stereotype threat affects working memory by creating distraction that reduces individuals' ability to actively maintain task goals or if they just rely on learned automatic processes while a stereotype threat is operating. In addition, by investigating strategy use it can be determined if individuals experiencing stereotype threat adopt a strategy that only operates at a general task level or if they adopt strategies that allow for successful performance on a trial-to-trial basis due to fluctuations in task demands.

Study 2 was designed to investigate whether individuals under stereotype threat utilize a *proactive control strategy* (active goal maintenance that is connected to individuals higher in WMC and successful task performance) or a *reactive control strategy* (reactivation of the goal after committing an error that is associated with lower performance and individuals with lower WMC) by manipulating ISPC.

Item-Specific Proportion Congruency

It is necessary to distinguish between ISPC effects and list-wide effects in order to better understand reactive and proactive control strategies. It is generally assumed that interference on conflict tasks arise through conflicting responses that are generated by

irrelevant and relevant response pathways. For instance, in Cohen, Dunbar, and McClelland's (1990), parallel distributed process model, greater conflict results when the irrelevant response pathway dominates the task. For example, the relevant pathway in the Stroop task would be the pathway that elicits the correct response that is consistent with the task goal of stating the ink color whereas the irrelevant response pathway would be the one that elicits the incorrect word reading response. A top-down task demand unit is argued to increase activation of information within a task-appropriate pathway and suppress information from a task-inappropriate pathway. Thus, within mostly congruent lists, in which larger proportion congruency effects occur, participants more heavily rely on the irrelevant pathway since it would better predict a correct response. Therefore, as the proportion of incongruent trials increase, interference from word reading decreases because repeated exposure to these high conflict trials triggers the need for more top-down control, providing continual maintenance of the task goal. However, if the task is made up of mostly congruent trials and fewer instances of conflict are likely to be encountered less control is needed. This places greater demand on WM to maintain the task goal so interference from word reading does not occur, which would result in increased intrusion errors on rare incongruent trials.

The top-down task demand unit explains list-wide proportion congruency effects in that a global task strategy can be adopted in order to perform the task successfully by biasing responses toward the appropriate stimulus dimension. For example, within a mostly incongruent list a task demand unit would be boosted that activates the color naming pathway while suppressing the inappropriate word-reading pathway. Attentional

control, would then be vital for maintaining and utilizing task goals when the proportion congruency of the task is manipulated, to help account for performance effects due to trial-by-trial fluctuations in control. This need for attentional control, especially in mostly congruent lists, has been evidenced in individual differences in WMC on Stroop performance across list-wide proportion congruency manipulations (Kane & Engle, 2003). When the proportion of congruent trials increases external reminders of the task goal decrease which results in increased errors when an incongruent trial is encountered. In these cases, the task demand unit begins to activate the irrelevant task pathway and the task goal is neglected. In order to explain list-wide differences Botvinick, Braver, Barch, Carter, and Cohen (2001) later proposed the addition of a conflict monitoring unit which could better explain differences in response competition depending on the degree of conflict within congruency tasks as well as trial-by-trial fluctuations in control (e.g. Gratton, Coles, & Donchin, 1992). This unit could therefore signal when more control was needed (i.e. incongruent trial in a mostly congruent list). Modification of these models was better able to explain the list-wide proportion congruency effects that were found by Kane and Engle (2003). Across five experiments, both list-wide and group-wide effects were found that demonstrated Stroop effects in interference errors and facilitated reaction time were primarily exhibited by those participants in a mostly congruent list whom had low WMC. However, there was little to no individual differences in performance in mostly incongruent lists where the goal is externally maintained via frequent exposure to incongruent stimuli. This was believed to be due to the increased conflict that participants experienced within a mostly incongruent list providing external

maintenance of the task goal. Within the mostly congruent list however, infrequent exposures to incongruent stimuli produced a greater demand on WMC because participants would have to actively maintain the task goal, thus participants that were lower in WMC were much more prone to interference errors and facilitated RT. This provides further evidence that increasing list-wide proportion congruency of congruent items reflects failures in top-down active control of the task goal.

In an intriguing study Jacoby, Lindsay, and Hessels, (2003) challenged the notion that list-wide congruency effects were due to top-down control processes and proposed instead that these effects may be due to ISPC effects which is confounded within a mostly congruent list. Prior manipulations of list-wide congruency confounded ISPC because mostly congruent or mostly incongruent lists are themselves made up of mostly congruent or mostly incongruent items. ISPC effects manifest in interference and facilitation at an item level, resulting in smaller interference and facilitation effects on incongruent items than congruent items. However, list-wide proportion congruency effects are produced at a word-reading level not color naming, whereby a task-demand unit can only account for a reduction in the influence of word-reading across the whole task. Thus, Jacoby et al. (2003), constructed lists in such a way that manipulated specific items which produced a 50% list-wide proportion congruency. The results from Jacoby et al., (2003) demonstrated increased Stroop effects on mostly congruent items but decreased Stroop effects on mostly incongruent items within the same 50% congruent list. Thus, it was suggested that these ISPC effects could not be due to a central task demand unit because the subject cannot predict proportion congruency of the upcoming

trial in an overall list-wide proportion congruency of 50% (Jacoby et al., 2003). Jacoby et al., (2003) therefore suggested list-wide proportion congruency effects in Stroop were due to automatic bottom-up processes occurring at the item level rather, than a general task level of top-down control of word reading. Specifically, Jacoby et al., (2003) proposed two automatic bottom-up control processes. The first was an item specific control process, which due to early processing of individual words, an inhibitory process is triggered that stops full-word reading and blocks access of any word-reading processes to the response system (Jacoby et al., 2003). The second automatic process Jacoby et al., (2003) proposed was an associative learning process, whereby pairings between color words and responses produced learned responses. These learned S-R contingencies are acquired rapidly and independently of the color in which words are presented (Jacoby et al., 2003).

To account for ISPC effects, later manipulations of Stroop began to construct lists in such a way that distinguishes between list-wide and ISPC effects, as well as to investigate automatic control of the influence of word reading (Bugg, Jacoby, & Toth, 2008; Hutchison, 2011). In order to account for these ISPC effects models began to include an item-specific control mechanism. This control mechanism was believed to exert influence in such a way that mostly incongruent words in the Stroop task automatically trigger a top-down suppression of word reading allowing for better performance. However, suppression of word-reading leads to increased incidental learning of S-R contingencies after conflict has occurred.

In addition, research has provided support for the alternative associative learning account by which participants' make use of the tasks stimulus-response contingencies (Hutchison, 2011). Within this account participants are believed to learn to give a specific color response to specific irrelevant words. This S-R contingency learning is exhibited in a decreased RT across the trials to those items that have high contingency responses (i.e. the word red is usually shown in blue). By orthogonally manipulating list-wide proportion congruency, ISPC, and stimulus contingencies Hutchison (2011) provided evidence that participants in fact rely on S-R contingencies to predict correct responses to specific items. He had proposed that if item-specific congruency effects were due to item-specific control then Stroop effects would be reduced for words that are usually shown in incongruent colors, regardless of whether the word contained a high contingency or low contingency response. In contrast, if Stroop effects were due to S-R learning, Stroop effects would be reduced to a greater extent for those trials involving the high contingency color. Indeed, Hutchison (2011) found that, within mostly congruent lists composed of high contingency items Stroop effects increased on mostly congruent items, but decreased in mostly incongruent lists on mostly incongruent items. In addition, this experiment found that these effects were more pronounced for those low in WMC.

When manipulating only list-wide congruency, the critical items are all either mostly congruent or mostly incongruent, thus if there are 4 critical words in a mostly congruent list all randomly presented words would be mostly presented in their respective congruent color, thus item-specific proportion congruency would be confounded. This list-wide proportion congruency allows for an investigation of the use of top-down

processes in determining performance. Specifically, it determines how well participants maintain the task goal and overcome response competition in regards to the task demand unit if ISPC is not confounded within the list. In contrast, manipulating ISPC by controlling which items are mostly congruent or incongruent allows for a distinction between the use of both a top-down control mechanism as well as a bottom-up context specific mechanism, most specifically proactive and reactive control.

Proactive and Reactive Control Strategies

Proactive control strategies (early selection) require an abundance of mental and physical resources, are resistant to stimulus response contingencies (i.e., a conditioned response to a stimulus), and tend to be used most often by individuals that are high in WMC (Braver et al., 2007). In addition, because they are metabolically taxing and indicative of prolonged goal maintenance, they are most often employed when the goal is externally maintained. Proactive control is believed to be a top-down control mechanism primarily used by participants that are higher in WMC (Braver et al, 2007; De Pisapia & Braver, 2006). It is distinguished by constant activation of the lateral prefrontal cortex, whereby active goal maintenance is used in order to prepare the participant for early selection of task-appropriate stimuli. This early selection decreases interference from irrelevant stimuli, thus reducing the learning of S-R contingencies along the irrelevant dimension pathway. However, proactive control is very metabolically taxing, requiring higher levels of WMC in order for it to be used over longer periods of time.

In contrast, reactive control (late correction) tends to be utilized by individuals who are lower in WMC and not actively maintaining the goal. Instead, individuals wait for conflict to arise (which is detected by the anterior cingulate cortex, ACC), which then transiently activates the lateral prefrontal cortex (PFC) to reactivate the goal. For example, when reading the word red in blue ink the ACC detects the conflict and activates the goal in the PFC (De Pisapia & Braver, 2006). Reactive control, is considered to be a bottom-up control mechanism whereby the goal of the task is only retrieved when conflict arises. Use of reactive control makes individuals more susceptible to learning S-R contingencies in the Stroop task because they fail to suppress word-reading. In addition, reactive control strategies are primarily used by participants that are low in WMC (Braver et al., 2007; De Pisapia & Braver, 2006; Hutchison, 2011). When using reactive control, rather than actively maintaining the goal, if an error occurs on the current trial, this conflict triggers the need for more control by activating the lateral PFC. In tasks in which there is a greater number of non-conflicting trials, less top-down control is needed, resulting in word reading that causes greater interference on the rare conflicting trials. Thus, reactive control is considered to be a late-selection mechanism used after word-reading has been suppressed and conflict has already occurred, increasing the likelihood of learning S-R contingencies.

Combining the work of Kane and Engle (2003) and Braver et al., (2007), would suggest that, regardless of both list-wide and ISPC manipulations, those participants that are higher in WMC will rely on the use of the early selection mechanism of proactive control. Individuals higher in WMC employ the use of proactive control regardless of the

task or context demands, and are less sensitive to list-wide proportion congruency effects as well as ISPC effects (Hutchison, 2011). Therefore, Study 1 results suggested that regardless of stereotype threat, when participants were higher in WMC, they utilized a proactive control strategy that allowed for top-down control of responses by actively maintaining the task goal in the face of distraction and ignoring the irrelevant stimuli dimensions. In contrast, if participants are lower in WMC, they are more reliant on the external task goal reminder provided by frequent incongruent trials (list-wide proportion congruency effects), suggesting the use of a late-correction reactive control process. In addition, those participants that are lower in WMC are also more sensitive to S-R contingencies whereby, unlike those higher in WMC, they are unable to suppress the irrelevant task dimensions and rely on transient activation of the goal after conflict occurs.

Research conducted by Hutchison (2011), which orthogonally manipulated list-wide proportion congruency (as in Study 1), ISPC and stimulus contingencies within mostly congruent or mostly incongruent lists, found evidence for reactive control strategy as well as evidence that list-wide proportion congruency effects are not just artifacts of ISPC. In addition, Hutchison (2011) found evidence suggesting that a participant's ability to use a reactive and/or proactive control strategy depends on trait levels of WMC. Specifically, individuals low in WMC use reactive control strategies to a greater extent than people that are high in WMC (Hutchison, 2011). A reactive control strategy is more sensitive to response contingencies (S-R contingencies) and is associated with lowered performance on infrequent incongruent trials. People using a reactive strategy are argued

to learn these stimulus-response contingencies between the irrelevant words and the ink colors due to frequent pairings. This is believed to be due to their inability to disregard the irrelevant dimensions of the task and thus demonstrate larger ISPC effects. Hutchison (2011) demonstrated that those low in WMC showed larger ISPC effects (indicative of reactive control), especially when the overall list was mostly congruent, which also encourages reactive strategies.

Study 2 Project Overview

Study 1 results suggest that participants experiencing stereotype threat and that are lower in WMC become so distracted that they cannot maintain the task goal, resulting in increased interference. This would suggest a reliance on reactive control mechanisms. However, because ISPC was controlled for in Study 1 the results cannot provide any conclusive results as to whether or not stereotype threat increases the use of a bottom-up control strategy whereby participants would rely on transient activation of the goal and be sensitive to S-R contingencies. Learned associations that result from reactive control use, are demonstrated by individuals lower in WMC. These learned associations also result in an increase in Stroop effects. Due to a greater reliance on the learned responses between the irrelevant task dimensions and the stimuli, individuals that have adopted a reactive control strategy, are more likely to neglect the task goal, and rely on conflict between to accessible responses to reactivate the goal. By distinguishing between list-wide and ISPC effects we can investigate whether or not participants experiencing stereotype threat rely on reactive control that is marked by an increase in associative learning to form a

response on tasks that require active goal maintenance. Also, by manipulating the ISPC one can investigate the type of control strategy individuals utilize when under stereotype threat in order to complete the task and if individual differences in WMC predict which strategy is used.

Study 2 was designed to replicate Study 1's findings as well as to examine what performance control strategies are engaged when individuals are under stereotype threat. Proactive and reactive control strategies relate to stereotype threat in many ways, for example the STEP model postulates that individuals adopt specific performance goals that predict how they approach the situation as well as the outcome (Smith, 2004). It is also known that stereotype threat interferes with women's problem-solving strategies on both word problem tests and math tasks when under stereotype threat (Quinn & Spencer, 2001). However, there have been no studies that use ISPC and the Stroop task to investigate the control strategies engaged while under stereotype threat. As such, a second goal of the current project was to see if individuals under stereotype threat wait for conflict to arise using reactive control strategies and if this is dependent on WMC, as Hutchison (2011) found.

In addition to examining which control strategy operates while individuals are experiencing stereotype threat and replicating Study 1, the current study aimed to examine participants' subjective experience while also using a different population (women) and a different stereotype (quantitative abilities). Thus, the current study focused on women's subjective experience of the stereotype that "women are bad at math" (Smith & White, 2002; Spencer, et al., 1999) by including additional exploratory

measures of task experience, achievement goal adoption and performance concerns that were given after the Stroop task was explained as a task of quantitative abilities.

It was predicted that women experiencing stereotype threat, would not only show evidence of distraction, but that reactive control strategies would be the cognitive process operating while performing the task. It was also expected that women who were in the stereotype threat condition would report less positive task experiences, would report greater performance-avoidance goal adoption, and increased performance concerns compare to the control condition. In order to introduce the new manipulations successfully, the WMC measure changed to the automated version of the Reading Span (Unsworth et al., 2005) so as not to interfere with the math stereotype, a new version of the Stroop was implemented that used numbers and quantities (Pavese & Ultima, 1998) rather than classic color naming Stroop. Also in order to investigate ISPC, the counting Stroop was modeled after past research involving mostly congruent and mostly incongruent items embedded within mostly congruent and mostly incongruent lists (Hutchison, 2011).

Unlike list-wide proportion congruency, which distinguishes between the dual processes of control, ISPC is composed of specific critical items being either mostly congruent or mostly incongruent and can determine the use of proactive or reactive control strategies. For example, by manipulating (ISPC), it can be investigated if participants use strategies such as actively maintaining the goal without external assistance (*proactive control strategy hypothesis*) or if they just wait for conflict to arise (*reactive control strategy hypothesis*). ISPC is manipulated such that, within lists, critical

items are either congruent (i.e., “4444” where quantity and item match) 85% of the time or incongruent (i.e., “6666” where quantity and item do not match) 85% of the time. List-wide proportion congruency is then controlled by inserting 100% congruent or incongruent fillers to boost or lower the overall congruency of the list. This results in a list-wide proportion congruency of 67% vs. 33% and an ISPC of 85% vs. 15%. Average Stroop effects on errors and RTs for the critical mostly congruent and mostly incongruent items that are manipulated within both lists are then assessed. This allows for an investigation of proportion congruency effects within each list, between items within a list, as well as any congruency effects across conditions.

Study 2 specifically aimed to replicate the results obtained from Study 1. It was predicted that list-wide proportion congruency Stroop effects in errors would emerge that replicated Study 1’s distraction hypothesis, such that while experiencing stereotype threat, women would produce significantly more intrusion errors on incongruent trials within the mostly congruent list, and that these effects would be more pronounced for those lower in WMC. In addition, it was predicted that Stroop effects in RTs would also replicate both Study 1 results as well as past Stroop research that manipulated both ISPC and list-wide proportion congruency (Hutchison, 2011). In particular, it was expected that the results would produce the typical interaction between list and WMC, whereby those that are lower in WMC have significantly slower RTs in the mostly incongruent list condition. In addition to replicating Study 1’s results, Study 2 aimed to test specific predictions in regards to the new ISPC manipulation and addition of subjective measures.

A preliminary analysis of list, item type, and congruency condition manipulations on Stroop effects predicted the typical main effect of congruency condition was also expected to reach significance, such that there would be significantly more errors committed within the incongruent congruency condition than the mostly congruent congruency condition. In addition, I predicted a list by congruency interaction would emerge resulting in significantly Stroop effects in the mostly congruent list versus the mostly incongruent list. A significant item type by congruency interaction was also predicted to produce the typical ISPC effect, such that significantly larger Stroop effects would be found for mostly congruent items versus mostly incongruent items. Preliminary analysis of Stroop effects RTs on the critical items were predicted to replicate past research (Hutchison, 2011; Jacoby et al., 2003) by producing the typical main effect of congruency condition, such that RTs would be significantly slower within the incongruent condition than the congruent condition. In addition, the typical two-way interaction between item type and congruency condition was predicted to emerge demonstrating an increase in reaction time on mostly congruent items within the incongruent condition relative to the three other groups. In sum, the preliminary analysis of critical items were predicted to produce the typical ISPC effects that past research have produced when ISPC and list-wide proportion congruency is manipulated (Hutchison, 2011; Jacoby et al., 2003).

The regression analysis on Stroop effects in errors and RTs with WMC included were expected to replicate Study 1 results as outlined above. Specifically, Stroop effects in errors were expected to produce the same results that support the distraction

hypothesis, such that women that were lower in WMC and experiencing stereotype threat would produce more errors in the mostly congruent list condition. This was predicted to be evidenced in a main effect of list, stereotype threat and WMC, which would be qualified by the significant two-way interaction between list and WMC, the significant two-way interaction between list and stereotype threat, and the significant three-way interaction between list, stereotype threat, and WMC. The results of the regression analysis on overall Stroop effects in RT were also expected to replicate Study 1's results, such that the typical main effects of list and WMC as well as their respective two-way interaction would replicate past Stroop studies that have manipulated list-wide proportion congruency (Hutchison, 2011; Kane & Engle, 2003). The regression analysis on mostly congruent and mostly incongruent Stroop effects in errors were predicted to replicate the typical main effects of list and WMC as well as their respective interaction, such that error rate would be significantly larger on mostly congruent items in mostly congruent lists and that this would be especially true for those lower in WMC. It was also predicted that there would be a main effect of stereotype threat such that individuals experiencing stereotype threat would have significantly more errors on mostly congruent items than mostly incongruent items. A predicted interaction between list and stereotype threat was also predicted such that those women that were experiencing stereotype threat would produce significantly more errors in the mostly congruent list condition on mostly congruent items. These results were also predicted to be more pronounced for those lower in WMC (i.e. significant three-way interaction between list, stereotype threat, and WMC).

It was predicted that the regression analysis on overall ISPC effects would replicate typical ISPC effects found in past research (Hutchison, 2011). Specifically, it was predicted that ISPC effects would be larger for both error rate and RTs in the mostly congruent list (i.e. significant main effect of list) and that WMC would moderate these effects such that those participants that were lower in WMC would have significantly larger ISPC effects in errors and RTs (i.e. main effect of WMC and significant two-way interaction). It was also predicted that a main effect of stereotype threat would emerge, such that those participants experiencing stereotype threat produced significantly larger ISPC effects. In addition, it was expected that stereotype threat would interact with list such that those experiencing stereotype threat would produce significantly larger ISPC effects in the mostly congruent list relative to the other three conditions. Lastly, it was also expected that there would be a significant three-way interaction between list, stereotype threat, and WMC, such that the women that were lower in WMC and experiencing stereotype threat would produce significantly larger ISPC effects in mostly congruent lists relative to those in the other three conditions.

The final regression analysis on the secondary dependent variables was predicted to replicate past stereotype threat research (Brodish & Devine, 2009; Marx & Goff, 2005; Smith & Johnson, 2006; Smith et al., 2007; Smith & White, 2002). Specifically, it was predicted that only a significant main effect of stereotype threat would emerge. Thus, it was predicted that women experiencing stereotype threat would report lower levels of perceived competence, less interest in the task and performing the task in the future, would not perceive the task to be valuable, would express less interest in pursuing Math

domains, would adopt a performance-avoidance achievement goal, and would report higher levels of concern that their performance would be judged based on the stereotype.

STUDY 2 METHOD

Participants

One hundred and fifty-four female ($M = 20.1$ years old, 93.5% Caucasian) Montana State University Introductory Psychology Students participated for partial course credit.

Procedure and Design

Participants were randomly assigned to one of four conditions: 2 (Stereotype threat: stereotype threat vs. no threat) X 2 (List: mostly congruent list vs. mostly incongruent list) X 2 (Congruency: congruent vs. incongruent) X continuous (working memory score) mixed design with stereotype threat and list being manipulated between-subjects and congruency being manipulated within-subjects. Eighty-four participants in the mostly incongruent condition ($n = 35$ control; $n=49$ stereotype threat) and 70 participants in the mostly congruent condition ($n=34$ control; $n= 36$ stereotype threat).

Prior to arrival, the research assistant prepared the computer to administer the tasks. The gender of the research assistant was again controlled such that a female research assistant ran all the female participants. Upon entering the lab, participants were informed that the study was devised in conjunction with the Office of Admissions. Participants were told that we were investigating a new portion of an entrance exam that was being developed for incoming freshmen. After reading the informed consent and agreeing to participate in the study, the participants were informed that they were going

to be taking part in a series of three tasks. Participants were then randomly assigned to one of 4 conditions. All participants completed the first task, which was the automated version of Rspan used to measure WMC (Unsworth et al., 2005) and was described to measure personality traits, specifically “integrative orientation.” The automated Rspan task required the participant to read a sentence, determine if the sentence makes sense in both syntax and grammar, and then receive a letter to remember until prompted to retrieve at later recall from working memory. After completing the Rspan, participants were then told they would be completing a second task described as a section of the exam that was being devised with the Office of Admissions. At this point, participants received the stereotype threat manipulation (described below). The research assistant then readied the counting Stroop task on computer and told the participant about the exam. Once the counting Stroop task was ready, the research assistant then read the instructions from the computer screen explaining to them that they must state the quantity of the item not the item itself and to respond as quickly and accurately as possible. After the counting Stroop, task all participants were then told they would be completing the third task that allowed them to give their opinion of exam which contained our secondary dependent variables. After the participants were finished they were debriefed, given a debriefing exit survey and thanked for their time.

Materials

All measures of WMC and the Stroop task were recorded using E-Prime Software (Schneider, Eschman, & Zuccolotto, 2002). The instructions for the WMC measure were

displayed in black on a white background. Instructions for the counting Stroop task were displayed in white on a black background which remained black throughout the task.

Working Memory Capacity Measure

Working memory capacity was measured with Automated Reading Span (Rspan) (Unsworth et al., 2005) and was presented in the same fashion as the Ospan from Study 1. The task consisted of 75 trials with sentence sets of 3-7. Participants had to determine if the sentence made sense (You must look down to see the sky. = False) and then received a letter that they had to retain in memory for later recall. Similar to Ospan, Reading span scores are determined by summing all the letters from each set that were recalled in the correct order.

The Counting Stroop Task

The counting Stroop task required the participant to state the quantity of the item instead of the item itself (i.e., 3333 = “4”) (Pavese & Umiltà, 1998). Participants were given 2 examples and 16 practice trials before the experimental manipulation portion of the task began. Response time was again recorded with E-Prime software with a 300 PST serial response box with responses recorded by the research assistant sitting next to the participant just as was done in Study 1. Answers were recorded by marking the number that the participant stated (i.e., “2”=2) and microphone errors were recorded by typing “0”.

Independent Variables

Stereotype Threat Manipulation

In order to trigger stereotype threat, a subtle manipulation was used that included telling the participants in the stereotype threat condition that, “Our goal today is to test gender differences in quantitative ability and performance in college students,” modeled after Spencer et al., (1999). Next, participants were given a sheet for “data processing” that was titled the “Massachusetts Quantitative Skills Battery” with the participant’s gender indicated and filled out prior to the participant entering the study (Spencer et al., 1999). The exam was described as measuring the “Quantitative Skills of Men and Women” and was titled the Quantitative Ability Test (Spencer et al., 1999). Before the counting Stroop task began, the participant was asked to indicate their gender by pressing “M” or “F,” which then advanced the screen to the counting Stroop task instructions (Danaher & Crandall, 2008).

Control Condition

Participants assigned to the control condition were informed that the second task was investigating processing ability and performance in college students. They were then given sheet for “data processing” as well (but without gender listed) that was titled the “Massachusetts Processing Skills Battery” (Siebt & Forster, 2004). In addition, participants in the control condition did not have to indicate their gender before the Stroop task advanced to the instructions.

Item Specific Proportion Congruency Manipulation

ISPC was manipulated within the counting Stroop task and modeled after Hutchison (2011), but composed of numbers (i.e., 1, 2, 3, 4, 5 and 6) and symbols (i.e., #) in order to manipulate ISPC so that reactive and proactive control strategies could be investigated. The participants first were given 16 random practice trials and then engaged in a total of 200 experimental trials. The numbers 1, 3, 4, and 6 were the critical items that were presented as either mostly congruent (i.e. “333”) 85% of the time or mostly incongruent (i.e. “3”) 85% of the time depending on list. The 200 trials were composed of 32 congruent trials, 32 incongruent trials, 24 neutral trials, and 112 filler trials. Fillers were either 100% congruent (i.e., 22 or 55555) or 100% incongruent (i.e., 22222 or 55). Participants were randomly assigned to the mostly congruent lists or mostly incongruent lists. Mostly congruent lists were composed of 144 congruent trials (i.e., 32 congruent and 112 congruent fillers), 32 incongruent trials, and 24 neutral trials. Mostly incongruent lists were composed of 144 incongruent trials (i.e., 32 incongruent trials and 112 incongruent fillers), 32 congruent trials, and 24 neutral trials. This created a list-wide proportion congruency of 72% vs. 18% and an ISPC of 85% vs. 15%.

Primary Dependent Variables

Overall Stroop Effects

As in Study 1, the primary dependent measures were overall Stroop effects in errors and RT (Hutchison, 2011). These effects are computed by taking the average RT on incongruent trials per individual and subtracting the average RT of congruent trials

(Incongruent-Congruent). Stroop error effects were as difference scores which were computed by taking the average error score on the incongruent trials and subtracting the average congruent error score (Incongruent Error – Congruent Error).

Mostly Congruent and Mostly Incongruent Item Stroop Effects

In order to investigate the control strategy hypotheses Stroop effects in RT and errors on mostly congruent and incongruent critical stimuli within the lists were also investigated (Hutchison, 2011). Responses to filler trials were ignored.

Item Specific Proportion Congruency Stroop Effects

In order to investigate the overall congruency manipulation, ISPC effects within both lists were computed as difference scores for overall Stroop effects in RTs and errors between mostly congruent items and mostly incongruent items. The overall ISPC effect was computed for both RT and errors by subtracting the mostly incongruent item effects from the mostly congruent item effects that were manipulated within both lists (Hutchison, 2011). Thus, mostly incongruent items effects within both lists were subtracted from mostly congruent item effects within both lists, to obtain an overall effect of the ISPC manipulation.

Secondary Dependent Variables

Secondary dependent measures were included for exploratory analyses to examine participants' domain identification, perceived competence, current task interest,

future task interest, task value, task involvement, future interest in pursuing Math, motivation/goal adoption, performance concerns in regards to stereotype threat, individual impression, group impression. Specifically variables assessed were as follows:

Domain Identification

Domain identification was included as secondary variable (Smith & White, 2001). This variable has been shown to be related to stereotype threat effects and predicts if, and to what, extent stereotype threat is experienced (Smith, 2006; Smith & Johnson, 2006). This survey consists of 9 items measured on a 5-point Likert scale with 1 being “Strongly Disagree” and 5 being “Strongly Agree” for the first 4 items (i.e., “Math is one of my best subjects”), 1 being “Not at all” and 5 being “Very much” for the following 4 items (i.e., “How much do you enjoy Math-related subjects”) and 1 being “Very poor” and 5 being “Excellent” for the last (i.e., “Compared to other students, how good are you at math exams”).

Task Rating Survey

The Task Rating Survey was adapted from Smith et al., (2007). This survey consisted of six subscales with a total of 28 items measured on a 7-point Likert Scale (1=Strongly Disagree, 7=Strongly Agree) that included 7 subscales. Four of the items were filler questions and three reversed scored items (1 current interest, 1 value, and 1 perceived competence).

Perceived Competence. The first subscale consisted of 4 items and measured how competent in the task participants perceived themselves (i.e., “I feel that I did this exam well.”).

Current Task Interest. The second subscale consisted of 5 items and measured how interesting the participants found the current task (i.e., “I would describe this exam as very interesting.”).

Future Task Interest. The next subscale consisted of 3 items and measured how interested the participants would be in completing the same task in the future (i.e., “I would like to do this type of exam in the future.”).

Value. The fourth subscale consisted of 7 items and measured how valuable the participants thought the task (i.e., “I think this is a valuable exam.”).

Task Involvement. The fifth subscale contained 3 items to measure how involved participants were in the task (i.e., “I became very absorbed in with this exam while I was doing it.”).

Future Interest in Pursuing Math. The last and final subscale included 2 items that measured participants’ future interest in pursuing Math (i.e., “How interested are you in learning more about being a Math Major.”).

Achievement Goal Questionnaire Revised

Achievement goals were measured using the Achievement Goal Questionnaire Revised (AGQ-R) (Elliot & Murayama, 2008). This survey consisted of 12 items measured on a 5-point Likert scale (1 being “Strongly Disagree” and 5 being “Strongly Agree”). Each of the four subscales contained 3 items.

Master-Approach Goal (MAV). This goal is defined as a desire to be successful and/or understand or master a skill. It is an individual’s motivation to improve performance and develop knowledge just for the sake of mastering the topic (i.e., “My aim was to completely master the material presented in this exam”).

Mastery-Avoidance Goal (MAP). This goal is the considered to be a person’s desire to avoid misunderstanding, failing to learn a skill, making errors, and making any mistakes. It is motivated by the desire to avoid not mastering a skill (i.e., “My aim was to avoid learning less than I possibly could”).

Performance-Approach Goal (PAP). PAP is operationalized as an individual’s desire to outperform in comparison to others. It is the motivation to improve performance and increase their knowledge compared to the performance of others (i.e., “I was striving to do well compared to other students”).

Performance-Avoidance Goals (PAV). PAV is an individual’s want to avoid performing poorly compared to others. It is the motivation to avoid appearing as though they did not improve their performance or increase their knowledge compared to the

performance of others (i.e., “My goal was to avoid performing poorly compared to others”).

Exam Concerns Survey

Performance and impression concerns were measured using the “Exam Concerns Survey” which consisted of three subscales adapted from Marx and Goff (2005). The three subscales included 3 items each measured on a 7-point Likert scale (1=Strongly Disagree-7=Strongly Agree). These three subscales measured participants’ performance concerns as being judged in regards:

Stereotype Concerns. This scale measured participants concern that their performance would be judged based on the stereotype that women are bad at math (i.e., “I worry that, because I know the negative stereotype about women and math ability, my anxiety about confirming this stereotype will negatively influence how I perform.”).

Individual Impression Concerns. This subscale measured to what extent participants were concerned about the impression they were portraying as an individual (i.e., “I am concerned about the impression I am making.”).

Group Impression Concerns. The last subscale measured the concern participants felt that their performance would reflect poorly on the image of their group (i.e., “I am concerned about people’s impression of my group.”).

STUDY 2 RESULTS

Only correct responses were considered for the RT analyses. Fillers were not included in this analysis. Individual means and standard deviations were computed for all types of trials for each participant. Outliers were removed from the analysis using the Van Selst and Jolicoeur (1994) non-recursive outlier removal procedure which adjusts the criteria for outlier removal based on sample size. This procedure removed 2.5 % of correct RTs.

Preliminary Analysis of Stroop Effects

Mostly congruent item and mostly incongruent item Stroop effects in errors and RT were submitted to a 2(List: mostly congruent vs. mostly incongruent) X 2(Item Type: congruent item vs. incongruent item) X 2 (Congruency: congruent condition vs. incongruent condition) mixed ANOVA. Errors were assessed first to establish item specific effects within congruency conditions between lists. The ANOVA revealed the typical significant main effect of list, $F(1, 142) = 4.53, p = .035$, whereby significantly more intrusion errors were committed in the MC list ($M = .032$) than the MI list ($M = .020$). The analysis also revealed a significant main effect of item type, $F(1, 142) = 12.93, p < .001$ such that significantly more errors were committed on mostly congruent items ($M = .034$) than mostly incongruent items ($M = .018$). The main effect of congruency condition also reached significance, $F(1, 142) = 56.61, p < .001$, revealing an overall Stroop effect, with more errors in the incongruent condition ($M = .045$) than the congruent condition ($M = .007$). In addition these main effects of list and condition

were qualified by a significant two-way interaction between the list and condition, $F(1,142) = 6.21, p = .014$. The significant two-way interaction between list and congruency condition revealed that Stroop effects were much larger in the mostly congruent list ($M = .009$) than the mostly incongruent list ($M = .04$). In addition the ANOVA revealed a significant two-way interaction between item type and condition, $F(1, 142) = 12.01, p < .001$. The significant two-way interaction between item type and congruency revealed the typical ISPC effect, such that Stroop effects were larger for mostly congruent items ($M = .06$) than mostly incongruent items ($M = .02$). The two-way interaction between list and item type and the three-way interaction between list, item type, and congruency condition did not approach significance (p values $> .10$).

Next RTs on specific items within congruency conditions between lists were assessed. The ANOVA revealed the typical significant main effect of congruency condition, $F(1, 142) = 185.03, p < .001$, such that RTs were significantly slower in the incongruent condition ($M = 745.76$) relative to the congruent condition ($M = 677.62$). In addition the analysis also revealed the typical two-way interaction between item type and congruency condition $F(1,142) = 86.52, p < .001$, whereby Stroop effects in RTs were significantly slower on mostly congruent items ($M = 100.20$) than mostly incongruent items ($M = 28.40$). The main effects of list and item type did not approach significance (p values $> .10$). In addition the two-way interactions between item type and list, congruency condition and list, nor the three-way interaction between list, item type and congruency condition did not approach significance (p values $> .10$). In sum, these results revealed the typical list-wide proportion congruency effects when ISPC is controlled as

well as typical ISPC effects, which replicates past research and lends support that list-wide proportion effects are not ISPC effects in disguise (Hutchison, 2011).

After these preliminary analyses regression models including WMC were assessed on all primary and secondary dependent variables. Automated Rspan scores in the stereotype threat conditions were 33.2 ($SE = 2.3$) for those participants who received the mostly congruent list and 27.6 ($SE = 2.3$) for those who received the mostly incongruent list. Within the control groups Rspan scores were 32.4 ($SE = 2.5$) for those in the mostly congruent list condition and 31.7 ($SE = 2.9$) for participants in the mostly incongruent list condition (see Table 4, pg. 75). Groups did not significantly differ from each other in WMC (t values $> .50$ and p values $> .50$). Correlations between all primary and secondary variables are provided in Table 1 (pgs. 73). In addition, Tables 2-4 display all means, standard deviations, and respective significance tests for list condition (Table 2, pg. 74), stereotype threat condition (Table 3, pg. 74) and the four experimental groups (Table 4, pg. 75).

Basic regression models were used to examine Stroop effects (i.e., incongruent – congruent conditions) in both errors and RTs. These were created in order to analyze the data as follows: main effects of stereotype condition (coded as stereotype threat = 1; no stereotype threat = -1), the main effect of list congruency (coded as mostly congruent = 1; mostly incongruent = -1), main effect of Rspan WMC (centered and measured continuously); the three two-way interaction terms and the one three-way interaction term. No participants committed over 25% microphone errors resulting in no exclusion of participants. Participants with a sentence accuracy under 85 % (3 participants),

participants with incomplete data (1 participants), participants that expressed suspicion (2 participants), participants with responses coded incorrectly (2 participants), and participants that expressed extreme anxiety before the study began (1 participants) were excluded from analyses, resulting in 144 participants. RTs and error rates to neutral items did not vary as a function of stereotype threat regardless of list (all t -values < 1.1) so neutral items are not discussed further.

Domain Identification was included in the survey packet as a possible exploratory variable. This was primarily because past research has shown that individuals highly identified with the domain are more susceptible to experiencing stereotype threat and its correlation with the AGQ-R motivation measure and the “Task Rating” survey (Smith & White, 2001) (see Table 1, pg.73). However DIM entered into the regression model did not contribute to the overall model nor change any of the significant, marginally significant, or non-significant results and thus was excluded from analysis.

Table 1. Correlations between WMC, primary dependent variables and secondary dependent variables.

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | |
|-----------------------------|-----|-----|------|-------|-------|-------|-------|--------|--------|-------|------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|-------|-------|------|
| 1. WMC | --- | .01 | -.03 | .01 | -.02 | .00 | -.04 | .01 | .01 | .22** | -.01 | -.03 | .03 | .02 | .19* | -.04 | -.01 | .14 | .00 | -.10 | .10 | .00 | .18* | |
| 2. Stroop Errors | --- | --- | .05 | .97** | 0 | .36** | .13 | .76** | -.09 | .13 | -.02 | -.09 | .03 | .07 | -.02 | .02 | .10 | -.06 | -.02 | .02 | -.00 | .04 | -.02 | |
| 3. Stroop RT | --- | --- | --- | -.00 | .93** | .21* | .52** | -.10 | .40** | -.05 | .09 | .20* | .13 | .03 | .04 | .01 | .13 | -.05 | .03 | -.06 | .03 | .08 | -.02 | |
| 4. MC Item Errors | --- | --- | --- | --- | -.05 | .14 | .12 | .89** | -.13 | .14 | -.02 | -.11 | .02 | .05 | -.05 | .00 | .09 | -.05 | -.01 | .01 | -.03 | .02 | -.05 | |
| 5. MC Item RT | --- | --- | --- | --- | --- | .22** | .17* | -.15 | .71** | -.07 | .12 | .19* | .14 | .03 | .03 | .01 | .16 | -.09 | .03 | -.08 | .06 | .08 | -.03 | |
| 6. MI Item Errors | --- | --- | --- | --- | --- | --- | .07 | -.32** | .13 | .00 | .01 | .03 | .07 | .11 | .09 | .08 | .09 | -.03 | -.05 | .03 | .10 | .09 | .10 | |
| 7. MI Item RT | --- | --- | --- | --- | --- | --- | --- | .08 | -.58** | .02 | -.03 | .08 | .02 | .01 | .04 | -.00 | -.02 | .07 | .02 | .03 | -.07 | .02 | .03 | |
| 8. ISPC Errors | --- | --- | --- | --- | --- | --- | --- | --- | -.18* | .13 | -.02 | -.12 | -.01 | -.00 | -.09 | -.04 | .04 | -.04 | .01 | -.00 | -.08 | -.02 | -.09 | |
| 9. ISPC RT | --- | --- | --- | --- | --- | --- | --- | --- | --- | -.07 | .12 | .10 | .10 | .02 | -.01 | .01 | .15 | -.12 | .01 | -.09 | .09 | .05 | -.05 | |
| 10. Perceived Competence | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .17* | .13 | .20* | .08 | .08 | .19* | .19* | .21* | .07 | -.03 | -.06 | -.07 | .17* | |
| 11. Current Task Interest | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .74** | .78** | .59** | .17* | .40** | .19* | .15 | .08 | .04 | .05 | .11 | .18* | |
| 12. Future Task Interest | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .76** | .50** | .27** | .49** | .22** | .26** | .21* | -.01 | .06 | .11 | .27** | |
| 13. Value | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .66** | .19* | .51** | .27** | .30** | .18* | .06 | .10 | .14 | .19* | |
| 14. Task Involvement | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .05 | .43** | .21* | .27** | .22** | .10 | .08 | .07 | -.06 | |
| 15. Future Interest in Math | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .10 | .01 | .18* | .06 | -.07 | .06 | .05 | .68** | |
| 16. MAP | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .36** | .48** | .25** | .10 | .14 | .14 | .09 | |
| 17. MAV | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .25** | .44** | .05 | .16 | .13 | .13 | |
| 18. PAP | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .58** | .10 | .17* | .17* | .23** | |
| 19. PAV | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .13 | .17* | .22** | .09 | |
| 20. Stereotype Concerns | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .08 | .21* | .01 |
| 21. Individual Concerns | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | .68** | -.00 |
| 22. Group Concerns | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | -.00 |
| 23. DIM | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

Note: * Indicates the correlation as significant at 0.05 level (2-tailed). ** Indicates the correlation as significant at 0.01 level (2-tailed).

Table 2. Mean, standard deviation, *t*- values, and significance levels for WMC and Stroop Effects within the Mostly Congruent List Condition and Mostly Incongruent List Condition.

| Variable | List Condition | | | | <i>t</i> (142) | <i>p</i> |
|-----------------------|-------------------------|-----------|-------------------------|-----------|----------------|----------|
| | MC List (<i>n</i> =77) | | MI List (<i>n</i> =67) | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | |
| WMC | 32.83 | 14.72 | 29.63 | 15.24 | 1.28 | 0.20 |
| Stroop Errors | 0.09 [□] | 0.12 | 0.04 | 0.10 | 5.94 | 0.02 |
| Stroop RT | 131.34 | 83.98 | 112.30 | 109.55 | 1.11 | 0.27 |
| MC Item Stroop Errors | 0.07 [□] | 0.11 | 0.04 | 0.10 | 2.19 | 0.03 |
| MC Item Stroop RT | 113.23 | 73.66 | 101.32 | 94.56 | 0.85 | 0.40 |
| MI Item Stroop Errors | 0.03 | 0.06 | 0.02 | 0.05 | 1.48 | 0.14 |
| MI Item Stroop RT | 34.44 | 64.06 | 22.35 | 80.67 | 1.00 | 0.32 |
| ISPC Errors | 0.05 | 0.12 | 0.02 | 0.11 | 1.39 | 0.17 |
| ISPC RT | 80.01 | 91.24 | 78.98 | 112.19 | -0.01 | 0.99 |

Note: Stroop Errors and MC Item Errors row and column not sharing a subscript differ at $p < .05$ level as determined by *t*-test.

Table 3. Mean, standard deviation, *t*- values, and significance levels for WMC and Stroop Effects within the Stereotype Threat Condition and Control Condition.

| Variable | Stereotype Threat Condition | | | | <i>t</i> (142) | <i>p</i> |
|-----------------------|-----------------------------|-----------|-------------------------|-----------|----------------|----------|
| | Stereotype (<i>n</i> =77) | | Control (<i>n</i> =68) | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | |
| WMC | 30.71 | 14.49 | 32.04 | 15.63 | -0.53 | 0.61 |
| Stroop Errors | 0.07 | 0.11 | 0.06 | 0.12 | 0.75 | 0.45 |
| Stroop RT | 123.16 | 93.63 | 120.92 | 100.89 | 0.14 | 0.89 |
| MC Item Stroop Errors | 0.06 | 0.11 | 0.05 | 0.11 | 0.83 | 0.41 |
| MC Item Stroop RT | 105.97 | 83.76 | 109.61 | 84.73 | -0.26 | 0.81 |
| MI Item Stroop Errors | 0.02 | 0.05 | 0.02 | 0.05 | -0.14 | 0.89 |
| MI Item Stroop RT | 34.37 | 77.01 | 87.01 | 66.58 | 0.98 | 0.33 |
| ISPC Errors | 0.04 | 0.11 | 0.02 | 0.11 | 0.86 | 0.39 |
| ISPC RT | 71.6 | 111.38 | 87.01 | 88.47 | -0.91 | 0.36 |

Table 4. Mean, standard deviation, F values, and significance levels for WMC and Stroop Effects within the four experimental groups.

| Variable | Stereotype Threat & List Conditions | | | | | | | | F | p |
|-----------------------|-------------------------------------|-------|----------------|--------|----------------|-------|----------------|--------|------|------|
| | Stereotype | | | | Control | | | | | |
| | MC List (n=42) | | MI List (n=34) | | MC List (n=35) | | MI List (n=32) | | | |
| M | SD | M | SD | M | SD | M | SD | | | |
| WMC | 33.21 | 14.89 | 27.62 | 13.55 | 32.37 | 14.72 | 31.70 | 16.77 | 1.11 | 0.29 |
| Stroop Errors | 0.08 | 0.10 | 0.06 | 0.12 | 0.09 | 0.15 | 0.02 | 0.07 | 1.42 | 0.24 |
| Stroop RT | 121.58 | 93.66 | 125.10 | 94.96 | 141.11 | 70.52 | 99.50 | 122.92 | 2.85 | 0.09 |
| MC Item Stroop Errors | 0.07 | 0.09 | 0.06 | 0.12 | 0.08 | 0.13 | 0.02 | 0.07 | 1.54 | 0.22 |
| MC Item Stroop RT | 103.03 | 80.72 | 109.61 | 88.45 | 125.48 | 63.15 | 92.78 | 101.13 | 2.67 | 0.10 |
| MI Item Stroop Errors | 0.03 | 0.06 | 0.01 | 0.04 | 0.03 | 0.05 | 0.02 | 0.06 | 0.01 | 0.91 |
| MI Item Stroop RT | 37.11 | 65.30 | 30.99 | 90.31 | 31.25 | 63.35 | 13.44 | 69.63 | 0.45 | 0.50 |
| ISPC Errors | 0.04 | 0.11 | 0.04 | 0.12 | 0.05 | 0.13 | -0.01 | 0.09 | 1.32 | 0.25 |
| ISPC RT | 65.92 | 91.63 | 78.62 | 132.93 | 94.23 | 89.64 | 79.34 | 87.94 | 0.73 | 0.40 |

Primary Dependent Variables

Overall Stroop Effects

Stroop effects in errors were first regressed on the model, as in Study 1. The overall model did not reach significance $R^2 = .24$, $F(7,136) = 1.23$, $p = .29$. The exploratory analysis did reveal the typical main effect of list whereby participants' Stroop effects in errors were larger in the mostly congruent list than in the mostly incongruent list ($\beta = .20$, $t = 2.41$, $p < .05$) (see Figure 2, pg. 76 and Table 2, pg. 74). This replicates Hutchison's (2011) finding of significant list differences even when ISPC is controlled within the list. The main effects of WMC and stereotype threat did not reach significance. In addition the three two-way interactions and the three-way interaction also did not reach significance.

Overall Stroop Error Rates

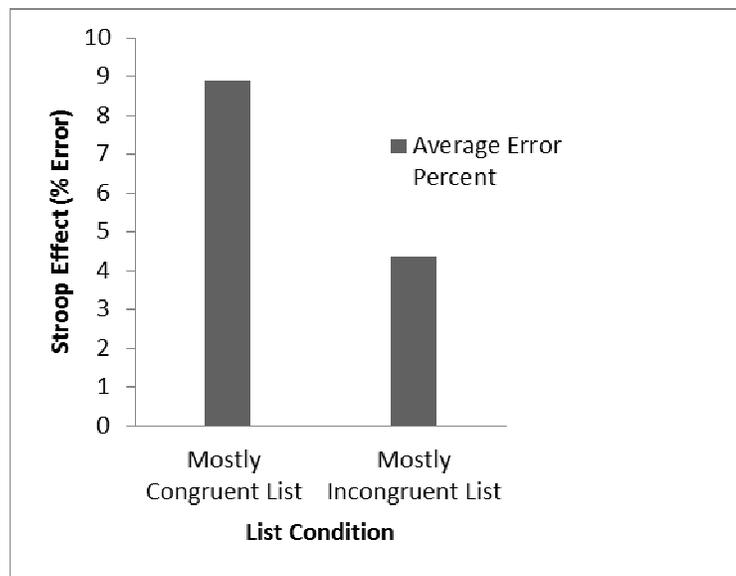


Figure 2. Average Stroop Effect Error Rates for the Mostly Congruent list and Mostly Incongruent list.

Stroop effects in RT were regressed on the model. Like errors, the overall model was not significant $R^2 = .22$, $F(7,136) = .94$, $p = .48$. The model also did not reveal any significant effects of list, threat, or WMC. None of the two-way interactions or the three way interaction reached significance.

Mostly Congruent and Mostly Incongruent Item Stroop Effects

Although the overall Stroop effects in errors and RT did not reach significance, in order to test the proactive and reactive strategy hypotheses we next conducted analyses of Stroop effects in errors and RT on mostly congruent items, mostly incongruent items, and the overall ISPC Stroop effects in errors and RT.

Stroop effects in errors on mostly congruent items were regressed on the model. Again the overall model did not reach significance, $R^2 = .22$, $F(7,136) = 1.11$, $p = .36$, however the exploratory analysis did reveal a significant main effect of list ($\beta = .21$, $t = 2.18$, $p < .05$) (refer to Figure 3, pg. 78 and see Table 2, pg. 74). Stroop effects in errors on mostly congruent items were greater in the mostly congruent list than the mostly incongruent list. Again, no other main effects or interactions reached significance.

Error Rate on Mostly Congruent Items

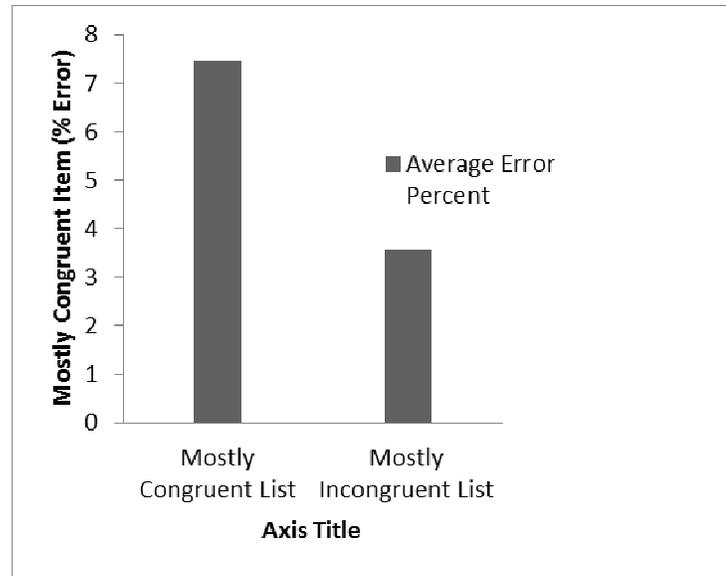


Figure 3. Average Error Rates on Mostly Congruent Items for the Mostly Congruent list and Mostly Incongruent list.

Stroop effects in RT on mostly congruent items were regressed on the model. The overall model was again not significant, $R^2 = .19$, $F(7,136) = .72$, $p = .66$. No other main effects or interactions reached significance.

Stroop effects in errors on mostly incongruent items were regressed on the model. The overall model did not reach significance, $R^2 = .15$, $F(7,136) = .47$, $p = .87$. Again no main effects or interactions reached significance.

Stroop effects in RT on mostly incongruent items were regressed on the model. The overall model did not reach significance, $R^2 = .21$, $F(7,136) = .91$, $p = .50$. The main effects of list, stereotype threat, or WMC did not reach significance, nor did any of the interactions between these variables.

Item Specific Proportion Congruency Stroop Effects

Although the preliminary analysis and the regression model on mostly congruent item and mostly incongruent item Stroop effects was significant we next assessed overall IPSC effects in error rate and RTs between lists.

ISPC error effects were regressed on the model. Again the overall model did not reach significance, $R^2 = .18$, $F(7,136) = .68$, $p = .69$. The model also did not reveal any significant effects of list, stereotype threat, or WMC with no significant interactions between these variables.

ISPC RT effects were regressed on the model. Again the overall model did not reach significance, $R^2 = .15$, $F(7,136) = .47$, $p = .87$, nor did the main effects of list, stereotype threat, or WMC. The model also revealed no significant interactions between these variables.

Secondary Dependent Variables

For the secondary survey measures that were included for exploratory purposes we first examined the seven Task Rating Survey subscales that measured participants' perceived competence, current task interest, future task interest, value, task involvement, and future interest in pursuing Math (Smith et al, 2007). We then examined participants' achievement goal adoption using the four subscales of the AGQ-R (Elliot & Murayama, 2008). Next we then analyzed our three subscales that measured participants' performance concerns (Marx & Goff, 2005). Tables 5-7 display all means, standard deviations, and respective significance tests for list conditions (Table 5, pg. 80),

stereotype threat conditions (Table 6, pg. 81) and the four experimental groups (Table 7, pg. 82).

Table 5. Mean, standard deviation, *t*- values, and significance levels for all Secondary Dependent Variables separated into their subscales and DIM within the Mostly Congruent List Condition and Mostly Incongruent List Condition.

| Variable | List Condition | | | | <i>t</i> (142) | <i>p</i> |
|-------------------------|-------------------------|-----------|-------------------------|-----------|----------------|----------|
| | MC List (<i>n</i> =77) | | MI List (<i>n</i> =67) | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | |
| Perceived Competence | 4.21 | 0.77 | 4.10 | 0.79 | 0.84 | 0.40 |
| Current Task Interest | 3.66 | 1.38 | 3.84 | 1.18 | -0.83 | 0.41 |
| Future Task Interest | 4.03 | 1.39 | 4.10 | 1.22 | -0.21 | 0.83 |
| Value | 4.09 | 1.19 | 4.34 | 0.89 | -1.40 | 0.17 |
| Task Involvement | 4.79 | 1.23 | 5.07 | 1.02 | -1.46 | 0.15 |
| Future Interest in Math | 2.32 | 1.40 | 2.53 | 1.70 | -0.79 | 0.43 |
| MAP | 3.61 | 0.86 | 3.66 | 0.86 | -0.29 | 0.77 |
| MAV | 3.01 | 1.04 | 3.05 | 0.92 | -0.25 | 0.80 |
| PAP | 3.91 | 0.92 | 3.92 | 0.92 | -0.07 | 0.94 |
| PAV | 3.81 | 0.99 | 3.86 | 1.11 | -0.28 | 0.78 |
| Stereotype Concerns | 1.97 | 1.26 | 1.99 | 1.10 | -0.08 | 0.94 |
| Individual Concerns | 4.17 | 1.61 | 4.74 [Ⓜ] | 1.44 | -2.22 | 0.03 |
| Group Concerns | 3.28 | 1.63 | 3.82 [Ⓜ] | 1.35 | -2.14 | 0.03 |
| DIM | 3.11 | 0.80 | 3.12 | 0.93 | -0.09 | 0.93 |

Note: Individual Concerns and Group Concerns row and column not sharing a subscript differ at $p < .05$ level as determined by a *t*-test.

Table 6. Mean, standard deviation, *t*- values, and significance levels for all Secondary Dependent Variables separated into their subscales and DIM within the Stereotype Threat Condition and Control Condition.

| Variable | Stereotype Threat Condition | | | | <i>t</i> (142) | <i>p</i> |
|-------------------------|-----------------------------|-----------|-------------------------|-----------|----------------|----------|
| | Stereotype (<i>n</i> =77) | | Control (<i>n</i> =68) | | | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | | |
| Perceived Competence | 4.19 | 0.74 | 4.12 | 0.82 | 0.53 | 0.59 |
| Current Task Interest | 3.81 | 1.28 | 3.67 | 1.30 | 0.66 | 0.51 |
| Future Task Interest | 3.91 | 1.35 | 4.22 | 1.25 | -1.40 | 0.16 |
| Value | 4.18 | 1.12 | 4.24 | 1.00 | -0.35 | 0.73 |
| Task Involvement | 4.94 | 1.14 | 4.89 | 1.15 | 0.29 | 0.77 |
| Future Interest in Math | 2.29 | 1.39 | 2.57 | 1.70 | -1.07 | 0.29 |
| MAP | 3.60 | 0.88 | 3.67 | 0.84 | -0.49 | 0.62 |
| MAV | 3.08 | 0.98 | 2.97 | 0.99 | 0.66 | 0.51 |
| PAP | 4.01 | 0.84 | 3.82 | 1.00 | 1.26 | 0.21 |
| PAV | 3.83 | 0.97 | 3.84 | 1.12 | -0.04 | 0.97 |
| Stereotype Concerns | 2.39 [Ⓜ] | 1.26 | 1.53 | 0.89 | 4.65 | 0.00 |
| Individual Concerns | 4.33 | 1.61 | 4.55 | 1.49 | -0.85 | 3.96 |
| Group Concerns | 3.47 | 1.49 | 3.59 | 1.58 | -0.45 | 0.66 |
| DIM | 3.08 | 0.82 | 3.15 | 0.91 | -0.48 | 0.63 |

Note: Stereotype Concerns row and column not sharing a subscript differ at $p < .001$ level as determined by a *t*-test.

Table 7. Mean, standard deviation, F- values, and significance levels for all Secondary Dependent Variables separated into their subscales and DIM within all four experimental groups.

| Variable | Stereotype Threat & List Conditons | | | | | | | | F | p |
|-------------------------|------------------------------------|------|----------------|------|----------------|------|----------------|------|------|------|
| | Stereotype | | | | Control | | | | | |
| | MC List (n=42) | | MI List (n=34) | | MC List (n=35) | | MI List (n=32) | | | |
| | M | SD | M | SD | M | SD | M | SD | | |
| Perceived Competence | 4.35 \square | 0.83 | 4.00 | 0.80 | 4.05 | 0.87 | 4.20 \square | 0.77 | 3.97 | 0.05 |
| Current Task Interest | 3.58 | 1.25 | 4.11 | 1.11 | 3.77 | 1.40 | 3.57 | 1.21 | 3.57 | 0.11 |
| Future Task Interest | 3.86 | 1.37 | 3.98 | 1.38 | 4.25 | 1.44 | 4.19 | 1.03 | 0.17 | 0.68 |
| Value | 4.02 | 1.33 | 4.40 | 1.01 | 4.18 | 1.20 | 4.31 | 0.76 | 0.43 | 0.51 |
| Task Involvement | 4.77 | 1.20 | 5.17 | 1.01 | 4.82 | 1.26 | 4.97 | 1.04 | 0.41 | 0.52 |
| Future Interest in Math | 2.27 | 1.22 | 2.31 | 1.50 | 2.39 | 1.51 | 2.77 | 1.89 | 0.44 | 0.51 |
| MAP | 3.56 | 0.81 | 3.65 | 0.96 | 3.68 | 0.92 | 3.67 | 0.75 | 0.10 | 0.75 |
| MAV | 3.02 | 0.90 | 3.15 | 1.08 | 2.99 | 1.19 | 2.95 | 0.72 | 0.25 | 0.62 |
| PAP | 4.14 | 0.71 | 3.84 | 0.95 | 3.64 | 1.07 | 4.01 | 0.89 | 4.86 | 0.03 |
| PAV | 3.83 | 0.89 | 3.83 | 1.08 | 3.79 | 1.11 | 3.89 | 1.15 | 0.06 | 0.81 |
| Stereotype Concerns | 2.31 | 1.31 | 2.48 | 1.22 | 1.57 | 1.10 | 1.48 | 0.62 | 0.46 | 0.50 |
| Individual Concerns | 4.10 | 1.60 | 4.68 | 1.58 | 4.31 | 1.63 | 4.81 | 1.30 | 0.07 | 0.80 |
| Group Concerns | 3.10 | 1.53 | 3.93 | 1.33 | 3.49 | 1.75 | 3.70 | 1.39 | 1.67 | 0.20 |
| DIM | 3.08 | 0.83 | 3.08 | 0.82 | 3.14 | 0.77 | 3.16 | 1.05 | 0.00 | 0.99 |

82

Note. Perceived Competence row and column not sharing a subscript differ at $p < .05$ level as determined by Simple effects contrast.

Task Rating Survey

Perceived Competence. First the measure of participants reported perceived competence on the task was regressed on the model. The overall model was marginally significant, $R^2 = .31$, $F(7,136) = 1.99$, $p = .06$. The exploratory analysis revealed a significant main effect of WMC, ($\beta = .19$, $t = 2.28$, $p < .05$), such that as WMC increased so did participants' reported perceived competence. In addition, the model revealed a marginal 2-way interaction between list and threat ($\beta = .16$, $t = 1.92$, $p = .057$) (see Table 7, pg. 82). Examining list threat separately revealed that perceived competence increased for those in the stereotype threat mostly congruent list conditions ($M=4.53$) relative to those in the stereotype threat mostly incongruent list conditions ($M=4.00$) ($\beta = .17$, $t = 2.07$, $p < .05$). This was not the case for those in the control conditions ($\beta = -.08$, $t = -.78$, $p = .47$).

Current Task Interest. Third, the subscale measuring participants' current interest in the task was regressed on the model. The overall model did not reach significance, $R^2 = .21$, $F(7,136) = .85$, $p = .55$. The main effects of stereotype threat, list, and WMC, as well as the interactions between these variables, did not approach significance.

Future Task Interest. The fourth subscale measuring participants' reported future interest in the task was regressed on the model. The overall model did not approach significance, $R^2 = .20$, $F(7,135) = .79$, $p = .60$, as did none of the main effects of the independent variables or the interactions between them.

Value. Participants perceptions of the value of the task was then regressed on the model, which again did not approach significance, $R^2 = .16$, $F(7,136) = .49$, $p = .84$. In addition the model did not reveal any significant main effects of stereotype threat, list, or WMC nor did any of the respective interactions between these variables.

Task Involvement. The sixth subscale that measured participants' task involvement was regressed on the model. Like task value, future interest, and current interest, the overall model did not approach significance, $R^2 = .25$, $F(7,136) = 1.30$, $p = .26$. However the exploratory analysis did produce a significant interaction between list and WMC ($\beta = -.17$, $t = -2.21$, $p < .05$), such that, as WMC increased participants in the mostly congruent list condition reported lower levels of task involvement. However, as WMC increased, participants in the mostly incongruent list condition reported higher levels task involvement.

Future Interest in Math. Finally, the seventh subscale measuring participants' future interest in pursuing Math was regressed on the model. The overall model was not significant, $R^2 = .26$, $F(7,135) = 1.34$, $p = .24$. However exploratory analysis did reveal a significant main of WMC ($\beta = .19$, $t = 2.27$, $p < .05$) such that, as WMC increased, so too did participants' reported future interest in pursuing Math (see Table 1, pg. 73).

Achievement Goal Questionnaire Revised

Next we analyzed the four subscales of the AGQ-R to examine the manipulations effects on achievement goal adoption.

Mastery-Approach Goals. First, MAP goals were regressed on the model. The overall model did not approach significance $R^2 = .14$, $F(7,135) = .39$, $p = .91$, as did none of the main effects of stereotype threat, list, WMC or their respective interactions.

Mastery-Avoidance Goals. Second MAV goals were regressed on the model. The overall model again did not reach statistical significance $R^2 = .14$, $F(7,135) = .36$, $p = .93$ nor did any of the main effects or interactions.

Performance-Approach Goals. Next we regressed PAP goals on the model, which the overall model also did not reach significance $R^2 = .27$, $F(7,135) = 1.49$, $p = .18$, as did none of the main effects. However exploratory analysis did reveal a marginally significant interaction between list and stereotype threat, ($\beta = .16$, $t = 1.89$, $p < .061$). Examining list separately between the stereotype threat condition showed participants' reported PAP goals significantly differed from each other in the mostly congruent list condition ($\beta = .25$, $t = 2.47$, $p < .05$) but not in the mostly incongruent list condition ($\beta = .19$, $t = -1.54$, $p = .13$). Participants within the stereotype threat mostly congruent list condition were more likely to report possessing a PAP goal than those in the other three conditions (see Table 7, pg. 82). No other main effects or their corresponding interactions between the independent variables reached significance.

Performance-Avoidance Goals. Finally, PAV goals were regressed on the model. The overall model did not reach significance, $R^2 = .24$, $F(7,135) = 1.17$, $p = .33$, but exploratory analysis did reveal a significant interaction between stereotype threat and WMC, ($\beta = .21$, $t = 2.4$, $p < .05$), such that, as WMC increased, participants that were in

the stereotype threat condition were more likely to report possessing a PAV goal. This was not so for participants in the control condition who were less likely to report possessing a PAV goal as WMC increased. No other main effects or the respective interactions approached statistical significance.

Exam Concerns Survey

In order to establish any kind of performance concerns were being experienced, we examined our three subscales of the impression concerns survey that measured participants stereotype threat, individual, and group performance concerns (Marx & Goff, 2005).

Stereotype Concerns. First, our measure of participants' concern about their performance being judged based on the stereotype that women do poorly on tasks of quantitative ability was regressed on the model. The overall model was significant $R^2 = .40$, $F(7,136) = 3.67$, $p = .001$. The model also revealed a significant main effect of stereotype threat ($\beta = .36$, $t = 4.54$, $p < .001$), demonstrating that participants in the stereotype threat condition were concerned that their performance would be judged based on the stereotype (see Table 6, pg. 81).

Individual Impression Concerns. Next, we examined the subscale that measured participants' concern about the impression their performance would portray about them as an individual. The overall model did not reach significance, $R^2 = .29$, $F(7,136) = 1.77$, $p = .098$. The exploratory analysis did however reveal a main effect of list such that participants in the mostly congruent list condition reported feeling less concerned about

the impression they were portraying as individuals than those participants in the mostly incongruent list condition ($\beta = -.21, t = -2.5, p < .05$) (see Table 5, pg. 80). The model also revealed a marginally significant three-way interaction between stereotype threat, list, and WMC ($\beta = -.15, t = -1.78, p < .077$).

Group Impression Concerns. Finally, the last performance concerns subscale measuring participants' concern in regards to image they were projecting of their group was regressed on the model. The overall model was significant $R^2 = .32, F(7,136) = 2.24, p = .035$. The model revealed a main effect of list, such that participants in the mostly congruent list reported less group performance concerns than those in the mostly incongruent list condition ($\beta = -.16, t = -2.26, p < .05$). The model also revealed a marginally significant interaction between list and WMC ($\beta = -.16, t = 1.89, p = .06$) such that, as WMC increased, group performance concerns increased for participants in the mostly congruent list but these same concerns decreased for those participants in the mostly incongruent list condition as WMC increased. Also the model revealed a significant three-way interaction between stereotype threat, list, and WMC ($\beta = -.18, t = -2.21, p < .05$). This interaction revealed, within the control condition that, as WMC increased, group concerns increased for those in the mostly congruent list ($r = .31$) but decreased in the mostly incongruent list ($r = -.39$). In contrast, within the stereotype threat condition as WMC increased group concerns increased, however these concerns did not significantly increase (r values $< .10$).

STUDY 2 DISCUSSION

Study 2 aimed to replicate Study 1 as well as to provide an understanding of what type of cognitive control strategy operates while a person is experiencing stereotype threat. Specifically, it was designed to determine if individuals under stereotype threat try to actively monitor their performance with proactive control strategies (early selection) or if they passively maintain the task goal by relying on reactive control strategies (late correction). Based on the dual processes of cognitive control (Braver et al., 2007), Study 2 tested whether stereotype threat increases the likelihood of engaging in reactive control, a less effective performance strategy on incongruent trials, whereby the goal is only activated after response conflict is detected or an error is committed. It was predicted that this would be especially true for those with lower WMC.

Study 2 tested two hypotheses. The first, the reactive control strategy hypothesis, would assume that stereotype threat increases the use of reactive control whereby participants do not actively monitor their performance, but instead reactivate the goal after detecting conflict between accessible responses. Presumably, such reactive control increases learning of S-R contingencies, due to suppressing word-reading, (i.e. learned associations between the correct response and irrelevant task dimensions) to aid performance. This late correction process would result in greater Stroop errors overall. In addition, this would also predict greater errors on mostly congruent trials under incongruent task conditions within a MC list, reflecting a greater ISPC effect due to learning S-R contingencies. The reactive control hypothesis would also predict that this would be most likely to occur for those lower in WMC. The second, proactive control

hypothesis, would predict that stereotype threat increases participant's likelihood of actively monitoring their performance by using a proactive control strategy. Thought of as an early selection process, participants actively maintain the task goal in order to decrease their mistakes. The proactive control strategy hypothesis predicted, only for those higher in WMC, that there would be less ISPC effects and less Stroop effects due to active goal maintenance and suppression of the irrelevant stimuli dimensions. Proactive control strategy use is distinguished by less interference from the irrelevant task information because it is goal driven, thus individuals only attend to the information that predicts successful performance. This results in less Stroop effects overall, as well as less ISPC effects because participants' attend to relevant task dimensions in service of the task goal. Thus, individuals using a proactive control strategy are less likely to learn S-R contingencies and are more likely to suppress the irrelevant dimensions of the task stimuli.

The preliminary analysis on critical items and Stroop effects did replicate past research on manipulations of ISPC suggesting that both controlled and automatic processes predict Stroop performance (Hutchison, 2011; Jacoby et al., 2003). Indeed, the traditional list-wide proportion congruency effects as well as ISPC effects, produced when ISPC was controlled, lends support to Hutchison's (2011) claim that list-wide proportion congruency effects are not ISPC effects in disguise.

The results of Study 2 did not replicate the findings from Study 1. Specifically the overall regression model for Stroop effects in errors did not approach significance. Although the results of Study 2 did not provide further support for Study 1's distraction

hypothesis, the list by congruency interaction was significant, replicating past research (Hutchison, 2011). A significant main effect of list when controlling for ISPC between the lists lends support to previous research, suggesting that list-wide proportion congruency is not just an artifact of ISPC effects which was confounded in previous Stroop manipulations (Hutchison, 2011; Jacoby et al., 2003; Kane & Engle, 2003). This also supports the belief that list-wide proportion effects are due to the use of top-down control strategies. These strategies aid in performance through a task demand unit that activates relevant task dimensions and suppresses irrelevant task dimensions in the service of the task goal. When these strategies are successful and the goal is actively maintained through controlled attention, less interference will occur resulting in fewer errors.

However, due to the fact that stereotype threat did not significantly effect Stroop performance, the question as to what type of strategy is employed while individuals' are experiencing stereotype threat is left unanswered. Due to these null results, no inferences can be drawn as to what type of cognitive control strategy operates while individuals are experiencing stereotype threat nor how WMC may moderate these effects. However, there were many changes to the manipulations in Study 2 that may have contributed to the null results that are discussed further in the general discussion.

New to Study 2 was the addition of several motivational self-report variables, and some intriguing patterns emerged revealing interesting implications for both the cognitive literature and stereotype threat literature in terms of motivation, goal adoption, task involvement, and perceived competence between tasks. For example, regardless of

whether participants were in the stereotype threat manipulation or not, as WMC increased participants reported higher levels of task involvement in the mostly incongruent list condition, but those within the mostly congruent list reported lower levels. This may suggest that when a task provides external goal maintenance, but on the surface appears to be a more difficult task (the conflict experienced when stating the ink color and not the word itself on frequent incongruent trials), individuals with higher levels of WMC are more engaged in the task. However, when the goal must be internally maintained, which actually makes the task more difficult but appear easier because of the ease of responding to frequent congruent stimuli, they may become bored. Cognitive literature to date has not investigated differences in reported task involvement on the Stroop task nor how WMC affects these perceptions.

Another interesting exploratory result was that of our perceived competence on the task. As participants WMC increased so too did their perceived competence on the task. However, the most intriguing result was that of the marginally significant list by stereotype threat interaction. Only those individuals in the stereotype threat mostly congruent list condition reported having higher levels of perceived competence. These same participants were also more likely to report having adopted a performance approach goal, increased performance concerns based on the stereotype, and less concern about the impression their performance was making of their group or themselves. The participants in the stereotype threat mostly incongruent list condition reported more concern about the image their performance was portraying about their group and themselves as individuals. These results may suggest that when a person is experiencing stereotype threat their

perceived competence increases when the task appears easier due to frequent exposures to mostly congruent stimuli which results in faster responses, but is actually the more difficult task due to the need for active internal goal maintenance. Increased perceived competence may contribute to the adoption of a performance approach goal. An alternative to this would be that when engaging in a difficult task, performance concerns based on the stereotype may cause the adoption of a performance approach goal, which in turn increases their perceived competence. The results of achievement goal adoption are contrary to past research that has shown a tendency for individuals experiencing stereotype threat to adopt performance-avoidance goals (Brodish & Devine, 2008; Smith, 2006; Smith & Johnson, 2006). However, as noted, past research in cognitive psychology has never explored these relationships between task involvement, perceived competence or achievement goal adoption. More research is needed in order to tease apart the possible interpretations and actual interactions between these variables within the context of WMC, the Stroop task, and stereotype threat.

Study 2 was designed to investigate whether individuals under stereotype threat utilize a proactive control strategy (active goal maintenance that is usually employed by individuals higher in WMC and which lead to successful task performance) or a reactive control strategy (reactivation of the goal only after conflict is detected, a strategy associated with lower performance and individuals with lower WMC). Unfortunately, Study 2 did not provide any conclusive evidence as to whether stereotype threat increases a person's use of reactive control, by which a person waits for an error to occur in order to trigger the need for more active goal maintenance. The results retain the null

hypothesis and do not provide evidence as to which type of cognitive control strategy operates when participants engage in a task while stereotype threat is present.

Specifically, it had been predicted that those that were experiencing stereotype threat and were lower in WMC would show evidence of using a reactive control strategy. This would have been demonstrated in increased intrusion errors overall, increased errors on mostly congruent trials, and decreased reaction time to item-specific trials. However, the exploratory analysis of Stroop error effects did reveal a significant main effect of list as well as did the exploratory analysis of Stroop error effects on mostly congruent items.

Participants' had significantly more intrusion errors on mostly congruent lists and mostly congruent items. In addition, the preliminary analysis revealed the typical ISPC effects as well as list-wide proportion congruency Stroop effects. Specifically, participants produced more errors on mostly congruent items and slower response latencies when under incongruent task conditions, but fewer errors on mostly congruent items and faster response latencies when under congruent task conditions. However, list-wide proportion congruency Stroop effects still emerged, even when ISPC was controlled. These results replicate past research and lend additional support that list-wide proportion congruency effects are not artifacts of ISPC effects in disguise (Hutchison, 2011).

GENERAL DISCUSSION

An abundance of research on stereotype threat effects has focused on determining how stereotype threat inflicts its deleterious effects. Investigations of this phenomenon have sought to find what mediating mechanisms influence negative stereotype threat effects and what variables may serve as moderators (for review see Nygugen & Ryan, 2008; Shapiro & Neuberg, 2007). However, much of the research has been restricted to what is causing individuals that are experiencing stereotype threat to be so adversely affected (for review see Smith, 2004). Due to overwhelmingly mixed results of research on this topic, researchers have proposed that there may be multiple mediating mechanisms involved in the experience of stereotype threat that predicts the outcome of performance situations (Schmader et al., 2009; Schmader, et al., 2008 Smith, 2004: Stone & McWhinnie).

Study 1 sought to determine if working memory is depleted because individuals are monitoring their performance, or whether they were so distracted by stereotype related cues and the process of trying to determine whether stereotype threat is present that they are left with limited attention to devote to the task. Based upon the dual processes of control framework (Kane & Engle, 2003), we used a well-established cognitive tool to conduct a straightforward examination of the *working memory depletion* account. This was accomplished by manipulating the proportion congruency of congruent and incongruent trials in the classic color Stroop task. In situations where the task at hand requires a great amount of attention in order to successfully complete the task, a distraction caused by stereotype-related thoughts possess many problems. If a stereotype

threat is operating, this could lead to a loss of the task goal, which will lead to increased errors on the task. Within domains in which achievement is measured based on how well an individual performs, such as earning a degree, this could lead to many consequential results. For example, if a person is highly identified with the degree they are pursuing, stereotype-related decreases in performance may increase perceptions of being less competent in the domain. Situations such as this might set in motion a chain of events that leads to a person distancing themselves from the domain. From a motivational perspective, increased errors on a task that is related to an aspect of a person's identity may increase the likelihood of performance-avoidance goal adoption, as well as decreased levels of task involvement and a decreased desire to pursue the domain in the future (Smith, 2004; Smith et al., 2007; Steele, 1997).

Study 2 was designed to investigate whether individuals under stereotype threat utilize a proactive control strategy (active goal maintenance that is usually employed by individuals higher in WMC and which lead to successful task performance) or a reactive control strategy (reactivation of the goal only after conflict is detected, a strategy associated with lower performance and individuals with lower WMC). Unfortunately, Study 2 did not provide any conclusive evidence as to whether stereotype threat increases a person's use of reactive control, by which a person waits for an error to occur in order to trigger the need for more active goal maintenance. The results retain the null hypothesis and do not provide evidence as to which type of cognitive control strategy operates when participants engage in a task while stereotype threat is present. However, the preliminary analysis revealed the typical ISPC effects as well as list-wide proportion

congruency Stroop effects. Specifically, participants produced more errors on mostly congruent items and slower response latencies when under incongruent task conditions, but fewer errors on mostly congruent items and faster response latencies when under congruent task conditions. However, list-wide proportion congruency Stroop effects still emerged, even when ISPC was controlled. These results replicate past research and lend additional support that list-wide proportion congruency effects are not artifacts of ISPC effects in disguise (Hutchison, 2011).

Limitations and Future Directions

Although the results of Study 2 did not replicate Study 1 or support the hypotheses, there were several significant but necessary changes in the manipulation to Study 2 that may have contributed to the null results. One major change was that of using the counting Stroop rather than the more well-known classic Stroop task. In order to trigger the stereotype that women perform poorly on tasks of quantitative ability, thus causing stereotype threat, the women had to believe that the task must pertain to the stereotype and resemble a stereotype relevant task. Thus, this necessitated the change from the more classic color Stroop, which resembles a task of verbal processing skills, to the counting Stroop, which resembles a quantitative ability task. Although very similar in composition in regards to congruency variations and trials, to date, past research has not manipulated ISPC within this version of Stroop.

It could therefore be possible, that even though the counting Stroop was constructed using exact replication of previously successful manipulations of ISPC that

utilized the classic color Stroop, the manipulations of ISPC within the counting Stroop may be fundamentally different. This may have caused variations in the phenomenological experience of the ISPC manipulation, possibly making the task exceptionally easy and lead to less need for active goal maintenance and response competition. For example, suppressing the habitual word reading response may pose to be more difficult while trying to actively maintain the goal of stating the color in the classic Stroop. In contrast, suppressing the response to read the number while maintaining the goal to state the quantity of the numbers, may not pose the same level of difficulty in the counting Stroop. This would result in significantly less Stroop effects in errors and reaction time in the counting Stroop as opposed to the color Stroop. Although the preliminary analysis did reveal significant Stroop effects and ISPC effects, which replicated past research (Jacoby et al., 2003; Hutchison, 2011), the ISPC manipulation in the counting Stroop may have been less sensitive and thus not produced the predicted results of stereotype threat effects.

Future research should investigate these possible variations in Stroop effects between the counting Stroop and the classic color Stroop. Investigations should also explore the possible differences in ISPC manipulations in the counting Stroop. For example, past research has demonstrated greater Stroop interference in the counting Stroop when the enumeration of the stimuli and response are closer in proximity (Pavese & Umiltà, 1998) However, Stroop interference decreases as the enumeration between the stimuli and response grow larger. It is possible that the very nature the counting Stroop does not allow for a successful manipulation of ISPC. If this is indeed so, future studies that investigate reactive control strategies may require the use of other conflict tasks, such

as the Antissaccade task, or manipulate ISPC within the color Stroop, but describing the task as a measure of spatial abilities or use a population (i.e. men) where the task can be described as stereotype-related task.

Another possible reason for the null results of Study 2 and unsuccessful replication of Study 1 is that of the women's level of domain identification with math. One exploratory variable that was included in Study 1 and Study 2 was participants' trait levels of domain identification which measures to what extent a participant identifies with the relevant domain (Smith, 2006; Smith & Johnson, 2006). Past research has demonstrated that women either do or do not identify with the math domain, and if they do not identify with the domain they do not care about their performance on the task. Both of these boundary conditions must be met in order for stereotype threat effects to occur. Thus, even a neutral score on the DIM scale would indicate that the women do not identify with math, and do not care about their performance (Smith, 2006; Smith & Johnson, 2006). Due to our sampling reporting very low DIM with math, it could very well be that they did not care about their performance and thus did not experience stereotype threat effects.

One last possible limitation to Study 2 is that due to the design of the study and the size of our sample, we risk the possibility to a Type II error. Since, the results consisted of data from 144 participants, which left an average of 10 participants per condition. Due to this, it is quite possible that our null results may do be due to a limit in statistical power rather than the lack of any genuine effect of stereotype threat. Future research can quite easily address this by testing a larger sample.

Implications

Although the results of the primary dependent variables did not reach significance in Study 2, the secondary dependent measures did reveal some interesting implications. The exploratory analysis of our secondary subjective measure provided intriguing insight for both the cognitive literature and stereotype threat literature. These insights are in terms of motivation, goal adoption, task involvement, and perceived competence between tasks of conflict. One measure that was particularly interesting and lends itself to the cognitive literature was that of reported task involvement. Regardless of whether participants were in the control condition or the stereotype threat condition, as WMC increased participants reported higher levels of task involvement in the mostly incongruent list condition. In contrast participants' within the mostly congruent list reported lower levels of task involvement as WMC increased. This suggests that when a task requires provides external goal maintenance for successful performance, but appears difficult on the surface because of repeated exposure two accessible responses (incongruent trials) participants with higher levels of working memory are more engaged in the task. In contrast, when the goal of the task requires internal goal maintenance because of repeated exposure to congruent stimuli, but appears easier on the surface (faster responses because of no conflict between two accessible responses) participants report being less engaged in the task as WMC increases. One theory that maps onto these results is that of the need for cognition (Cohen, Stotland, & Wolfe, 1955). Individuals that possess a higher need for cognition tend to seek out and enjoy tasks that are challenging versus those that are lower in the need for cognition. This need for cognition

may possibly explain increased perceptions task involvement in the mostly congruent list as WMC increases. Future research should investigate subjective measures of task involvement between mostly congruent and mostly incongruent list conditions as well as how trait levels of the need for cognition may contribute to perceptions of task involvement.

Another interesting finding was that of participants' perceived competence on the task. The results of our exploratory analysis showed that as participants working memory capacity increased so too did their perceived competence on the task. What presented to be the most interesting however, was the marginally significant list and stereotype threat interaction for perceived competence on the task. These higher levels of reported perceived competence only occurred for those participants who were in the stereotype threat mostly congruent list condition, which is considered to be a more difficult task by researchers but appears easier to participants. In addition these same participants were more likely to report having adopted a performance approach goal. Participants in this condition, as compared to the other three conditions, were also more likely to report higher levels of concern that their performance would be judged based on the stereotype than any other performance impression concerns. In contrast, the participants in the stereotype threat mostly incongruent list condition reported more concern about the image their performance was portraying about their group and themselves as individuals. These results may suggest that when a person is experiencing stereotype threat their perceived competence increases because on the surface the task appears to be easier (i.e. they are responding faster), but is in actuality more difficult because it requires active

goal maintenance. This increased perception of perceived competence may cause them to report that they had adopted a performance-approach goal, even though these individuals were also making more errors. It is also possible that when engaging in a difficult task while worrying about being judged based on the stereotype may cause the adoption of a performance-approach goal, in which the goal and motivation to perform better increases their perceived competence.

Although contrary to past research (Smith et al, 2007), these results present an interesting possibility. Adoption of a performance-approach goal, which would suggest a desire to perform better compared to others, as opposed to the adoption of a performance-avoidance goal, which suggests a desire to avoid the appearance of failing in comparison to others, would be preferable in most situations. However, if a person possesses higher levels of perceived competence as well as increased task involvement due to the adoption of a performance-approach goal, finding out that they actually performed worse in comparison to others may pose problematic. Upon discovering that performance on a task in which there is higher levels of perceived competence and task involvement was actually worse in comparison to others may lead to decreased desire to pursue the domain in the future. However, more research is needed in order to tease apart the possible interpretations and actual interaction between the perception of task difficulty, perceived competence, and performance approach goals.

Although more research and replication is needed to fully support the findings, the results from Study 1, as well the results of task involvement from Study 2, both can be incorporated with research on mind-wandering (McVay & Kane, 2009). The mind-

wandering literature has demonstrated individuals that are lower in WMC are more prone to mind-wandering, which predicts increased intrusion errors due to task irrelevant thoughts interfering with active maintenance of the task goal (McVay & Kane, 2009). In addition, research on mind-wandering in the stereotype threat literature has demonstrated that this mind-wandering plays an essential role in the relationship between stereotype threat and negative performance effects (Mrazek, et al., 2011). Integrating the results from Study 1 with the mind-wandering account would suggest that when a person has lower WMC and is experiencing stereotype threat, their mind may begin to wander during the task due to intrusions of stereotype-relevant thoughts. This impairs the ability to actively maintain the goal of the task and suppress inappropriate responses.

In addition, the results from Study 2 on the reported decreases in task involvement from participants that were lower in WMC and in the stereotype threat mostly congruent list condition may also suggest that mind-wandering has occurred because of focusing on stereotype-relevant thoughts, resulting less task engagement. Although it is still unclear whether mind-wandering is due to metacognitions of performance (Schmader, et al., 2009), bias detection (Kaiser et al., 2006), or interpretation of internal affective responses (Schwarz & Clore, 1983) future research should focus on integrating these results with the mind-wandering literature to achieve a better understanding of the effects of intrusive thoughts on performance and working memory.

The inability for those individuals experiencing stereotype threat to maintain the task goal sheds insight onto what cognitive process is triggered and how it may be contributing to performance as well as motivation. Although our results did not provide

support for the response competition hypothesis, future research should aim at replicating these results with other stereotypes (i.e. race related competence or age memory loss) as well as with populations that are more commonly victims of stereotype threat (i.e. African Americans and Latinos/as). It is possible that between different stereotypes and stereotyped populations the phenomenological experience may vary and produce more/less differences in the ability to maintain the task goal and deficits in response competition may emerge. In addition depending on the manner in which the stereotype threat is triggered may also affect the experience of stereotype threat as was outlined in the introduction section.

These differences in the presentation of stereotype threat may also result in differences in impairment of goal maintenance or response competition. As the “dual process’ model proposed by Stone and McWhinnie (2008) suggests, depending on whether the stereotype is presented in a subtle manner or a more blatant manner may result in differences in the level of stereotype threat activation which causes stereotype processes affecting performance. Specifically, when a subtle manipulation is employed, such as having participants indicate the gender or race, negative performance effects are due to working memory impairments because the individual is trying to detect any bias. In contrast when stereotype threat is triggered with a blatant manipulation, such as explicitly stating woman perform poorly on a math exams, performance is impaired due to worry and the adoption of performance avoidance goals (Smith et al., 2007). Study 1 tested our hypothesis using a subtle manipulation on a majority group that are less commonly the targets of stereotype threat (see Smith & White, 2002). The results still

however suggest that even when the stereotype is not fully clouding the air that the distraction hypothesis still holds true.

Study 1 provided research that allows the stereotype threat literature to move past just stating that negative stereotype threat performance effects are due to depletion of cognitive resources. Specifically, stereotype threat interferes with a person's ability to maintain the goal of the task, especially if they have a lower WMC. Stereotype related thoughts increase the likelihood of mind-wandering in individuals with lower WMC which results in increased goal neglect impairing performance on a variety of attention demanding situations.

Individuals that are targets of stereotyping are likely to encounter stereotypes in a variety of situations ranging from seeing a derogatory cartoon in the newspaper (Oswald & Harvey, 2000), watching television commercials (Davies, et al., 2001), to situations where their group representation is minimal (Inzlicht & Ben-Zeev, 2000). In addition these same stigmatized individuals are more likely to seek out and attend to environmental cues in the environment that may indicate that a stereotype threat is present (Kaiser et al., 2006; Murphy et al., 2007). Not only do stereotyped groups encounter stereotypes in all different types of situations triggered in a multitude of ways, stereotype threat does not just affect performance on stereotype relevant tasks. Stereotype threat has been evidenced to interfere with financial decision making (Carr & Steele, 2010), generation of problem solving strategies (Quinn & Spencer, 2001), aspiring to advance to leadership roles in work settings (Davies, Spencer, & Steele, 2005), as well as spilling over into non-stereotypic behaviors such as aggression and eating behaviors in

women after a difficult math task (Inzlicht & Kang, 2010). Thus it is impossible to rid the world of stereotypes and extremely difficult to nullify the stereotype in every situation. Stereotype threat interventions have aimed at educating the stereotyped group and nullifying the stereotype, alleviating the pressure caused by the testing situation, focusing on multiple identities, and motivating a promotion focus (DeCaro, Rotar, Kendra, & Beilock, 2010; Gresky, Ten Eyck, Lord, & McIntyre, 2005; Johns, Schmader, & Martens, 2005; Keller & Dauenheirmer, 2003). However, our results suggest a need for interventions to be based around changes in the environment to decrease situational cues that create distraction and the need for tasks constructed in such a way that externally maintains the task goal. By providing a better understanding of how stereotype threat operates, researchers and educators alike can begin to focus on ways to minimize distraction that impairs cognitive resources when devising interventions as well as strategies to decrease mind-wandering. For example, devising exams where the goal is externally maintained. This could be accomplished by removing cues from the environment that can trigger stereotype threat, such as requiring individuals to indicate gender and/or race on an exam, or simply naming the task something novel that does not relate to the stereotype. This is particularly important because even the most structured of testing situations (e.g., GRE, ACT, SAT) create situations where distraction can occur due to stereotype threat being triggered by marking race and gender on the test. Our results suggest that this will be especially true for those populations already prone to mind-wandering and distraction (i.e. lower levels of WMC and individuals with ADHD). However the findings also demonstrate that mind-wandering and distraction will also

affect individuals at times when working memory is not at full capacity such as when attention is divided, during ego-depletion (Carr & Steele, 2010), or during varying times of a person's circadian rhythm (Hasher, Zacks, & May, 1999). These negative stereotype threat effects target a wide range of populations as well as encompass a variety of different stereotypes within these populations. Thus having a better understanding of how stereotype threat affects an individual, and what process it triggers, is important for decreasing the performance gaps between various social groups.

One reason in particular that demands a better understanding of stereotype threat is the gaps that this phenomenon creates in socioeconomic status (SES). SES is determined by many factors such as income, occupation, and education all of which are associated with access to resources, for example health care (American Psychological Association, 2011). The most important in regards to stereotype threat effects is how it negatively affects performance in education settings for stigmatized groups such as less access to higher education, less motivation to pursue higher education in areas related to the stereotype relevant tasks, and higher dropout rates (American Psychological Association, 2011). By knowing how stereotype threat occurs and how it affects cognitive resources, researchers can establish interventions to be taught in the classroom or during testing periods to decrease stereotype threat effects. Our research has provided evidence of stereotype threat's true disruptive nature on working memory. Now the idea that stereotype threat disrupts working memory by interfering with the individual's ability to maintain task goals in the face of distraction may be largely admitted.

REFERENCES CITED

- American Psychological Association (2011). *Socioeconomic status*. Retrieved from <http://www.apa.org/topics/socioeconomic-status/index.aspx>.
- 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. doi: 10.1006/jesp.1998.1371
- Baddeley, A. D. (1998). The central executive: A concept and some misconceptions. *Journal of the International Neuropsychological Society, 4*, 523-526.
- Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science, 17*, 339-343. doi: 10.1111/j.1467-8721.2008.00602.x
- 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., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the casual 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, alleviations, and spillover. *Journal of Experimental Psychology: General, 136*, 256-276.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652. doi: 10.1037//0033-295X.108.3.624
- Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. Conway, C., Jarrold, M. J., Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 76-106). New York, NY, US: Oxford University Press.
- 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. doi: 10.1016/j.jesp.2008.08.005.

- 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.
- Bugg, J. M., Jacoby, L. L., & Toth, J. P. (2008). Multiple levels of control in the Stroop task. *Memory & Cognition, 36*, 1484-1494. doi: 10.3758/mostlycongruent.30.8.1484.
- 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.
- Carr, P. & Steele, C. M. (2010). Stereotype threat affects financial decision making. *Psychological Science, 21*, 1411-1416.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332-361.
- Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for cognition. *The Journal of Abnormal and Social Psychology, 51*, 291-294. doi: 10.1037/h0042761
- Croizet, J., Despres, G., Guazins, M., Huguet, P., Leyens, J., & Meot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin, 30*, 721-731. doi: 10.1177/0146167204263961.
- Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*, 1639-1655. doi:10.1111/j.1559-1816.2008.00362.x
- Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*, 1615-1628.
- Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology: Interpersonal Relations and Group Processes, 88*, 276-287. doi: 10.1037/0022-3514.88.2.276.
- DeCaro, M. S., Rotar, K. E., Kendra, M. S., & Beilock, S. L. (2010). Diagnosing and alleviating the impact of performance pressure on mathematical problem solving.

The Quarterly Journal of Experimental Psychology, 63, 1619-1630. doi: 10.1080/17470210903474286.

- De Pisapia, N. & Braver, T. S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing*, 69, 1322-1326. doi: 10.1016/j.neucom.2005.12.100.
- Elliot, E. S. & Murayama, K. (2008) On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, 100, 613-628.
- Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121, 480-506. doi: 10.1037/0096-3445.121.4.480.
- Gresky, D. M., Ten Eyck, L. T., Lord, C. G., & McIntyre, R. B. (2005). Effects of salient multiple identities on women's performance under mathematics stereotype threat. *Sex Roles*, 53, 703-716. doi: 0360-0025/05/1100-0703/0
- Harkins, S. G. (2006) Mere effort as the mediator of the evaluation-performance relationship. *Interpersonal Relations and Group Processes*, 91, 436-455. doi: 10.1037/0022-3514.91.3.436.
- Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and cognition. In D. Gopher, & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 653-675). Cambridge, MA: The MIT Press.
- Hutchison, K. A. (2011). The interactive effects of listwide control, item-based control, and working memory capacity on Stroop performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 851-860. doi: 10.1037/a0023437.
- 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. doi: 10.1111/1467-9280.00272.
- Inzlicht, M. & Kang, S. K. (2010). Stereotype threat spillover: How coping with threats to social identity affects aggression, eating, decision-making, and attention. *Journal of Personality and Social Psychology*, 99, 467-481. doi: 10.1037/a0018951.
- Jacoby, L. L., Lindsay, D. S., & Hessels, S. (2003). Item-specific control of automatic processes: Stroop process dissociations. *Psychonomic Bulletin & Review*, 10, 638-644.

- Jamieson, J. P. & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, *93*, 544-564. doi: 10.1037/0022-3514.93.4.544.
- Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, *137*, 691-705.
- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, *16*, 175-179. doi: 10.1111/j.0956-7976.2005.00799.x
- Kaiser, C. R. & Hagiwara, N. (2011). Gender identification moderates social identity threat effects on working memory. *Psychology of Women Quarterly*, *35*, 243-251. doi:10.1177/0361684310384102.
- Kaiser, C. R., Vick, S. B., & Major, B. (2006). Prejudice expectations moderate preconscious attention to social identity threatening cues. *Psychological Science*, *17*, 332-338.
- 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 task interference. *Journal of Experimental Psychology: General*, *132*, 47-70.
- 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. doi: 10.1023/A:1021003307511
- Keller, J. & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin*, *29*, 371-381. doi: 10.1177/0146167202250218.
- Kiefer, A. & Shih, M. (2006). Gender differences in persistence and attribution in stereotype relevant contexts. *Sex Roles*, *54*, 859-868.
- Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, *71*, 1092-1107. doi: 10.1037/0022-3514.71.6.1092
- Marx, D. M., & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on targets' test performance and subjective experience. *British Journal of Social Psychology*, *44*, 645-657.

- McVay, J.C., & Kane, M.J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 196-204.
- Mrazek, M.D., Chin, J.M., Schmader, T., Hartson, K.A., Smallwood, J., & Schooler, J.W. Threatened to distraction: Mind-wandering as a consequence of stereotype threat. *Journal of Experimental Social Psychology* (2011), doi: 10.1016/j.jesp.2011.05.011.
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*, 879-885.
- Nguyen, H.-H. D. & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314-1334.
- Osbourne, J.W. (2001). Testing stereotype threat: Does anxiety explain race and gender differences in achievement. *Contemporary Educational Psychology*, *26*, 291-310. doi: 10.1006/ceps.2000.1052.
- Oswald, D. L. & Harvey, R. D. (2000). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology: Developmental, Learning, and Personality*, *4*, 338-356.
- Pavese, A. & Umiltà, C. (1998). Symbolic distance between numerosity and identity modulates Stroop interference. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1535-1545.
- 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.
- Regner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J. M., & Huguet, P. (2010). Individual differences in working memory moderate stereotype-threat effects. *Psychological Science*, *21*, 1646-1648. doi:10.1177/0956797610386619.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Psychology*, *38*, 194-201. doi: 10.1006/jesp.2001.1500.

- Schmader, T., Forbes, C.E., Zhang, S., & Mendes, W.B. (2009). A meta-cognitive perspective on cognitive deficits experienced 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., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336-356.
- Schneider, W., Eschman, A. & Zuccolotto, A. (2020). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*, 513-523.
- Shapiro, J. R. & Neuberg, S. L. (2007). From Stereotype threat to stereotype threats: Implication of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality of Social Psychology Review*, *11*, 107-130. doi: 10.1177/1088868306294790.
- Siebt, B. & Forster, J. (2004). Stereotype threat and performance: How self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology*, *87*, 38-56.
- Smith, J. L. (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. L. (2006). The interplay among stereotypes, performance-avoidance goals, and women's math performance expectations. *Sex Roles*, *54*, 287-296.
- Smith, J. L. & Johnson, C. S. (2006). A stereotype boost or choking under pressure? Positive gender stereotypes and men who are low in domain identification. *Basic and Applied Social Psychology*, *28*, 51-63.
- 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, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61*, 1040-1058.
- Smith, J. L. & White, P. H. (2002). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology, 99*, 99-114. doi: 10.1037/0022-0663.99.1.99.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28.
- Steele, 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., Spencer, S.J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34*, 379-440.
- 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.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.
- 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 Selst, M., & Jolicoeur (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology, 47A*, 631-650.