Theses and Dissertations (All)

9-16-2009

Stereotype Threat in Text Comprehension

Cole McCracken
Indiana University of Pennsylvania

Follow this and additional works at: http://knowledge.library.iup.edu/etd

STEREOTYPE THREAT IN TEXT COMPREHENSION

A Dissertation Submitted to the School of Graduate Studies and Research in Partial

Fulfillment of the Requirements for the Degree Doctor of Psychology

Cole McCracken

Indiana University of Pennsylvania

August 2009

Indiana University of Pennsylvania
The School of Graduate Studies and Research
Department of Psychology

We hereby approve the dissertation of

Cole McCracken

Candidate for the degree of Doctor of Psychology

_____   _____
                                         Susan Zimny, Ph.D.
                                         Professor of Psychology, Advisor

_____   _____
                                         Donald Robertson, Ph.D.
                                         Professor of Psychology

_____   _____
                                         Lynda Federoff, Ph.D.
                                         Professor of Psychology

ACCEPTED

_____   _____
Michele S. Schwietz, Ph.D.
Assistant Dean for Research
The School of Graduate Studies and Research

Title: Stereotype Threat in Text Comprehension
Author: Cole McCracken

Dissertation Chair: Dr. Susan Zimny

Dissertation Committee Members: Dr. Donald Robertson

                                    Dr. Lynda Federoff

This study explored the phenomenon of stereotype threat in a population of

undergraduate psychology students at a small, rural, state university in Pennsylvania. The

primary goal was to attempt to establish a clear and concise mechanism of operation to

explain the origin of the phenomenon, using Walter Kintsch's (1986) Construction-

Integration model of text processing as a base.

Participants were asked to read a series of brief texts. Following each text,

participants were asked to determine whether or not a number of sentences had appeared

in the preceding texts. The analysis of the pattern of reading times and recognition of

sentences was used to test the hypothesis that participants would experience disruption of

performance, primarily through the activation of threatening evaluative information at the

situation model level of representation. Long-Term Working Memory structures

associated with stereotypic information were thought to be a key factor in decreased

processing speed and memory disruption under conditions of stereotype threat.

Black participants primed to approach a task as if it were a measure of their

intelligence were expected to suffer performance deficits through the activation of task-

unrelated memory constructs associated with stereotypes. Meanwhile, those individuals

operating under stereotype threat who were exposed to a stereotype-relevant text were

expected to show heightened performance, due to the relevance of the stereotypic information.

The results from the analyses of the reading time and recognition data suggested a consistent pattern of activation of stereotype threat among the White participants, with little reaction to this manipulation among the Black students. Black students were noted to consistently focus their attention on the surface level of representation, while the White students focused primarily on the situation model level. This pattern had not been previously investigated in stereotype threat research and several potential explanations were discussed. Of significance, these patterns would not have become evident were it not for the application of a well-defined model of text processing, such as the Construction-Integration model.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The purpose of the current research project is to attempt to provide an explanation of the cognitive mechanisms behind the phenomenon of "stereotype threat." Stereotype threat is defined by Steele (1999), one of the original researchers into the phenomenon, as "the threat of being viewed through the lens of a negative stereotype, or the fear of doing something that would inadvertently confirm that stereotype" (p. 46). Stereotype threat has been implicated as potentially being one of the primary contributors to the persistent achievement gap noted between men and women, as well as between Caucasian Americans and a variety of ethnic and racial minority groups. In addition to understanding the obvious socio-cultural ramifications of stereotype threat, the educational and psychological fields must be aware of these effects in order to improve assessment measures and techniques.

This research project will incorporate a number of distinct areas of research literature in an attempt to identify the mechanisms underlying stereotype threat. The first section will be dedicated to the initial wave of research establishing stereotype threat as a replicable phenomenon. The second section will focus on exploring previously suggested mechanisms by which stereotype threat may operate. The third section will address anxiety, one of the most commonly cited potential mechanisms underlying the stereotype threat process. This section will focus on establishing a solid, operational definition of anxiety, and on exploring how it may produce stereotype threat effects. Then a brief overview of Kintsch's (1988, 1998) Construction-Integration (CI) model of text comprehension will be presented, in which the potential avenues by which stereotype

threat may operate in a reading comprehension task will be highlighted. Finally, a relatively new theoretical advance in memory research, Ericsson and Kintsch's (1995) theory of Long Term Working Memory (LTWM) will be discussed as it pertains to stereotype threat.

CHAPTER 2

REVIEW OF LITERATURE

Stereotype Threat

There exists a substantial body of literature on the psychological phenomenon

dubbed stereotype threat. The first wave of research into this phenomenon focused on

establishing its existence. The second wave explored factors that may explain its effects.

The purpose of the current research project is to investigate how it is that a threat of

confirming a negative stereotype regarding the participant's in-group actually acts to

impair the participant's cognitive performance, and to describe this process in greater

detail than has been the case in the existing literature.

The results of standardized testing have long indicated a discrepancy in

performance between African Americans and Caucasian Americans. This divergence has

also been seen in school achievement, grades, and retention/graduation rates for all levels

of academic achievement, from kindergarten through college (Steele & Aronson, 1995).

These differences in mean performance levels have been noted in areas outside of

academia as well, including general cognitive ability testing (Mayer & Hanges, 2003;

Nguyen, O'Neal, & Ryan, 2003). Likewise, similar performance differences have been

noted between Caucasian Americans and Latino Americans (Schmader & Johns, 2003).

Not limited to matters of race, disparities in performance have also been seen on

standardized testing in the realm of mathematics and mathematics-based sciences

between males and females of diverse ethnicities (Cadinu, Maass, Frigerio, Impagliazzo,

& Latinotti, 2003; Hufnagel, Deming, Landato, & Hodari, 2004; Quinn & Spencer,

2001). These observed differences were consistent with commonly held stereotypes

regarding the performance of each stigmatized group, whether inclusion was based on gender or race.

Stereotype threat begins with the existence of stereotypes within a culture. All people utilize cognitive heuristics to simplify their daily lives. Although this tendency saves people from having to use finite cognitive resources to store information on people with whom they have limited contact, it results in generalizations to social groups based on limited information made readily available through the dominant culture (Dovidio, Kawakami & Gaertner, 2000). It is no exaggeration to claim that all people are members of a group about which some stereotype has been formed, be group membership based on race/ethnicity, gender, religion, sexual orientation, socioeconomic status (SES), or some other characteristic. These stereotypes can be positive or negative, but this research will focus on the negative stereotypes. If these negative stereotypes are sufficiently pervasive to create a general societal perception of their truth, those living under these stereotypes can suffer significant detriments (Steele, 1999)**.**

The focus for the present research was the deficits in cognitive performance brought about through the threat of confirming a negative stereotype. The basic concept of stereotype threat, as articulated by Steele and Aronson (1995), was that members of a stigmatized group feared confirming the negative stereotype. When these individuals were led to believe that a particular measure was indicative of their level of ability, they then operated under a stereotype threat, and became highly concerned with avoiding this potential confirmation. This avoidance reaction was then thought to lead to behavioral changes that altered performance on a cognitive task. The cognitive mechanism by which this operated was not the specific target of Steele and Aronson's original work, but was

4

hypothesized to be a decline in the processing efficiency of those individuals who were operating under the stereotype threat condition. The authors noted that this decline was similar to that seen in prior research regarding test anxiety, and likely involved the participants' frequent shifting of attention from the task of providing accurate answers, to the task of assessing the situation for any internal, in-group relevant explanation for their difficulty on the measure.

The first issue addressed in the consideration of stereotype threat was whether it was a real phenomenon or not. Steele and Aronson's (1995) seminal work in the area examined the impact that stereotype threat had on the performance of Black college students on a difficult verbal task. This was accomplished through a series of four experiments.

Their first study was a 2 (Black, White) x 3 (threat, challenge, non-challenge) design, and examined both Black and White students' responses to high-level verbal items taken from the Graduate Record Exam (GRE). Stereotype threat was manipulated by placing participants into one of three groups. The stereotype threat condition participants were told that the measure was highly indicative of intellectual ability, and the items were difficult because it was necessary to provide an accurate test of the participants' verbal abilities. The experimenters presumed that this would induce a stereotype threat response in African American participants. Participants in the challenge condition were told that the measure was designed to examine psychological variables involved in solving verbal problems, and was not representative of verbal abilities. Furthermore, they were told that the task would be difficult, but that the difficulty was best viewed as a challenge to be overcome. The final, non-challenge group was also told

that the test would be difficult, but that the difficulty was inherent in the test design. This division into challenge and non-challenge groups reflected the experimenters' desire to examine whether an emphasis on the challenge of a task may enhance task performance (Steele & Aronson, 1995).

In the threat condition, Black students underperformed in comparison to White students. No other levels of the race by condition effects reached significance. The race by condition interaction effect was noted to be only marginally significant however, and then only when the participants in the challenge condition were excluded from the analysis. This weak support for the existence of stereotype threat said little about how it operated. Self-report measures of academic competence, self-worth, disruptive thoughts, and ratings of test bias resulted in only one main effect. African American participants consistently rated the test as being more biased than did Caucasian American participants. The experimenters' hypothesis that distracting thoughts may disrupt performance for those under threat conditions was not supported (Steele & Aronson, 1995).

The second study replicated the results of the first study, and expanded upon them by recording time spent per item, and assessing for the participants' level of post-test anxiety using a self-report measure. Self-reported levels of potential stereotype threat mediators were also requested post-test. These included effort expended, extent of guessing, persistence, time pressure, extent of rereading, test bias, and frustration/giving up. None of these self-reported variables significantly affected test performance. Likewise, the anxiety results were not significant (Steele & Aronson, 1995).

The same threat condition by race interaction effect noted in the first study was replicated. The effect was also more robust in this study than previously. Furthermore, Black participants in the threat condition completed significantly fewer items than did White participants or members of both races in the non-threat condition. Thus the stereotype threat effects of the test impaired both speed and accuracy for African American participants. The experimenters believed that this did not occur in the prior study because they had shortened the amount of time allowed in the second experiment. Because the test items were arranged in order of decreasing difficulty, the easier items near the end of the test were presumed to have not been reached. This would have amplified performance differences between the two groups. This observation was supported by the fact that Black participants in the threat condition took an average of 94 seconds to answer each of the first five test items. Those participants of both races in the non-threat condition took between 71 and 73 seconds (Steele & Aronson, 1995).

Steele and Aronson (1995) suggested that the decreased processing speed was similar to that seen in test anxiety. This was consistent with their conceptualization of stereotype threat as having its effect through the induction of an avoidance response.

The third study added several elements to the established pattern. These included a measure of the activation of racial stereotype knowledge, a measure of stereotype-related self-doubts, a measure of dissociation from stereotypically relevant ideas and activities, and a measure of self-handicapping. The experiment was conducted with a 2 (Black, White) x 3 (threat, non-threat, control) design. The participants in the threat and control conditions were led to believe they would be completing two verbal measures, and would be evaluated on their performance. Those in the non-threat condition were told

they would complete these tasks, but that their personal performance would not be evaluated. Both groups were then shown a sample item from a fragment completion task, and three samples from a set of verbal GRE tasks that were similar to those used in previous studies. The control group was informed that the experimenter could not be present, and were asked only to finish a fragment completion task and a stereotype avoidance or dissociation measure, described below. They were not led to believe there would be any further measures (Steele & Aronson, 1995).

The participants in the threat and non-threat conditions first completed the word-fragment task. This was designed to be a measure of cognitive activation of stereotype relevant concepts. The authors sought to confirm that those operating under stereotype threat conditions were more aware of the stereotypes. The task involved adding letters to word fragments to make complete words. Many of these fragments could have been completed in ways that would relate either to race or to negative racial stereotypes about African Americans (B-L-A-C-K, B-R-O-T-H-E-R, L-A-Z-Y). This task also included several fragments that could be completed in a fashion that indicated significant self-doubt (L-O-S-E-R, D-U-M-B; Steele & Aronson, 1995).

Participants were then asked to complete a number of measures designed to determine the degree to which they may be avoiding stereotype confirmation. This included rating a number of activities (including stereotype-laden activities like rap music and basketball) in terms of preference. They were also asked to rate a number of possible descriptors (including stereotypical terms like aggressive) in terms of degree of self-identification. A final dissociation measure came in the form of a demographic questionnaire, on which they were given the option of recording their race or leaving that

question blank. The final additional measure was viewed as a self-handicapping pretest. Participants were given the opportunity to rate a number of areas that may impair their performance on the upcoming test. These included such things as focus, physical health, and fatigue. As Steele and Aronson (1995) predicted regarding the fragment completion task, an interaction resulted between race and condition. Blacks in the stereotype threat condition produced more words pertaining to race or racial stereotypes, as well as to self-doubt statements, than did Blacks in the non-stereotype threat condition or Whites in either condition. The second set of measures addressed the authors' hypothesis that those in the threatened condition would protectively dis-identify with stereotypes. As predicted, Black students in the threat condition avoided self-identifying with both activities and personality traits traditionally linked to racial stereotypes. This was not true for either Blacks in the non-threat condition or Whites in either condition. When given the option of reporting race on a demographic questionnaire, only 25% of Blacks in the threatened condition reported their race, whereas 100% of all participants in other conditions reported it. This combination of results supported the hypothesis that those operating under conditions of stereotype threat experienced an avoidance or dissociation reaction (Steele & Aronson, 1995).

The final addition to this third study involved a post-test report of self-handicapping. The authors' hypothesis had been that those who were operating under stereotype threat conditions would be less certain of their chances of success, and would thus engage in self-handicapping before the test. Black students in the threat condition reported fewer hours of sleep, less ability to focus, and an increased perception of test

bias. This supported the authors' hypothesis regarding self-handicapping (Steele & Aronson, 1995).

The fourth and final study examined the ability of stereotype threat to interfere with performance using a very subtle manipulation of stereotype threat. The procedures were similar to the non-threat conditions of the earlier experiments. Half of the participants were asked to complete a brief questionnaire that contained race as an item, whereas the other half completed the same questionnaire with the race item removed. The verbal test from the first two studies was then administered, followed by a series of self-report measures addressing confidence, guessing, effort, time limitations, persistence, rereading, frustration, and perception of test bias. Participants also completed a self-report assessment of their degree of agreement or disagreement with a series of statements that addressed level of stereotype threat experienced. These included such statements as "some people feel I have less verbal ability because of my race," and "my race does not affect people's perception of my verbal ability." Several other items assessed the participants' degree of self-identification with mathematical and verbal abilities (Steele & Aronson, 1995).

An interaction between race and condition resulted. Black participants in the racially-primed condition performed worse than participants in the other conditions, including the non-racially primed Black group. They completed fewer items, and were less accurate. Although the completion scores reached significance, the accuracy effects only reached a $p < .08$ level of significance. This trend was consistent with the similar speed and accuracy data noted in the earlier studies. This occurred in spite of using only a

very minimal racial prime, which implied a process of automatic activation anytime that race became relevant (Steele & Aronson, 1995).

As expected on the self-report measures of subjective stereotype threat experience, Black participants reported a higher degree of stereotype threat than did White participants. Against predicted results, African Americans in the racially primed category did not show an elevated activation of stereotype threat in comparison to those Blacks in the non-primed condition. This result is in contrast with the results of the fragment completion task in Experiment 3, where those Black participants in the stereotype threat condition showed an increased tendency towards activation of stereotype-relevant knowledge that was expected to correlate with stereotype threat. The authors suggested this failure to show an increase in perception of stereotype threat with a minimal prime reflected a process of dis-identification and denial of stereotypes by those in the primed condition. Finally, an increase in anxiety was noted for those operating under stereotype threat. This discrepancy with the results of Experiment 2 was suggested to result from differences in the length of the two tasks, with the decreased time allowed for the task in the final experiment perhaps exacerbating the participants' anxiety, and possibly producing threat effects through a different mechanism (Steele & Aronson, 1995).

In an effort to expand the scope of stereotype threat research, Croizet and Claire (1998) approached the issue of stereotype threat using a socioeconomic stereotype. Their hypothesis was that a "poverty = ignorance" stereotype could be used to induce similar performance deficits to those noted for African Americans. Each participant was asked to complete a pre-test questionnaire containing basic, demographic information. This pretest

was used to access university records in order to covertly classify the students in terms of SES. The questionnaire also included a self-handicapping section similar to that used by Steele and Aronson (1995). After completing the questionnaires, the participants were asked to complete a series of difficult, high-level items taken from the Verbal portion of the GRE (Croizet and Claire).

The experiment was a 2 (high SES, low SES) x 2 (SES primed, SES not primed) x 2 (threat, non-threat) design. Division into low and high SES populations was made covertly based on parental occupation information contained in student records, with those in the middle SES ranges excluded from data analysis. Priming of SES was achieved by asking those in the SES primed condition to provide information on parental education and occupation. Manipulation of stereotype threat was accomplished by informing those in the threat condition that the task was designed to get an accurate assessment of their verbal abilities. Those in the non-threat condition were informed that the task was designed to examine the role that attention plays in lexical memory (Croizet and Claire, 1998).

The analysis of the performance data on the verbal measure supported the primary experimental hypothesis. Those participants in the low SES condition performed worse than other participants only when the measure was presumed to be diagnostic of verbal ability. In addition, low SES participants in this stereotype threat condition completed fewer items in the time allotted. These data supported the generalizability of the stereotype threat concept to economic stereotypes. However, these results again did not address the mechanisms behind the threat effects (Croizet & Claire, 1998).

The results of the self-handicapping measures showed that, contrary to experimenter expectation, those in the low SES group did not actually show an increased tendency to self-handicap in anticipation of failure. The hypothesis that self-handicapping may impair performance for those operating under stereotype threat was not supported. The authors saw this as evidence that stereotype threat operates at a subconscious, automatic level (Croizet & Claire, 1998).

These results however, were in contrast to the results of Steele and Aronson's (1995) Experiment 3, in that Black students operating under conditions of stereotype threat did engage in self-handicapping. The discrepancy may have been a function of the populations involved. The Steele and Aronson research involved African American participants, whereas the Croizet and Claire (1998) research involved French students. The possibilities of cultural differences in self-handicapping, or categorical differences in types of stereotype threat cannot be ruled out.

Croizet and Claire (1998) had predicted that participants asked to report their parental education and occupation (the SES primed condition) would have an increased awareness of SES, and would be more susceptible to stereotype threat effects. Although priming the stereotype had the expected effect, priming SES did not have an effect on the outcome. One possible interpretation offered by the authors was that, for participants in a low SES group, knowledge of their SES status is always salient at some cognitive level. That is, priming this serves little purpose, because it is always primed in the minds of in-group members. In essence, the members of this group are always aware of their relative poverty, though not necessarily always attending to relevant stereotypes. If this were the

case, it would be reasonable to expect that members of racial minority in-groups would be similarly aware of race.

While race has been the focus of much of the stereotype threat research, a body of literature addressing stereotypes regarding women and their abilities to perform on mathematical tasks has also been established. As with Blacks and Latinos in the realm of general intelligence, women have been stereotyped as being less capable than men on mathematical endeavors. This stereotype has also been accompanied by underperformance in comparison to men on standardized mathematical examinations. Quinn and Spencer (2001) conducted a series of studies to examine female performance in relation to the concept of stereotype threat. In the first study, they examined two divergent hypotheses. One of those hypotheses was that stereotype threat interfered with the ability to conceptualize and respond to complex word problems. The other was that there was some truth to the stereotype, and women simply were not as good as men at these tasks. The authors assumed that distinguishing between these two competing hypotheses was accomplished by removing one step from the process of solving a series of complex problems.

One experimental group, comprised of both men and women, was given a test consisting of high-level mathematical word problems. The other group was given the same set of problems, without the verbal context. The mathematical skills required on both sets of questions were identical, but the conceptual steps were removed (Quinn & Spencer, 2001).

There is a considerable literature that documents the increased difficulty of mathematical problem solving when problems are presented in a verbal context (Kintsch

& Greeno, 1985). What makes word problems more difficult than their arithmetic counterparts is the necessity of integrating linguistic and semantic knowledge with arithmetic knowledge, a process that makes heavy demands on various aspects of memory (See Kintsch (1988) for a model of these processes).

In Quinn and Spencer's (2001) first study, it was assumed based on existing literature that stereotype threat was the baseline condition, and did not need to be induced. The problems used were taken from a practice test for the Graduate Mathematics Aptitude Test. As predicted, women underperformed on the word problem variation of the test, and no differences were seen on the numerical only variations.

The second study more directly addressed the priming issue by adding a manipulation of threat. Half of the participants were given the standard instructions (presumed to be high in stereotype threat) and half were given a modified set of instructions that referred to the test as being gender-fair. The authors were also interested in examining the ways in which both men and women attempted to strategize their answers to the word problems. High-scoring math students were asked to complete difficult word problems from the Student Aptitude Test (SAT) Mathematical exam. The participants were asked to reason out loud while solving the problems, and these strategies were then coded by the experimenters. All participants completed the same set of problems (Quinn & Spencer, 2001).

The results indicated that men outperformed women in the high threat condition whereas men and women performed equally well in the low threat condition. When examining the coded problem-solving strategies, one significant difference was noted. Women in the high threat condition were more likely than both women in the low threat

condition and men in either condition to be unable to formulate a problem solving strategy. This supported the hypothesis that stereotype threat impaired women's ability to perform on these tasks. It also supported the hypothesis that an assurance of test-fairness would diffuse stereotype threat effects (Quinn & Spencer, 2001).

The majority of research on stereotype threat addressed issues of an intellectual nature, but support has also been found for the existence of an effect of stereotype on athletic performance. The overarching stereotypes about racial performance in athletic endeavors suggested that Black athletes were more naturally gifted in a strictly physical fashion, whereas White athletes were more gifted in an "athletic intelligence" fashion. In the first of a series of experiments conducted by Stone, Lynch, Sjomeling, and Darley (1999), participants were first asked to complete a pre-test questionnaire that contained a brief self-report measure of anxiety, a pair of self-handicapping questions, and a question on their general perception of bias in standardized testing. A demographic questionnaire was included for three of the four primed conditions that are outlined below. This was followed by a 10-hole trial of golf, and a post-test survey repeating the pre-test anxiety measure, and asking about their perceptions of their performance. Athletic performance was measured by the number of strokes necessary to complete the golf course.

The experiment was a 2 (Black, White) x 4 (primed condition, defined below) design. Black and White participants were asked to read a handout that explained the experiment, the contents of which varied based on the experimental condition primed. The four test conditions were athletic ability, athletic intelligence, race, and a non-primed control. Participants in the athletic ability condition read that the test measured purely physical abilities. Participants in the athletic intelligence condition read that the test

measured intellectual abilities related to athletic performance. Participants in the racially primed and non-primed condition read that the test measured "psychological factors" (p. 1220), with no reference made to either intelligence or ability. Participants in the non-primed condition completed their demographic questionnaire as a post-test item, in order to avoid mentioning race prior to the measure (Stone et al., 1999).

The results indicated that level of anxiety was unrelated to performance. Black participants performed significantly worse when intelligence was primed, as well as when race alone was primed (Stone et al., 1999). It was not made clear why this effect was seen for the race only prime, though it may well have been the case that the confounding stereotype that African Americans are not good golfers complicated the outcome. Although White participants in the athletic intelligence condition performed similarly to the racially primed and no-prime groups, those in the athletic ability-primed condition performed worse, and also tended to discount the validity of the measure in the post-test. There was no significant effect of the racial priming for White participants. The experimenters interpreted this as indicating a possibly unreliable effect of stereotype threat on White participants in the athletic ability-primed condition (Stone et al.). An alternative interpretation may be that, because Whites are typically not stereotyped as being bad golfers, the racial-only prime was not strong enough to produce the same effects it did for Black participants.

A second experiment used only White participants to test the reliability of stereotype threat effects for White participants, and to address a hypothesis regarding the role anxiety played in producing stereotype threat effects. The study was a 2 (identified with athletics, not identified with athletics) x 2 (threat, non-threat) x 2 (internal vs.

external attribution of anxiety) design. All participants were self-described as being athletic in a pre-test used to select participants. They then completed a self-report scale that addressed the importance of athletics to the participants' identities. They were classified as either high or low in this characteristic. They also completed a brief anxiety measure, and a demographic questionnaire. Participants in the external attribution condition were then informed that the laboratory space had been recently renovated, and were asked to complete a brief survey as to whether the new space made them tense or uncomfortable. Those in the internal attribution condition simply proceeded to the threat priming. As with the first experiment, threat was primed by telling the participants (again, all of whom were White) that the task was a measure of physical athletic ability. Those in the non-threat condition were told it was a measure of psychological factors. The task was the same golf task described above. A post-test contained several anxiety measures and the disengagement scale from the pre-test (Stone et al., 1999).

The results replicated the induction of stereotype threat in White participants such that those who were primed to see the measure as representative of athletic ability performed worse than those in the other conditions. The authors also indicated that the threat effects were stronger for those participants whose identification with athletics was stronger. These results were further support for the generalizability of stereotype threat effects to non-academic situations (Stone et al., 1999).

There was also an indication that the availability of an external attribution decreased the impact of stereotype threat. Although participants in this group did not report significantly lower subjective experience of anxiety, their performance was not as impaired as those in the high threat, internal attribution group. The authors interpreted

18

this as a distraction from stereotype threat. They suggested that the participants were equally aware of the threat, but were focused on attending to the environment due to the cues offered by the researchers, that served to distance them from the anxiety (Stone et al., 1999).

In a pair of experiments testing the limits of stereotype threat on influencing performance, Steele and Brown (1999) tested the hypothesis that stereotype threat does not require a history of stigmatization or internalized feelings of inferiority. The experimenters were also interested in the hypothesis that strong identification with the domain of interest increased the impact of stereotype threat.

In the first study, the authors asked White participants to complete a very challenging mathematical exam. All participants had previously received high scores on the Arithmetic section of the SATs, and were at least neutral when asked about their opinions of the importance of mathematics skill. Prior to taking the exam, those in the threat condition were presented with a series of newspaper articles emphasizing the skilled performance of Asians on mathematic tasks, and were also told that the study wanted to examine the differences between Asians and Whites. Those in the control condition did not see these articles, and were told that the study was simply a measure of mathematical ability. A computer tracked the time participants spent on each question, and recorded their answers. Following the test, participants completed an anxiety measure and a questionnaire about their effort, sense of pressure, confidence, and estimation of performance (Steele & Brown, 1999).

The results of an ANCOVA of the performance measure indicated that the participants solved fewer items in the threat condition than in the control condition, in

spite of putting forth more effort. The authors interpreted this as support for the

hypothesis that a stereotype threat could be induced in a population that was not likely to

have internalized a negative self-stereotype. Their rationale for this hypothesis was that

stereotype threat was not the result of chronic exposure to and internalization of negative

stereotypes. They suggested that stereotype threat was actually the result of a drive for

feelings of competence (Steele & Brown, 1999).

One issue with this experiment was the fact that there exists a stereotype

unfavorably comparing Whites to Asians on mathematical ability. Although it was likely

true that Whites were less prone to chronic concern over this stereotype than the

comparable ones concerning Blacks, Latinos, women, and the poor, that certainly does

not indicate that the stereotype was not known to the participants. The authors addressed

this in their general discussion, indicating that perhaps the amount of exposure to

negative stereotypes somehow affects the impact they have on the stereotyped population

(Steele & Brown, 1999).

The second study served to replicate the first, but also to elaborate upon it by

addressing the authors' hypothesis that identification with the domain of interest was

necessary in order for stereotype threat to affect performance. The authors used a 2 (high

or moderate math identification) x 2 (threat or non-threat) design. Participants completed

a pre-test 3 weeks prior to the test date that addressed their degree of self-identification

with mathematical ability. On the test date, participants in the threat condition were

presented with a written description of the study's purpose, which described the high

level of mathematical abilities possessed by Asian students, and the experimenters'

interest in comparing scores between White and Asian students. Those in the control

condition read a description that indicated the test was designed to understand mathematical thought, and that the results would be compared to students across the country. The post-test contained the same items as in the previous study (Steele & Brown, 1999).

Results indicated that only those with high math identification scored lower in the threat condition. The authors again saw this as support for their hypothesis that chronic stereotype exposure was not necessary to induce stereotype threat. They also suggested that identification with the domain is important in the mediation of stereotype threat (Steele & Brown, 1999). Again, the question was raised whether or not a stereotype disparaging White mathematical abilities in relation to Asian's would be considered common knowledge. The effects of domain identification were similar to earlier observations by Steele and Aronson (1995).

In summary, the body of literature produced in the first wave of stereotype threat research provided substantial evidence that it was a true psychological phenomenon. A great many other studies were conducted. This selection was chosen because it provided a cross-section of the variations on stereotype threat that had been examined. Stereotype threat effects were demonstrated for African Americans, Latin Americans, women, and low SES individuals. The second challenge that faced researchers was the exploration of how stereotype threat altered performance. A number of plausible explanations were addressed in the literature.

Proposed Mechanisms Underlying Stereotype Threat

One of the first proposed explanations for stereotype threat was motivational. The mechanism proposed was a withdrawal of effort based on decreased belief in the

possibility of success. Steele and Aronson (1995) partially addressed this by having the participants respond to a series of Likert-type scales after the task that addressed participants' perceptions of the amount of effort put forth. These scales showed no significant effects. Computer tracking of time spent per item showed Black participants did reread more questions and spent more time on each item. Although this could be interpreted as an increase in motivation for Black participants, since motivation itself is not clearly defined. It was clear however, that Black participants did not decrease the effort expended under conditions of stereotype threat.

McFarland, Lev-Arey, and Ziegert (2003) sought to investigate the question of motivation and stereotype threat more fully. They were also interested in examining the operation of stereotype threat in a real-world setting, such as a job interview, where a tangible reward was offered for success. Finally, they were interested in exploring how threat effects generalized from one test to another in multi-test situations (McFarland et al., 2003).

The experimenters asked participants to complete measures assessing two domains: personality and intelligence. They also informed participants that those who scored in the top 15% across both measures would receive a monetary prize that was meant to simulate being hired for a job. Because the experiment was designed to simulate a job interview, good performance on the personality measure was assumed to be reflected in high scores on elements that had previously been correlated with good job performance (extroversion) and low scores on those that were not (neuroticism). Participants in the threat condition were asked to complete a demographic questionnaire and racial and domain (intelligence) identification measures before the test. They were

informed that the experimental tasks were correlated with intelligence and ability. Racial and domain identity were assessed in a post-test for all participants. The experiment was a 2 (Black, White) x 2 (intelligence test first, personality test first) x 2 (threat, non-threat) design (McFarland et al., 2003).

The results indicated that, contrary to expectation, there was no race by threat effect. When the authors controlled for identification with race and the domain of intelligence however, an effect of stereotype threat was found. Those participants of both races who were low in domain identification and high in racial identification did perform worse under threat conditions. This was truer for Black participants than for White. Furthermore, those Black participants high in race identification in the pre-test performed better than those who were low in race identification. The opposite trend was true for the post-test, where Black participants who performed well were actually lower in racial identification. This indicated that these individuals experienced a process of dis-identification. No similar effect was noted for White participants (McFarland et al., 2003). This process of dis-identification was similar to that noted in Study 3 by Steele and Aronson (1995).

There were a number of other significant effects noted. Black participants scored significantly worse than White participants across conditions. White participants scored higher when strongly identified with the domain of intelligence. All participants strongly identified with race scored worse in the threat condition. Finally, all participants performed better on the personality measure when in the threat condition (McFarland et al., 2003).

The authors interpreted this pattern of results to indicate that the expected stereotype threat effects were somewhat altered by the motivation of the monetary reward. They also suggested that evidence had been found for individual differences that may modify the way in which stereotype threat had its effect. They suggested that racial identity buffered the impact of stereotype threat, but that post-test only measures may fail to replicate this, due to the dis-identification process (McFarland et al., 2003). This left several questions unanswered. What is motivation? How does it actually affect stereotype threat? How does identification, with race or domain, alter outcome measures? A more thorough understanding of the mechanisms underlying stereotype threat may resolve some of these questions.

Stangor, Carr, and Kiang (1998) investigated another hypothesis regarding the mechanism underlying stereotype threat that addressed the participants' level of confidence. They suggested that stereotype threat interfered with performance by undermining the expectation of success, which in turn affected actual performance. It was suggested that expectancy of success may have served as a buffer against threat effects, and that low expectations may have established a self-fulfilling prophecy.

The first of the experiments was a 2 (high expectation, low expectation) x 2 (threat, non-threat) design. A group of all-female participants were presented with a pair of tasks. The first of the tasks was a word search task that was presented as gender-fair. Before beginning this task, participants were asked for an estimate of how well they expected to perform. They then completed the task, and were given feedback about their performance. Feedback was manipulated to lead the participants to believe their performance was either average or excellent, in order to establish an expectation of

success on the next task. Participants then provided a self-rating of their task performance for the first task to reinforce the manipulation of expectation. They were then informed that the next task, a spatial-abilities task, was correlated with intelligence. Those in the threat condition were told that men outperformed women on this task, whereas the non-threat participants were told it was gender-fair. Participants were asked to estimate their performance, and to indicate their expected affective reactions to the task (Stangor et al., 1998).

The analysis showed that there was an interaction between threat condition and expectation. Performance feedback on the first task influenced performance expectations for the second task, but only in the non-threat condition. Those participants in the threat condition did not benefit from improved expectation of performance due to positive feedback. The authors saw this as support for the hypothesis that stereotype threat undermined expectations of future success, even when evidence of past success was fresh (Stangor et al., 1998).

The second study was quite similar, but it used what the authors suggested was a more subtle stereotype manipulation. Male and female participants were told that they would participate in four different word-finding tasks. They were led to believe that they would be working alone, in groups, or both, though in actuality all participants were assigned to groups. Threat was manipulated by demographic composition of assigned groups. The participants then completed a baseline measure word finding task, which served to provide feedback in the same fashion as in the prior study. Participants were then asked their preference to work alone or in groups. They were told that those who worked in groups would have their performance reported to the other group members. A

demographic sheet containing information about the group the participants were possibly

joining was provided, in order to manipulate similarity and threat. Participants in the

similar condition were assigned to a group containing men and women with similar

majors to the participant. Those in the dissimilar condition were assigned to a group that

was composed of individuals of the opposite gender and with different majors from the

participant. The expectation was that those in the dissimilar condition would feel more

threatened. A questionnaire about performance expectations and expected affective

reactions was provided (Stangor et al., 1998).

Analysis of the results indicated that, although positive feedback improved

expectations in general, those who expected to work in dissimilar groups, and were thus

presumed to be operating under threat conditions, did not experience increased

expectations. The authors interpreted this as further evidence that stereotype threat served

to undermine task performance expectations (Stangor et al., 1998). Although the data did

seem to support this hypothesis, the question of mechanism was not addressed, nor was

there a clear description of how this expectancy would alter behavior. Furthermore, it was

not demonstrated that decreased expectation for success was causal in the decreased

performance observed for those operating under stereotype threat. In addition to the

concerns mentioned above, the authors did not actually include a performance measure.

A series of follow-up experiments investigated the role that expectancy played in

mediating performance under conditions of stereotype threat. In the first experiment,

female participants were assigned to one of three levels of a threat condition. Those in the

high threat condition were told that men outperformed women on logical-mathematical

problems, whereas those in the low threat condition were told that women outperformed

men on these tasks. Those in the neutral condition were told that men and women performed equally. Before performing a difficult math task, the participants were asked to provide an estimate of their expected performance. Data on the participants' identification with mathematical ability, identification with the female gender, and locus of control was gathered in a pre-test. The experimental hypothesis was that those in the high threat condition would have lower expectations, and would also have lower performance scores (Cadinu et al., 2003).

The results indicated that there was an interaction such that women highly identified with math showed a decrease in expectation when they were operating under stereotype threat. This was not true of women with low math identification, who performed consistently across threat levels. This same effect was observed in the performance data. In a test for mediation, the authors found that expectancy was a partial mediator for performance, which supported their primary hypothesis. They also found that domain identification moderated the effects of stereotype threat (Cadinu et al., 2003).

The second experiment sought to replicate these results using a different population. A group of Black Americans living on a NATO base in Italy participated in a 2 (race prime, nationality prime) x 2 (threat, non-threat) experiment. The authors proposed that these participants were not only members of a Black minority group, but were also members of a respected American majority group on this base. Data from White participants was collected but not analyzed to avoid overtly priming race by including only Black participants. The participants first completed an identification scale that addressed race for those participants assigned to the Black level of the in-group condition, and nationality for those assigned to the American level. The participants were

then told that their in-group (Blacks or Americans) performed better or worse than the out-group (Whites or Italians) on tasks of verbal abilities. Participants then rated their performance expectations before completing a test of verbal abilities taken from the GRE (Cadinu et al., 2003).

A main effect of expectation was found such that both the Black-primed and American-primed groups performed significantly better under non-threat conditions. There was no interaction effect. This was true for both expectation and performance data. The mediation test was repeated, and the partial mediation of performance by expectancy was again supported, but only for participants in the Black level of the in-group condition. The authors suggested this may have been the result of different stereotype threat mechanisms for minority versus majority group members. They suggested the self-fulfilling prophecy for minority group members, but were uncertain regarding what mechanism may have been in effect for the majority (Cadinu et al., 2003). Again however, this did not address a specific, cognitive mechanism.

Another possible mechanism postulated for stereotype threat was the interference of anxiety on performance. Yerkes-Dodson, a well-known psychological law, states that arousal and performance have a curvilinear relationship (Teigen, 1994). Regarding stereotype threat, the pressure to prove oneself and one's in-group might be expected to raise the level of anxiety in an individual. If the level of anxiety rises past the optimal point, performance would begin to decline. There was mixed support for anxiety influencing stereotype threat in the work of Steele and Aronson (1995). Specifically, they noted that self-reported anxiety levels had no impact in Experiment 2, but that higher anxiety was correlated with decreased performance in Experiment 4. Other research has

indicated no connection between anxiety and performance under conditions of stereotype

threat (Stone et al., 1999). In an effort to clarify this issue, Blascovich, Spencer, Quinn,

and Steele (2001) investigated the effects of stereotype threat on blood pressure. Blood

pressure is a commonly accepted physiological measure of emotional reactivity.

Black and White participants in the high threat condition viewed a video in which

a Caucasian experimenter briefly discussed the issue of bias in standardized testing. They

were told they would be taking part in normative research for a new intelligence test.

Those in the low threat condition viewed a similar message from an African American

experimenter who informed them that they would be taking part in normative research for

the development of a culturally unbiased test, and that pilot studies had shown this test to

be culturally-balanced. They were then asked to complete two trials of a word-association

task. Blood pressure was monitored throughout (Blascovich et al.).

The results supported the hypothesis that stereotype threat would increase blood

pressure for African Americans under conditions of stereotype threat. This effect

persisted through a rest period, and carried over into the repeated task trial. Performance

scores were also lower for Blacks in the threat condition. This was seen to support the

role of anxiety in stereotype threat (Blascovich et al.). The issue that was still not

addressed was how anxiety actually interfered with performance on this cognitive task.

A study by Mayer and Hanges (2003) sought to investigate four different factors

thought to influence stereotype threat effects. They identified cognitive interference,

anxiety, performance expectations, and evaluative apprehension as the variables of

interest. The goal was to examine the combined impact of these factors, in case individual

effects faded when viewed in isolation. The researchers were also interested in examining

29

a non-verbal task that was still correlated with intelligence, but that may be more believably presented as bias-free. Finally, the authors suggested the mixed results of prior research may have been a reflection of two distinct types of stereotype threat. The first of these was a specific, focused threat effect that was generated by some specific element of the test or setting to which the participants responded (i.e. a racial or threat prime). The second type was a generalized level of stereotype threat that cuts across time and setting. This was defined as a characteristic inherent in the participant and reflected a constant awareness of stereotypes.

The experiment was a 2 (Black, White) x 2 (threat, non-threat) design. In order to make the experiment ecologically valid, the researchers did not prime race directly. A pre-test given 1 month before the task included a racial identification scale. On the day of the experiment, participants were informed that they would be taking a test that was commonly used as part of a job interview. The test was a pattern-identification measure that was presented as being correlated with intelligence (threat condition) or perception (non-threat). The participants then completed further pre-tests including anxiety, cognitive interference, self-efficacy, and evaluation apprehension scales. Following the experimental measure, participants again completed cognitive interference measures as well as measures of perception of both general and specific stereotype threat. Finally, a brief manipulation check was conducted (Mayer & Hanges, 2003).

The first hypothesis tested was that African Americans would report higher levels of general stereotype threat in their self-report post-tests than would Caucasians. This expected result was supported (Mayer & Hanges, 2003).

The second hypothesis was that Black participants in the threat condition who identified strongly with race would experience high levels of specific stereotype threat. This hypothesis was also supported (Mayer & Hanges, 2003). This was consistent with results discussed above (McFarland et al., 2003).

The third hypothesis was that Black participants in the threat condition who identified strongly with race would experience decreased performance in comparison to Whites. This interaction was not found to be significant (Mayer & Hanges, 2003). This replicated the results of an earlier study, in which a simulated job interview failed to elicit the standard stereotype threat effects (McFarland et al., 2003). The researchers conducted follow-up analyses using their manipulation check that revealed that both Black and White participants in the threat condition did not believe that the test was actually diagnostic of intelligence. Black participants were more likely to disbelieve this manipulation. Thus, it is possible that the interaction did not occur because the manipulation was ineffective (Mayers & Hanges).

The fourth hypothesis was that self-reported levels of both specific and general stereotype threat would be negatively correlated with task performance. This hypothesis was supported. The relationship was stronger for general stereotype threat than for specific. This indicated that the theoretical separation into two types of stereotype threat may have some validity (Mayer & Hanges, 2003).

The final hypothesis was that the relationship between threat and performance would be mediated by anxiety, cognitive interference, self-efficacy, and evaluation apprehension. Structural equation modeling was used to test these hypotheses for both general and specific stereotype threat. For specific stereotype threat, these hypotheses

31

were not supported. The only construct other than test score that was directly connected to stereotype threat was evaluation apprehension, which was not itself correlated to test score. For general stereotype threat, a nearly identical model was produced, save that there was no correlation between threat and evaluation apprehension. The combination of these two models added the expected correlation between general and specific threat, but was otherwise not significant. The authors suggested this may have been due to the relative weakness of the stereotype threat effect in comparison to previous research (Mayer & Hanges, 2003). Another potential explanation is that some researchers (Barlow, 2004; Borkovec, Ray, & Stöber, 1998) suggest that nonverbal tasks such as were used in this study may not be as strongly affected by anxiety as are verbal tasks.

Seibt and Förster (2004) expanded the concept of stereotype threat effects by investigating situations in which these effects may actually enhance performance. Their research focused on the concept of stereotypes inducing regulatory foci. These were conceptualized as a sort of mental set that was designed to enhance performance on certain types of tasks. Their hypothesis revolved around the idea of two opposing goals, approach and avoidance, dictating the way in which the participant responded to various stimuli. The approach focus served to move the participant towards some desired goal, whereas the avoidance focus served to move them away from an undesirable one. In the case of a stereotype threat situation, the authors hypothesized that a negative stereotype that the participant identified as potentially self-relevant created an avoidance reaction. This idea also allowed for positive stereotypes, which created an approach reaction. They further suggested that an avoidance reaction encouraged caution and accuracy, whereas

an approach reaction fostered speed and creativity. Prior research, which is beyond the scope of the current work, was cited to support these assertions.

Stereotype threat was defined by the authors to be a stimulus that would induce an avoidance focus. If this were the case, the authors hypothesized that an individual operating under stereotype threat conditions would perform poorly on creative tasks, but may excel on analytic tasks, whose goals could be met using time-tested heuristics with minimal probability of failure. The authors tested this hypothesis over the course of five separate studies (Seibt & Förster, 2004).

In the first study, the researchers asked their participants to read a text containing events from a person's life, each of which exemplified either an approach or avoidance method of achieving the protagonist's goals. The experimental task was presented as representative of verbal intelligence in order to encourage identification with the domain of the task. All participants were non-psychology students. Participants in the negative stereotype threat condition were told that psychology students did comparatively well on the experimental task. Those in the positive stereotype threat condition were told that other majors outperformed psychology students. Participants were then asked to write down their majors, and complete a brief assessment of their expectations of performance. They then read a brief text containing grammatical errors, which they were asked to correct. The sentences in the text were a mix of approach and avoidance statements involving the protagonist. Participants then completed a surprise recall task (Seibt & Förster, 2004).

The results supported the hypothesis that those in the positive stereotype condition would better recall information of an approach nature, whereas those primed

with a negative stereotype would better recall avoidance information. This provided general support for the authors' initial hypothesis (Seibt & Förster, 2004). The indication was that there may have been multiple pathways by which stereotype threat operated.

In the second study, the authors used a well-researched stereotype to examine these pathways. Participants were asked to complete a word-selection task, in which they chose one word that did not belong in a list of five. Participants in the stereotype condition were primed to believe that the experiment was investigating gender differences, whereas participants in the control condition were led to believe it was simply a measure of individual performance. A control group was not informed of the purpose of the study. Motivation and expectation for success were then assessed, and demographic information was gathered. Participants in the stereotype condition were informed they would be asked to note age and gender on each page of the test. All participants were told that speed and accuracy were equally important (Seibt & Förster, 2004).

Analysis of the data revealed that priming a negative stereotype produced slower, more accurate performance in comparison to the control group. Priming a positive stereotype led to faster, less accurate performance in comparison to the control group. These results were less reliable for the positive stereotype (Seibt & Förster, 2004). This is consistent with Steele and Aronson's (1995) work, in which they suggested that positive stereotype threat effects were more difficult to induce.

The third study attempted to replicate and clarify the results of the second study. A task that had been predetermined to be sensitive to changes in regulatory foci was used, and the investigators ensured that both the positive and negative valence were equally

powerful. The experiment was presented as an investigation of imagination. Stereotype was induced in the same way as in the first study. A control condition was included, with no stereotype threat prime included. The participants were asked to provide demographic information, including major, and to indicate their expectations for success (Seibt & Förster, 2004).

Participants in the negative stereotype condition had increased accuracy relative to the control and positive stereotype groups. Speed was higher for the positive stereotype group (Seibt & Förster, 2004). Again, support was found for the existence of differential effects based on the cognitive focus of the participants.

Study 4 sought to examine the impact of stereotype threat on analytic versus creative tasks. The hypothesis was that positive stereotypes would improve creativity and impair analytic thinking, whereas negative stereotypes would have the opposite effect. During prior research, analytic elements from the GRE were found to benefit from an avoidance prime, whereas a creative task (generating novel uses for a brick) was found to benefit from an approach prime. Participants were provided with the same set of instructions used in Experiment 1 in order to induce the stereotypes, and were again asked to rate their expectations of performance. All participants were presented with the analytic task, followed by the creative task (Seibt & Förster, 2004).

As predicted, the nature of the required task determined the influence of the stereotype. For the brick-usage task, a negative stereotype impaired performance whereas a positive stereotype improved it. The opposite pattern was noted for analytic tasks. This was the first time a stereotype threat model was able to accurately predict a complex

pattern of results such as this (Seibt & Förster, 2004). This again supported the differential effects hypothesis.

The final study of the series was a replication of Study 4, with the substitution of an alternative creative task. Participants were asked to provide "goodness-of-fit" ratings for a number of items and relevant categories. The authors suggested that creativity entailed the activation of several networks to produce connections between the objects and the categories. Although uncreative responses would involve seeing a solid fit only for objects (e.g., brick) that obviously fit a category (e.g., building material), a creative response would involve connecting objects (e.g., brick) with less obvious, but still potentially applicable categories (e.g., paperweight). The stereotype threat and pre-test procedures from the first experiment were repeated. Specific, regulatory-focus relevant questions were included that assessed tendency toward avoidance and approach. Participants completed the creative task, followed by the analytic task (Seibt & Förster, 2004).

The results again supported the hypothesis. Participants in the positive stereotype condition performed better on the creative task, whereas those in the negative stereotype condition performed better on the analytic task (Seibt & Förster, 2004). This provided further support for the primary hypothesis. Furthermore, a pattern was emerging that indicated that participants' mental model of the task influenced the effect that stereotype threat had upon performance.

One further potential explanation, explored by Schmader and Johns (2003), examined a cognitively oriented explanation for the reduction in performance. They suggested an explanation of stereotype threat effects based on the idea that stereotype

threat not only acts as a stressor, but also reduces the capacity of working memory. Baddeley's (1986) classic definition of working memory was used. This consisted of a mechanism comprised of two parts (the phonological loop and the visuo-spatial sketchpad) controlled by a central executive. The purpose of these mechanisms was to allow for the rehearsal and manipulation of a small number of items until this information reached long-term storage. The central executive coordinated this mechanism, and dictated the focus of attention.

Schmader and Johns' (2003) first experiment was designed to test the hypothesis that women in a stereotype threat condition would under-perform on a working memory task in comparison to men and to women in a control condition. The working memory task was an operation-span task that required participants to simultaneously solve a simple mathematical problem while memorizing a word presented concurrently with the problem. During the introduction to the experiment, participants in the non-threat condition were told the experimental measure they would be completing was a test of working memory capacity. Those in the threat condition were told the task was a measure of mathematical performance under distraction conditions, and that men tended to outperform women. An anxiety scale and a self-report of test difficulty and personal effort were presented as a post-test. The outcome measure was called the "absolute span score" (p. 443). This was a summation of the words recalled in those trials in which the participant obtained a perfect score. That is, when the participant forgot one word from a trial, no words from that trial were counted towards the total.

The results indicated that women in the threat condition recalled fewer words than women in the control condition, and men in either condition. Performance on the

mathematical tasks was also analyzed, and there was no difference in performance across gender or threat conditions (Schmader & Johns, 2003). This was most likely due to a ceiling effect resulting from the use of a simplistic task. No significant effects were noted regarding anxiety scores, though women in the threat condition did rate the test as being more difficult than did men or women in the control condition. The authors saw these results as support for their hypothesis that stereotype threat would lead to a measurable decline in working memory resources (Schmader & Johns).

The second study sought to replicate the above effects with a different population and different cognitive domain. Furthermore, the authors sought to rule out the possibility that their blunt priming of stereotype may have caused the results. Thus, they used a more subtle stereotype prime to compare Latino and White performance on the same working memory task used above. The prime for the threat condition consisted of asking participants to provide their ethnicity as part of a pre-test questionnaire. Although the task was described as being a working memory measure for all conditions, those in the threat condition were also told it correlated strongly with intelligence (Schmader & Johns, 2003).

The results of this second study were largely consistent with those seen in the prior study. There was an interaction between race and condition, such that Latinos in the threat condition recalled fewer words. There was also an unexpected gender effect such that women in the threat condition recalled fewer words. There was no effect on mathematical performance due to race or threat condition, but there was an effect due to gender such that women underperformed relative to men. This was significantly amplified for the threat condition. Finally, Latino participants in the threat condition

38

reported significantly more anxiety, but this was not correlated with performance on either the memory task or the mathematics task. The authors saw these results as a successful replication of the earlier experiment, and an expansion of the threat effect to a more subtle manipulation (Schmader & Johns, 2003).

The final study in the series was designed to determine whether working memory capacity mediated the effects of stereotype threat on performance on an academic test. To test for mediation, female participants were asked to complete a working memory measure and a standardized arithmetic test. The working memory task was modified by replacing the arithmetic task with a vowel-counting task. This was done to reduce the possibility of participants tending to focus on the arithmetic portion of the task, given its possible confusion with the threatened domain of intelligence. The final modification involved changing the threat manipulation from explicitly drawing a connection between the working memory task and the stereotyped domain, to an even more subtle, environmental manipulation. Specifically, women in the threat condition were the sole female present in a group of male confederates, and were told they would be completing a math test following their working memory task. Those in the control condition were run with other female participants, and were told they would be completing a problem-solving task. The experimental hypothesis was that any performance decrements for the stereotype threat condition would be accounted for by the scores on the working memory task (Schmader & Johns, 2003).

As with the previous experiments, women in the threat condition recalled fewer words. There were no differences in terms of accuracy of vowel counting or performance on the subsequent math test. Anxiety levels also did not differ. With regards to the

mediation tests, stereotype threat was not related to math task performance when the working memory performance was taken into account. The authors saw this as support for their experimental hypothesis that working memory capacity was a mediator of threat effects. Furthermore, they took the overall outcome of this series of studies as an indication that the effects of stereotype threat were produced, at least in part, through a reduction of working memory capacity. The authors stated the importance of examining the role that affective arousal played in these effects, and of determining where the working memory resources were consumed (Schmader & Johns, 2003).

<div align="center">Anxiety and Stereotype Threat</div>

Many researchers have postulated some form of anxiety as a causal mechanism of stereotype threat. Others have suggested a more capacity-oriented cognitive explanation. Schmader and Johns (2003) suggested both anxiety and working memory capacity as two possible mechanisms. But this separation of anxiety and cognitive factors into distinct mechanisms may have been artificial, because it is possible that anxiety processes interfere directly with cognitive factors such as working memory.

Barlow's (2004) comprehensive, empirically-based theory of anxiety and its disorders emphasized the cognitive elements of anxiety. He identified apprehension and worry as the primary characteristics of anxiety. Furthermore, he pointed out the research indicated that distraction and worry played a greater role than autonomic arousal in poor test performance. The research described in Barlow's book indicated that situational cues associated with negative affect led to a shift towards an internal focus of attention on somatic sensations and self-evaluative aspects of the situation. The author suggested this evaluative self-focus could not be separated from affective and cognitive descriptions of

anxiety. Furthermore, Barlow (2004) indicated individuals with high levels of anxiety tended to respond with heightened avoidance of anxiety-relevant cues. This response bias led to a narrowing of attentional focus that decreased the number of contextual cues used to assess a given situation, and changed the interpretation of those cues that were used. This has clear implications for research into stereotype threat.

Worry was also explored in Barlow's (2004) work. It was described as a potentially adaptive mechanism that allowed the individual to plan for difficult situations in the future, serving as a pre-emptive coping response. This strategy can become problematic when the worry moves past the point of allowing the individual to plan, and actually begins to interfere with concentration and attention. At this point, worry becomes a maladaptive cognitive pattern that serves to protect the individual from addressing the unpleasant affective core of their anxiety. The data indicated the worrying itself was a verbal/linguistic phenomenon that interfered with verbal tasks, but not with more visually oriented tasks. The interference occurred through a combination of attentional shifts and hypervigilance for threatening elements. Barlow also explored the intrusion of irrelevant material into cognitive processing during anxiety.

A literature review by Borkovec, et al. (1998) investigated worry specifically as a cognitive phenomenon. Their findings supported Barlow's (2004) assertion that the cognitive dimension of anxiety had more influence over performance than did the affective dimension. This could be an explanation for the mixed findings regarding self-reports of anxiety and stereotype threat effects. For example, in Steele and Aronson's (1995) work, self-reports of anxiety were elevated for those operating under stereotype threat in the fourth experiment, but not in the second. Both experiments resulted in

stereotype threat effects on performance. Although Steele and Aronson suggested this may have indicated multiple mechanisms, an alternative interpretation may be that the measures used to assess anxiety typically evince responses for the affective elements, but not for the potentially more powerful cognitive intrusions. Thus, the lack of a solid operational definition of anxiety may have corrupted the measures. Borkovec, et al. (1998) also suggested that worry was primarily a verbal/linguistic phenomenon that affected performance through intrusions. This process resulted in a decrease in attention to details of the given task. Consistent with Barlow's work, those with tendencies to worry had more elaborate networks of associations built around the topic of their worries. These networks were thought to contribute to the observed delay in processing speeds for those high in worry. Furthermore, worriers displayed fewer inhibitory processes and less executive control of their cognitive activities than non-worriers. This is consistent with Steele and Aronson's observations that those operating under stereotype threat tended to take longer to complete the tasks.

An important observation from this work is that much of the research cited involved the induction of these performance effects in a non-clinical population. The cognitive impairments were often noted in individuals who did not meet diagnostic criteria for anxiety disorders (Borkovec et al., 1998). This suggests continuity in the anxiety process between clinical and non-clinical populations.

Recently, Watkins (2004) examined the mechanisms underlying rumination and worry. He had participants complete a series of self-report measures of worry, rumination, anxiety, and intrusive thoughts. Correlational and factor analyses were conducted on these data. The results indicated that increased rumination was related to

the personal relevance of the material, and a need to understand the situation through analysis focusing on its causes and meaning. Increased worry was associated with a negative response to worry-related intrusive thoughts and an attempt to displace these thoughts with other negative thoughts.

The recent research on the cognitive characteristics of anxiety holds a number of suggestions that are relevant to understanding stereotype threat. Watkins' (2004) research suggested that these effects were demonstrable in both clinical and non-clinical populations, indicating that these anxiety effects were a part of normal processes.

Watkins (2004) observed that personally important material tended to inspire worry and rumination. This observation suggested that stereotype threat, a highly personal phenomenon for those with strong domain identification, would likely produce these worry and ruminative effects.

Finally, observations from several sources suggested a process by which anxious cognitive patterns interfered with verbal task performance through a combination of hypervigilance, distraction, cognitive intrusions, and decreased executive control (Barlow, 2004; Borkovec et al., 1998). This is consistent with Steele and Aronson's (1995) assertion that stereotype threat involved an avoidance mechanism. By focusing on other aspects of the material or situation, the participants were able to distance themselves from the primary concern of confirming a negative stereotype.

A Model of Text Processing

Stereotype threat has been established as a solid, replicable psychological phenomenon. Although the mechanisms by which it operates have not yet been made clear, many authors have focused on affective mechanisms such as anxiety and

evaluation apprehension. A few theorists have also investigated working memory mechanisms, with a predominant focus on working memory capacity reductions.

A potential problem in the research however, has been that simple experimental tasks have been used to test the proposed causal mechanisms behind stereotype threat. Although these have served to produce stereotype threat effects, it is questionable whether these effects would generalize to a more complex, real-world situation. For example, the capacity for operation-span tasks (Schmader & Johns, 2003) to accurately represent a more real-world task such as discourse comprehension is questionable.

A number of the experiments reviewed above have used a high-level reading task. However, they did not have a detailed task analysis to aid in interpreting the results. Thus, at the experimental end of the work, simple tasks were used. At the interpretive end, presumptions were made regarding complex processes. What is needed is a well-developed model of comprehension applied to a real-world task (text processing in this case) that has been shown to be applicable across a variety of situations (text types). The empirical support for Kintsch's (1988, 1998) CI model of comprehension is robust across a variety of texts, and has been found applicable in a variety of tasks beyond text comprehension. This comprehension model is briefly outlined below. This is a greatly simplified presentation of a complex model of comprehension, which is detailed much more thoroughly in Kintsch's own work.

Reading and comprehending a text is a complex cognitive task. It involves many ongoing processes, including but not limited to: selection and activation of appropriate knowledge of language, semantics, contextually appropriate word meanings, proposition construction, and knowledge from semantic memory. For a skilled reader, these cognitive

processes are strategic, well-learned, highly practiced, and therefore operate relatively automatically when the text is about a well-known topic. Under such circumstances, these processes result in an integrated cognitive structure that combines information from the text with information from a reader's semantic network (Kintsch, 1988).

The CI model proposes that text is processed in cycles, because the limitations of working memory would not allow anything but the shortest of texts to be processed at once. Each cycle processes roughly a single sentence. In each cycle there is a construction and integration phase. During the construction phase, strategic construction rules generate propositions that represent the meaning of the words in the text. When all possible propositions from one cycle have been produced, they comprise a coherence matrix. Each element in this matrix then undergoes an elaboration process in which close associations of text information from the reader's long-term memory are incorporated into the coherence matrix. Furthermore, demands specific to the readers' conceptualization of the task also activate long-term memory structures that are associated with those demands, and that may be included in the coherence matrix. Connections are made between elements of this matrix, based on connection strengths inherent in the text or in the relevant long-term memory structures. The result is a large, all-inclusive coherence matrix containing many potential interpretations of the text (Kintsch, 1988).

The integration phase is used to cull the irrelevant, poorly connected elements from this matrix. This follows a connectionist constraint satisfaction mechanism in which activation builds on propositions that are strongly interconnected, and deactivates those that are weakly connected. Activation is passed through the network until it stabilizes into a set of the most highly active elements. The most strongly connected of these elements

are held over in working memory, to be combined with information taken in during the next processing cycle. This carry-over maintains continuity between earlier text and that contained in later cycles, and allows for the integration of a large text despite the constraints of a limited working memory capacity (Kintsch, 1988).

These processes operate at all levels of text representation, which produces a multi-tiered memory representation of the text. It was proposed that at least three levels of representation are generated for each text: a surface level, a textbase level, and a situation model. The surface level produces a representation of the exact phrasing and structure of the discourse. The textbase produces a representation of the meaning of the text. The situation model level of representation generates a mental representation of the text combined with prior knowledge and task demands. It merges the text proper with what the text was actually about (Kintsch, 1988). This level is particularly important in terms of exploring stereotype threat effects because the complex verbal tasks often used in demonstrations of stereotype threat measured performance in terms of knowledge of what the text was about. They assumed that good performance involved a high-level understanding of the text, and adopted performance measures that reflected that assumption. The verbal section of the GRE test asks questions involving knowledge of the situation model level of representation. Furthermore, because the situation model activates and incorporates knowledge from long-term memory, this level of representation should be one that is most susceptible to intrusions and misunderstandings.

This is a highly complex task. There are numerous points at which this task can be influenced by a reader's knowledge and goals. For example, if the reader interprets a text with a situation model influenced by stereotype threat, the reader will likely place a

greater emphasis on their own prior knowledge of stereotypes and associated knowledge than the text was intended to produce. Furthermore, many of the comprehension tasks occur in working memory. As Schmader and Johns (2003) suggested, working memory capacity may be reduced by stereotype threat.

<div align="center">Memory and Text Processing</div>

A recent theory regarding working memory proposed by Ericsson and Kintsch (1995) is of some relevance to potential explanations of stereotype threat effects and how it may operate during text comprehension. It is one of several theoretical and research advances that currently expand our understanding of working memory beyond the usual textbook definitions derived from Baddeley (1986) and his colleagues. Briefly, Baddeley proposed that working memory is a multi-component system comprised of two highly specialized subsystems (phonological loop and the visuo-spatial sketchpad) and a central executive whose function was to temporarily store current information from the environment in the service of some more complicated cognitive activity. The phonological loop and the visuo-spatial sketchpad were thought to function with restricted capacities similar to original short term memory conceptualizations. Schmader and Johns' (2003) work discussed earlier basically adheres to such a conceptualization.

Ericsson and Kintsch (1995) noted that this common concept of working memory is inadequate to explain text comprehension where the demands on working memory clearly exceed the traditional capacity limits. Yet skilled readers seem to possess very good memories of the texts they read if those texts are about well-known topics. Thus, Ericsson and Kintsch have developed a theory of long-term working memory (LTWM) to

<div align="center">47</div>

explain how the traditional working memory constraints can be by-passed in text comprehension.

Long-term working memory was defined as knowledge structures within long-term memory that became linked to material currently active in working memory. For example, as a skilled reader progresses through a text, different propositions from the text are *active* during each cycle, as described previously. These active nodes in working memory may act as retrieval links to other currently *inactive* information in long-term memory. If that information in long-term memory is well-learned, with strong connections between its elements, the information currently *active* in working memory can act as a retrieval cue for a large amount of information in long term memory with a single retrieval function. Such a process requires that the reader be skilled and that the topic of the text be one for which the reader has well-established knowledge in long term memory. Thus, although it would be a time- and effort-consuming process for an intern to list all the characteristics of a given disease, a highly-skilled specialist with expert levels of knowledge of that disease could do so quickly and easily, using a single retrieval function, and producing a prodigious amount of information about the symptoms and diagnostic criteria using LTWM. Similarly, a reader with well-established knowledge and expertise on a given topic can produce a LTWM structure when an appropriately connected proposition is active in working memory. This greatly expands the function of traditionally defined working memory, and allows for the comprehension and robust memory of complex texts (Kintsch, Patel, & Ericsson, 1999).

With regards to stereotype threat, it is reasonable to suggest that those most likely to experience this phenomenon will have developed a degree of expertise in the nature

and content of existing stereotypes. They are, in a sense, experts in the stereotype that is pertinent to them. Therefore, they should have prodigious amounts of stereotype relevant information available in a single retrieval function when reading a text they perceive to be about some aspect of the stereotype.

Research by Zwann and Truitt (2000) suggested the potential applicability of the LTWM theory to text comprehension. They investigated the relative ability of smokers and non-smokers to suppress smoking relevant information. Using single sentences rather than complete texts, they provided data supporting the hypothesis that personal knowledge can negatively affect text processing. Participants were asked to read a sentence, then to judge quickly whether a subsequent word was related to the meaning of that sentence. The terminal word of each sentence had several possible meanings, which meant that participants had to first comprehend the sentence before addressing the relation to the subsequent word. Those participants who were identified as smokers had longer response times and lower accuracy scores when the target word was related to cigarettes (tar, nicotine, ash) than did non-smokers. The implication was that smokers, who presumably have a more elaborate knowledge structure related to smoking, were less able to suppress smoking-relevant information. The failure to suppress this information of course implies this knowledge was activated in the first place.

Overall Summary

The literature supporting the existence of stereotype threat as a psychological phenomenon is solid and varied. The effects of stereotype threat have been noted across stereotyped populations including gender, race, and SES. These effects have been induced on academic as well as athletic tasks.

The literature exploring the mechanisms by which stereotype threat operate are not so clear, however. Two main hypothesized mechanisms for stereotype threat reported here are anxiety-related intrusions and reductions in working memory capacity. The purpose of the current research is to investigate if and how these mechanisms may generalize to the real-world task of reading comprehension. An unexplored possibility is that these two processes operate conjointly. Using the framework of the CI model of comprehension, the processes of understanding or misunderstanding a text are potentially made clearer.

It was hypothesized that, when operating under a stereotype threat condition, participants would experience disruption of performance primarily through the activation of threatening evaluative information at the situation model level of representation. When Black participants were primed to approach a task as if it were a measure of their intelligence (under stereotype threat conditions), Watkins' (2004) work suggested their need to understand the task in these terms would activate strong evaluative elements from long-term memory that would be incorporated into their situation model level of representation of the text. This information, because it is not the intended meaning of the text, may interfere with their memory representations of the text, reducing their overall levels of performance. Depressed performance levels would be seen in terms of slower processing of the texts and more false recognitions of test sentences that are evaluative in nature or reflect their situation model representation of the text and not the author's intended model. This process should occur regardless of text topic.

A second mechanism through which stereotype knowledge may be activated is through strongly connected LTWM knowledge structures related to stereotypes. If the

text is about stereotypic information but the participant is not operating under conditions of stereotype threat, those with high levels of this knowledge may actually be aided in their high level understanding of the text by the activation of relevant knowledge from long-term memory via the LTWM process previously discussed. In such a case, test performance should be more accurate, but perhaps slightly slower due to the added retrieval function.

If the Black participants are under conditions of stereotype threat and the text is about stereotypic information, both the situation model level activation of intrusive thoughts and the LTWM access of stereotype knowledge should occur. It is unclear at this point what the influence of these two processes operating simultaneously may have on memory and processing of the texts.

CHAPTER 3

METHODS

Participants

The stereotype used in this research was that of African-Americans'
underperformance on academic tasks. Therefore, 34 African-American students attending
a rural, predominantly Caucasian university in southwestern Pennsylvania were included
in this study. This group was selected based on the hypothesis that Black individuals
attending a predominantly White university would have a high degree of experience in
dealing with the processes of stereotype threat. In addition, 29 Caucasian students were
used for a control condition. All participants were members of general psychology
courses, and participated in partial fulfillment of course obligations.

Materials

Measures

Participants were asked to complete the form G or H version of the Nelson-Denny
Reading Test (NDRT). The specific form was selected randomly by alternating which
form was handed to the participants as they entered the testing area. The NDRT is a brief
assessment tool used to measure vocabulary, reading comprehension, and reading rate.
The normative population consists of high school and college students, and the measure
is often used by collegiate admissions departments, including the one at which this study
was conducted, to determine placement of incoming students in appropriate courses
(Brown, Fishco, & Hanna, 1993). This measure was chosen because research on text
comprehension has identified two primary sources of individual differences in reading
comprehension. One source of variance in test scores is the readers' overall

verbal/reading abilities (See Ericsson & Delaney, 2003, for further discussion). Thus, NDRT scores were obtained to serve as a baseline measure of reading skill, which was used during statistical analyses as a covariate to control for individual differences in reading abilities. The other source of variance comes from the reader's knowledge of the text topic (Recht & Leslie, 1988; Walker, 1987; Schneider, Korkel, & Weinert, 1989), which was preseumed to be equivalent across texts.

Texts

Participants read three experimenter-devised texts modeled to reflect activities and topics with which the participants likely possessed similar degrees of familiarity (e.g., walking to class at a university). The knowledge of the text topic was, for the purposes of the current research project, presumed to be roughly equivalent for the non-stereotypic texts, which revolved around the subjects of air travel and a species of marine mammal (see Appendices B and C respectively for the complete texts). For the subject of air travel, there exists a consistent and widely accepted knowledge base regarding what can be expected to occur while flying. This holds true even for those who do not regularly fly (Galambos, 1982). Thus, the travel text was presumed to be equally familiar and accessible to all readers.

The text on the life of orcas was likely less familiar to the participants, which increased the difficulty somewhat for all participants. This enhanced difficulty was expected to show the greatest impact on those participants operating under conditions of stereotype threat, because it required not only the processing of unfamiliar information, but the processing of irrelevant evaluation concerns as well. It was expected that the

NDRT covariate would have a greater effect here, because most participants had to rely solely on reading ability.

In regards to the stereotype text, the scriptal content of the text focused on a topic familiar to all participants: starting out at college (see Appendix D for the full text). The presumably greater knowledge of racial stereotypes possessed by African American students in general served as the hypothesized primary source of variance in outcome measures. This provided African American students with greater stores of LTWM knowledge pertaining to the overarching difficulties experienced by the protagonists in this story, who are two Black students struggling with the adjustment involved in attending a predominantly White university. Furthermore, this greater access to relevant LTWM structures was presumed to increase the number of memory traces available, and was expected to produce a higher "yes" response rate for African American participants than Caucasians, when viewing test sentences with an evaluative component. When African Americans operating under conditions of stereotype threat read this text, it was unclear what would happen. While these individuals possessed more LTWM structures relating to the topic, they also were subjected to the flooding of working memory with the evaluative concerns that seem to typify stereotype threat.

A text length of approximately 600 words was chosen because the length of SAT essays, a task that is realistic but challenging for undergraduate students, provided an involved reading comprehension task. A shorter practice text was also included to allow participants to acclimate to the procedures. The text titles were used to separate the texts from one another in the presentation.

Recognition test sentences were constructed for each text (see Appendices B, C, and D for the test sentences). They followed the general format of Zimny (1987) and Schmalhofer and Glavanov (1986). There were several types of test sentences. "Old" sentences were actually included in the texts ("Orcas have no natural enemies"). "Paraphrase" sentences included minor modifications of the wording of actual sentences from the text with meaning intact ("In fact, she knew that driving was generally considered to be less safe than flying"). "Inference model-correct" were sentences that were not included in the text but accurately reflected the situation described in the text ("Marcus and Tyra grew up in an urban area, far from the countryside"). "Inference model-stereotype" sentences were sentences not included in the actual text but were consistent with a representation of the text colored by a tone of worry or anxiety ("The constant threat of terrorism is always a concern when flying"). These inference model-stereotype sentences focused on the evaluative and anxious rumination element central to stereotype threat. Finally, "inappropriate" sentences, which were unrelated to any content of the texts, were included as controls to establish baseline error rates for individual participants ("The brown cat slept peacefully on the rug"). The rationale for these sentence types is discussed below.

For the purposes of recognition data, the proportion "yes" responses reflected the contributions of varying memory traces to the participant's response. An old sentence was represented at the verbatim, textbase, and model-correct levels of representation. A paraphrase sentence was represented at the textbase, and model-correct levels of representation. An inference model-correct sentence was represented only at the situation

model correct level of representation. An inference model-stereotype threat sentence was represented only at the situation model-stereotype threat level of representation. Each trace level can be conceptualized as contributing to the overall recognition trace strength of each of the test sentences. Thus, test sentences that are represented at more trace levels should produce higher levels of recognition memory. The "inappropriate" sentences, which are not represented at any level of the text, served as a baseline indication of the error rate for each participant. See Table 1 for a visual representation of these memory traces. These traces were added to produce varying likelihoods for a "yes" recognition response (Kintsch, Welsh, Schmalhofer, & Zimny, 1990).

<center>Procedures</center>

A brief demographic questionnaire was distributed to students enrolled in general psychology courses on the first day of class. This included questions regarding general demographic information such as age, race, and area of origin. The students' identification numbers were also requested. The identification numbers and race information were used to choose participants for the racial groups. The balance of the survey was filler information that was used only to mask the nature of the questionnaire, and was discarded when the participants for the second phase were selected. The questionnaire can be found in Appendix G.

Participants that were randomly chosen from the appropriate racial lists generated by this survey were called to participate in the study. They were identified from this point on by an experimental number, and all identifying information was destroyed. It was explained that they would be participating in a brief study involving a reading task, and that they would first be asked to complete the NDRT. The NDRT was administered in a

racially heterogenous group setting in order to facilitate a more rapid processing of the results. Participants were scheduled at this time for completion of the experimental task at a later date, within two weeks.

Participants were run individually through the completion of the next phase of the experiment. In order to control for any potential effects of the experimenter's race, half of the participants in all conditions were supervised by a Caucasian-American experimenter, and half were supervised by an African-American experimenter. Each participant was first read a brief introduction by the experimenter. They were told that they would be reading some short stories on the computer, and then completing a sentence recognition task. Stereotype threat was manipulated by informing the participants in the threat condition that the experiment was designed to measure verbal intelligence. It was emphasized that verbal abilities are very important to academic success. The control condition participants were informed that the study was designed to aid textbook manufacturers, who were interested in shifting to computer-based texts. The participants were told that these designers wanted to know how eye strain would affect students asked to read such a text. These sets of instructions can be found in Appendix E for the control condition, and Appendix F for the experimental condition. The operation of the text presentation system on the computers was described, and an opportunity was provided for questions.

Participants each received the same set of texts. The order of text presentation was randomized, with the exception that the shorter practice text was always presented first. This practice text (Appendix A) was a simple reiteration of the procedures for advancing the text. The last sentence of each text was separated from the recognition sentences by a

brief pause and a message asking participants to press the space bar to proceed. Texts were presented one sentence at a time, with participants using the space bar to advance to the next sentence. Reading times per sentence were collected by the computers.

Recognition sentences were blocked by relevant text, but were randomized within the block. Each block was cued by the relevant text title, and was presented immediately following the related text. Participants were asked to confirm or deny having seen each sentence in the relevant text. This was accomplished by an on-screen prompt directing the participant to press the "F" key for an affirmative response and the "J" key for a negative response. The responses for each sentence were recorded by the computer. Sentence types were distributed between old, paraphrase, model-correct, model-stereotype threat, and inappropriate sentences. After the final recognition item, the title of the next text was presented. Following completion of the recognition sentences for the final text, participants were fully debriefed both verbally and in writing, and thanked for their participation.

<div align="center">Design</div>

<div align="center">*NDRT Data*</div>

The data from the NDRT were analyzed using a one-way ANOVA (with no covariate included) and ANCOVA (with the covariate included) with two levels of the independent variable of race. This was done in order to identify any preexisting differences between the racial groups and to control for variations in reading ability in the interpretation of reading time data.

Reading Time Data

The reading times for the texts were analyzed using a 2 stereotype condition (threat, non-threat) x 2 race (Black, White) x 3 text type (travel, orca, stereotype) ANCOVA with repeated measures on the last variable. The NDRT measures were utilized as the covariate.

Proportion "Yes" Recognition Responses

The sentence recognition data was analyzed using a 2 stereotype condition (threat, non-threat) x 2 race (Black, White) x 3 text type (travel, orca, stereotype) x 5 sentence type (old, paraphrase, model-correct, model-stereotype threat, inappropriate) ANOVA with repeated measures on the last two variables as the proposed analyses. The NDRT measure was the covariate.

The predicted results for the recognition test sentences were somewhat complex. Under most circumstances, the pattern of recognition would be old sentences > paraphrase sentences> model-correct sentences. This is due to the fact that the combined strength of the memory traces endorsing old sentences is greater than that of paraphrase sentences, which is greater than that of situation model sentences. This pattern was expected to hold for all participants, both Black and White, under non-stereotype threat conditions, for non-stereotype texts. For the stereotype texts, Black participants were expected to possibly show heightened levels of yes responses to old, paraphrase, and model-correct sentences because of their superior preexisting knowledge of this topic. In essence, their access to LTWM structures provided them with a greater base upon which to build the stereotype-relevant text, and created larger memory traces for stereotype relevant information.

The general discourse pattern was expected to be altered by the stereotype threat manipulation. Overall, Black participants under stereotype threat were likely to show depressed levels of old, paraphrase, and model-correct sentences as compared to White participants in general and Black participants under the non-stereotype condition. This was expected to occur due to the interference of stereotype threat on working memory. But they were also likely show heightened levels of recognition for model-stereotype threat sentences because this information was expected to be activated through LTWM during text processing.

For the stereotype text under stereotype threat conditions, it was unclear what the results would be. Although intrusive, non-text thoughts were activated by the threat condition, superior stereotype knowledge was also be activated by the text per se. It was unclear which process would predominate in this case.

CHAPTER 4

RESULTS

First, a brief overview of the organization of the different analyses is presented. The first group of analyses addresses the issue of the pre-experimental equivalency of our major factor, race, by analyzing the participants' differences on the NDRT and its various subscales. The next set of analyses examines the reading time results for the experimental texts both with and without using NDRT- R as a covariate. These analyses provide information regarding the efficiency and ease with which participants access relevant information and process a text. Next, the "proportion yes" responses to test sentences from the experimental texts were analyzed with and without using NDRT-T as a covariate. These analyses explore the different levels of representation for the texts as outlined by Kintsch (1998). Finally the "proportion yes" responses were transformed using signal detection theory into d', β, and trace strength measures and analyses were conducted on these transformed measures. The d' and β analyses provide a distinction between strength of memory trace and the decision processes of the readers. The trace strength analysis has been used in prior discourse research to examine the relative contribution of different sources of information to memory for sentences (Kintsch, Welsch, Schmalhofer, & Zimny, 1990). Given the complexity of the experimental design, marginally significant results were reported when they are consistent across analyses and interpretable.

The race of the experimenter implementing the project with participants and the participants' genders were initially included in all phases of these analyses. The inclusion

of these variables resulted in no main effects or interaction effects. Therefore, these variables will not be addressed further here.

<center>Nelson Denny Reading Test Analyses</center>

The NDRT generates four scores for each participant: Vocabulary (V), Comprehension (C), Total (T), and Reading Rate (R). The V score reflects performance on the subtest measuring general lexical knowledge. Scores on the Vocabulary measure indicated that White students ($M = 217.28$ points) performed significantly better than did Black students ($M = 197.5$ points), $F(1, 62) = 17.37$, $p < .01$.

The C score reflects performance on the subtest measuring integrative skill and general understanding of the texts. As described above, text comprehension is a complex and active process. This process, if successful, culminates in a coherent understanding of the text's meaning. The C scores for White participants ($M = 215.72$ points) were significantly higher than those of their Black peers ($M = 196.91$ points), $F(1, 62) = 15.38$, $p < .01$. The T score is an average of the V score and two times the C score, and is thus consistent with the pattern of White students ($M = 216.24$ points) outperforming their Black peers ($M = 197.11$ points). This result again was statistically significant, $F(1, 62) = 19.75$, $p < .01$.

The R score is the measure of how much each participant was able to read within the first minute of exposure to the first text of the NDRT series. White students ($M = 202.44$ points) and Black students ($M = 195.85$ points) did not differ significantly in their R scores.

These results indicate that the participants in this research project have pre-existing differences in their skill at effectively reading and comprehending texts of the

sort used in the NDRT. White students in this sample consistently outperformed their Black peers on both the vocabulary and comprehension subtests. Therefore, the total score (T) was used as a basis for the covariate in recognition analyses. The Black and White students did not differ in terms of their reading speeds for the first minute of reading. Nevertheless, it was decided to perform the reading time analysis calculations both with and without the covariate included. This was decision was made in order to keep the procedures consistent. It is likely the case that the lack of differences noted in the pretest reading time speeds is due to a combination of simpler texts and measuring only the first minute of reading.

Various results were changed by the inclusion of these covariates. Much of the prior research on stereotype threat has not used a covariate to control for preexisting differences in reading skill levels. Steele and Aronson (1995) did use a self-report of the students' SAT scores as their covariate. However, self-report responses are questionable due not only to the possibility of faulty memory for SAT results, but also due to the potential for social desirability to inflate participants' reporting of their scores. Quinn and Spencer (2001) and Schmader and Johns (2003) used a restricted range of participants, selecting only those who had achieved a score surpassing a cutoff point on the relevant portion of the SAT. The majority of other projects reviewed used no covariate. Results have been presented both with and without the covariates included when a significant discrepancy was generated.

<div align="center">Reading Time Analyses</div>

The reading time data were generated by recording the amount of time taken to read each sentence. These results were then divided by the number of words in the

sentence to produce a reading time per word per sentence. The reading times per word per sentence were averaged for each participant to produce a reading time per word per text.

The texts were written to present a variety of challenges that might be exposed in the reading time analyses. The first text (Air Travel) was a narrative describing over-seas travel by air in a job context. The second text (Orca) was an expository article on the life and habits of orcas, and reflected a style of writing common in educational texts. The third text (Big Changes) was another narrative containing information that would be familiar to many of these participants, especially the African American students. It was about moving from a predominantly urban environment to a rural, largely Caucasian environment to attend college. The Orca text would presumably be the most difficult, while the Big Changes text would be presumably the most familiar topic to the readers. The Air Travel text would likely be a mix of both familiar and unfamiliar material.

The reading times for the experimental texts were analyzed using a three way ANOVA (participant race x threat condition x text type) with repeated measures on text type. The data were first analyzed with the R score from the NDRT as a covariate for the reading time data. The dependent measure was average reading time, in milliseconds, per word across each text. There was a marginally significant main effect of race $F(1, 58) = 3.56, \rho = .06$. The White students ($M = 314$ msec/word) took less time to read the texts than did their Black peers ($M = 348$ msec/word). When analyzed without the inclusion of the NDRT-R covariate, this effect was stronger, $F(1, 59) = 5.69, \rho = .02$. The experimental text reading time results were not consistent with the R score analysis of the NDRT, which indicated that the Black and White participants read at the same rate. This

difference may be due to the fact that the NDRT only measures the first minute of the reading task. It is possible that, while the Black students may not have fallen behind in the first minute of the R task of the NDRT, a task that took approximately 10 minutes might show a difference. This result suggested a limitation in the use of the NDRT-R for the investigation of longer, more complex texts.

There was a significant main effect of text, $F(2, 57) = 3.68$, $\rho = .031$. The Orca text ($M = 377$ msec/word) did take significantly longer to read than the Air Travel text ($M = 315$ msec/word) and the Big Changes text ($M = 300$ msec/word). This result was stronger when the analysis was run without the covariate, $F(2, 58) = 59.82$, $\rho < = .01$. A Bonferroni procedure indicated that the Orca text was read more slowly than both the Air Travel and Big Changes texts, which were not significantly different from each other. This result supports the hypothesis that the Orca text was indeed more challenging than the two narrative texts. This relationship is illustrated in Figure 1.

While there was no main effect of threat on reading times, the threat condition did interact with other variables. There was an interaction between text and threat condition, $F (2, 57) = 3.51$, $\rho = .04$. This result was not altered significantly when the analyses were run without the covariate. These results are illustrated in Figure 2. A test of the simple effects on the univariate results indicated that there were significant increases in reading times in the threat condition for the Air Travel and Orca texts. There was no difference between threat conditions for the Big Changes text.
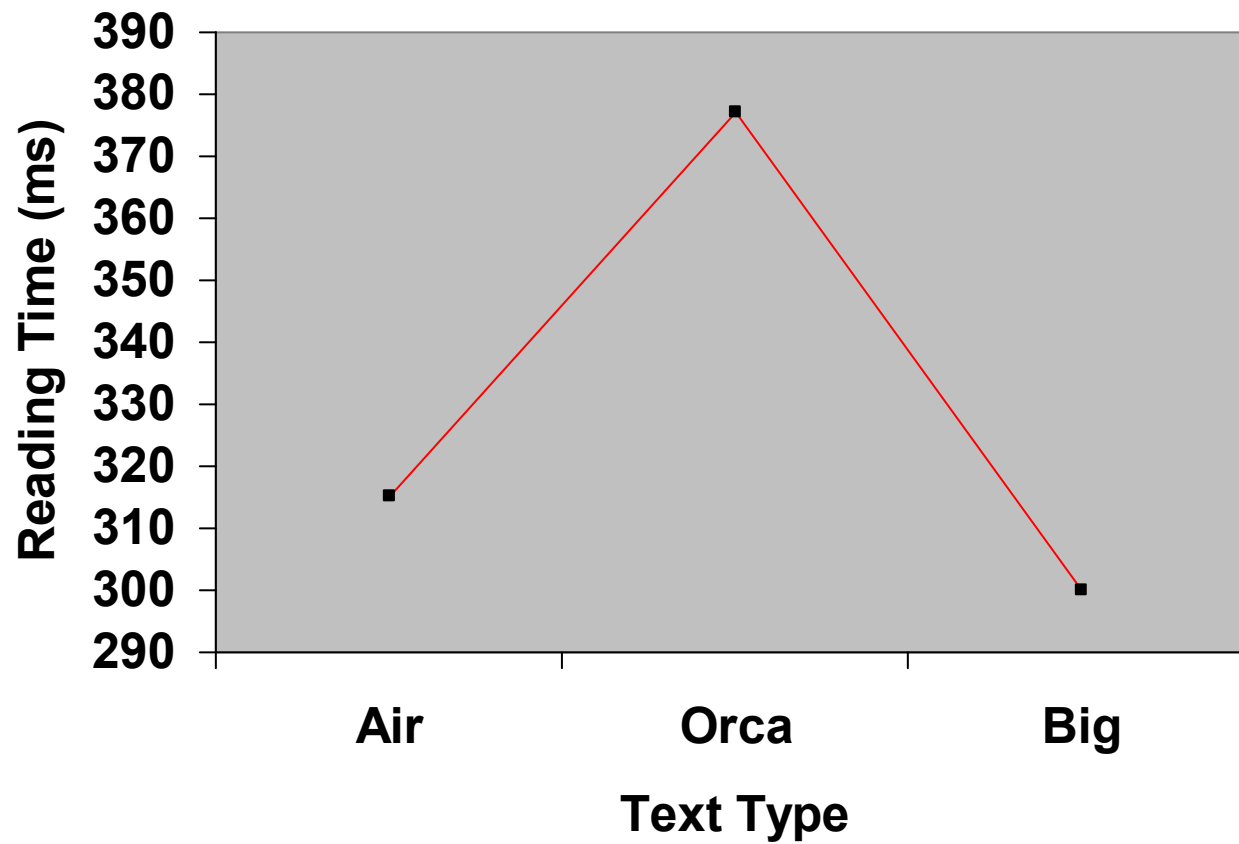
Figure 1. Nelson-Denny reading time as a function of type of text being read.
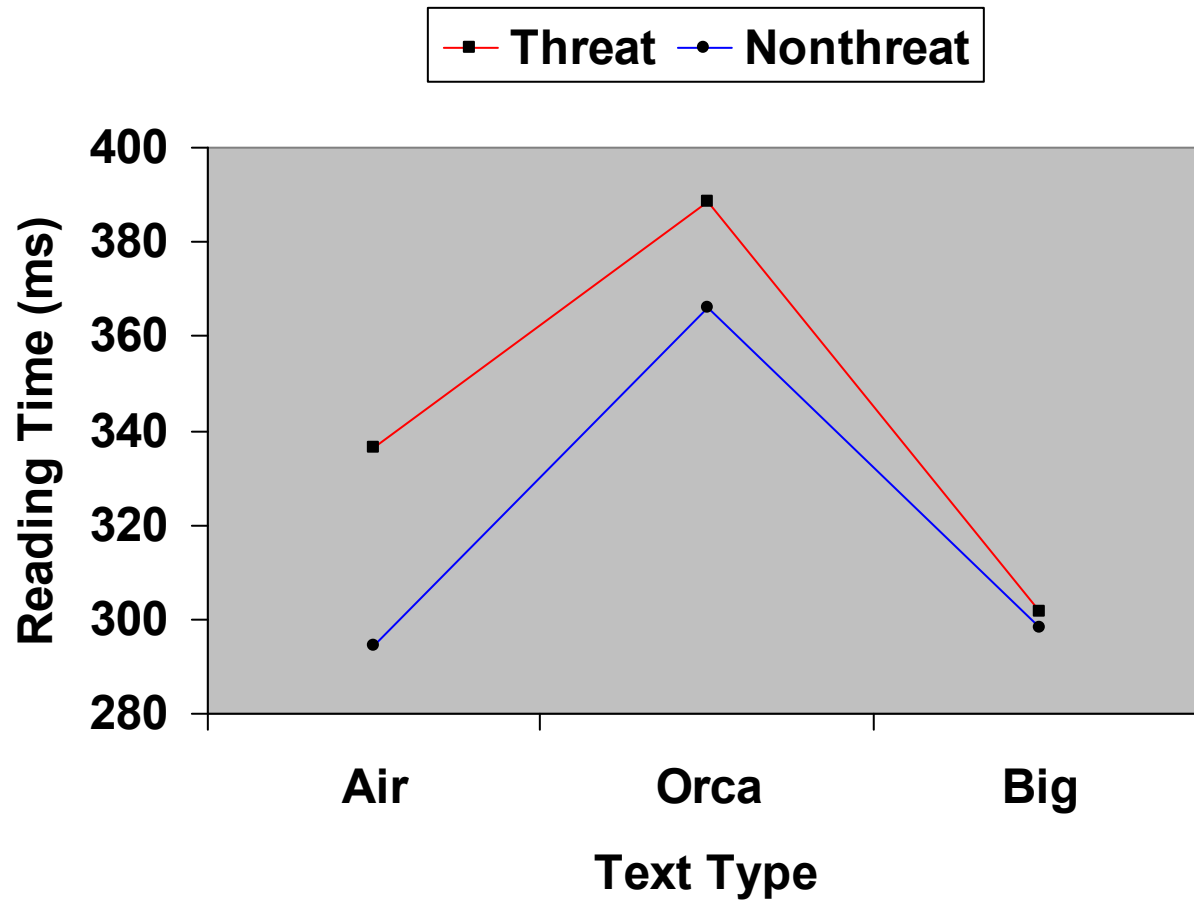
Figure 2. Nelson-Denny reading times as a function of threat condition and text type.

In summary, the reading time data indicated that the narrative texts were processed more quickly than the expository text. The Black students read more slowly across texts than did their White peers. Given that there were no group differences according to the NDRT-R, this suggests the R measure of the NDRT may not be adequate as a measure of longer and more complex texts. Text and threat interacted to produce an increased reading time for both races in the threat condition for the Orca and Air Travel texts, but not the Big Changes text. These results indicate that threat, as measured by reading time, was experienced by both Black and White participants in this study. Threat had an effect when the text was expository rather than narrative, and/or when the text pertained to unfamiliar rather than familiar subject matter.

<center>Recognition Response Analyses</center>

The recognition data consists of "proportion yes" responses to the recognition sentence stimuli. Each text was accompanied by 23 recognition sentences falling into one of five categories: exact duplicates of sentences presented in the reading task (old), synonym substitutions or slight word order changes to the sentences presented in the reading task (paraphrase), sentences whose content could reasonably be inferred from the reading task (inference), sentences whose content reflects an anxiety or fear-based inference drawn from the reading task (anxiety), and sentences whose content is clearly not connected to the reading task in any fashion (inappropriate). There were 5 old, 5 paraphrase, 5 inference, and 5 anxiety sentences. There were also 3 inappropriate sentences, for a total of 23 test sentences per text.

A "yes" response, indicating that the reader had read that exact sentence in the text, was thus only technically accurate for the old sentences. The research on text

processing (Schmalhofer &Glavanov, 1986; Kintsch, et. al., 1990) shows that the proportion "yes" ratios for these types of stimuli is highest for the old sentences, then for the paraphrase, then the inference, and lowest for inappropriate. The anxiety type sentences were a new addition to the format, and were expected to fall between the inference and inappropriate in terms of frequency of "yes" responses. They were intended to capture the intrusion of anxiety that has been hypothesized as a potential causal mechanism behind stereotype threat (Schmmader & Johns, 2003). The inappropriate sentences were included as insurance against random responding, and all participants showed a very low rate of response to these sentences.

These sentence types were selected to represent the various levels of memory trace that develop as a text is read. As a text is processed, the reader is building a representation of the text at three levels. A surface level representation corresponds to memory for the exact words contained in the sentence. A textbase model represents memory for the meaning of the text as separate from the precise wording of the text. The situational model reflects the relation between memory for the information in the text as it is integrated with the participants' prior knowledge. These three levels of representation combine to produce an overall meaning of a sentence for the reader (Zimny & Robertson, 1997).

The recognition results were analyzed using a four-way ANOVA (participant race x threat condition x sentence type x text type) with repeated measures on text and sentence types. The data were analyzed both with and without the NDRT included as a covariate. Without the covariate, there was a main effect of race $F(1,59) = 4.08$, $\rho = .05$. The White students ($M = .38$ proportion yes) responded "yes" to more of the items than

did the Black students (M = .33 proportion yes). This difference was stronger with the inclusion of the covariate, F(1,58) = 8.45, ρ <.01. The removal of reading skill as a variable made the effect of race on "proportion yes" responses stronger.

There was no main effect of threat condition, F(1, 59) = 2.02, ρ = .16 without the covariate included. With the NDRT-T covariate, removing the variable of reading skill, a marginally significant main effect of threat F(1, 58) = 3.01, ρ = .09 was noted. Those participants operating under the non-threat condition (M = .37 proportion yes) endorsed more of the recognition sentences than did those in the threat condition (M = .33 proportion yes).

Without the NDRT-T covariate, there was a significant main effect of text F(2, 58) = 3.21, ρ = .05, such that the Air Travel text generated a higher proportion "yes" than either of the other texts. This effect was eliminated when the NDRT-T covariate was included. This supports the utility of the NDRT-T as a control for reading skill, since the removal of readers' skill removed this difference in performance.

There was a significant main effect of sentence type F(4, 55) = 603.57, ρ < .01. This result is consistent with previous research and indicates that the standard old-paraphrase-inference-inappropriate distribution of sentences is sound. The anxiety sentences fell into the pattern at the expected point between inference and inappropriate. Participants endorsed the recognition sentences in the predicted pattern of Old (M = .8 proportion yes) > Paraphrase (M = .48 proportion yes) > Inference (M = .38 proportion yes) > Anxiety (M = .09 proportion yes) > Inappropriate (M = .01 proportion yes). This effect remained significant with the inclusion of the NDRT-T covariate, F(4, 55) = 4.35, ρ < .01. This is illustrated in Figure 3.

Without the NDRT-T covariate, there was an interaction between text type and sentence type F(8, 52) = 4.09, $\rho < .01$. The Big Changes text, arguably the easiest of the three experimental texts, produced a different pattern of responses to the sentences than did the Air Travel and Orca texts, with a decreased number of inferences being generated, and a higher number of "yes" responses to the old sentences. These effects were lost when the NDRT-T was included. This would indicate that these differences were due to the effects of reading skill level.

There was an interaction between sentence type and race F (4, 56) = 2.6, $\rho = .05$. This result was not appreciably altered by the inclusion of the NDRT-T, F (4, 55) = 2.41, $\rho = .06$. This is illustrated in Figure 4. A test of the simple effects on the univariate data showed that the Black participants made significantly fewer "yes" responses to the paraphrase and inference sentences and marginally significantly lower responses for the old sentences than their White peers.

There was no significant interaction between threat and race when the NDRT-T was not included. With its inclusion, there was a marginally significant interaction F(1,58) = 3.62, $\rho = .06$, shown in Figure 5. A test of simple effects on the univariate data indicated that White students made fewer "yes" responses under threat conditions ($M = .35$ proportion yes) than under non-threat conditions ($M = .44$ proportion yes). Threat had no effect on Black "yes" responses ($M = .31$ proportion yes across both conditions), and these remained lower than White "yes" responses across threat conditions. One possible explanation for this finding is that the Black students in our sample are operating under a
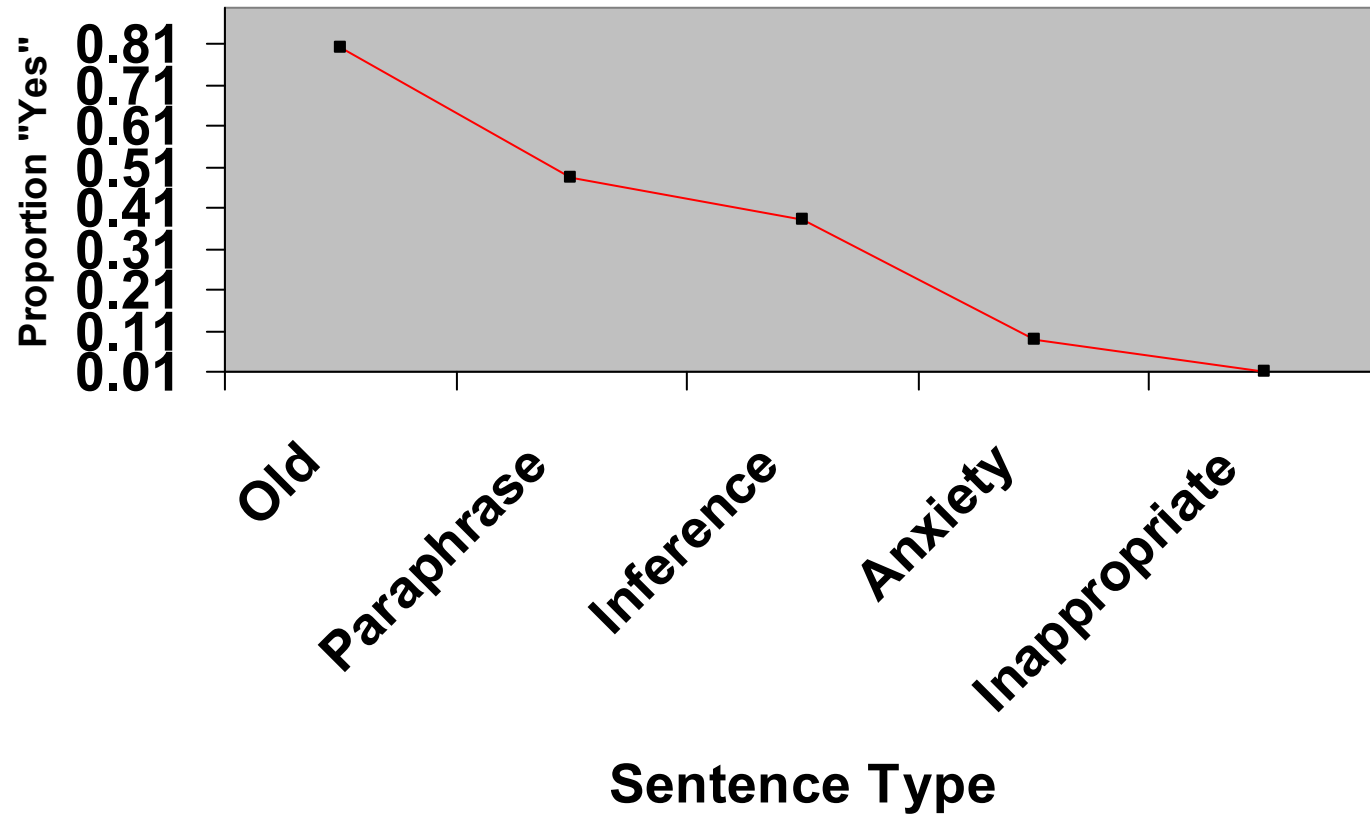
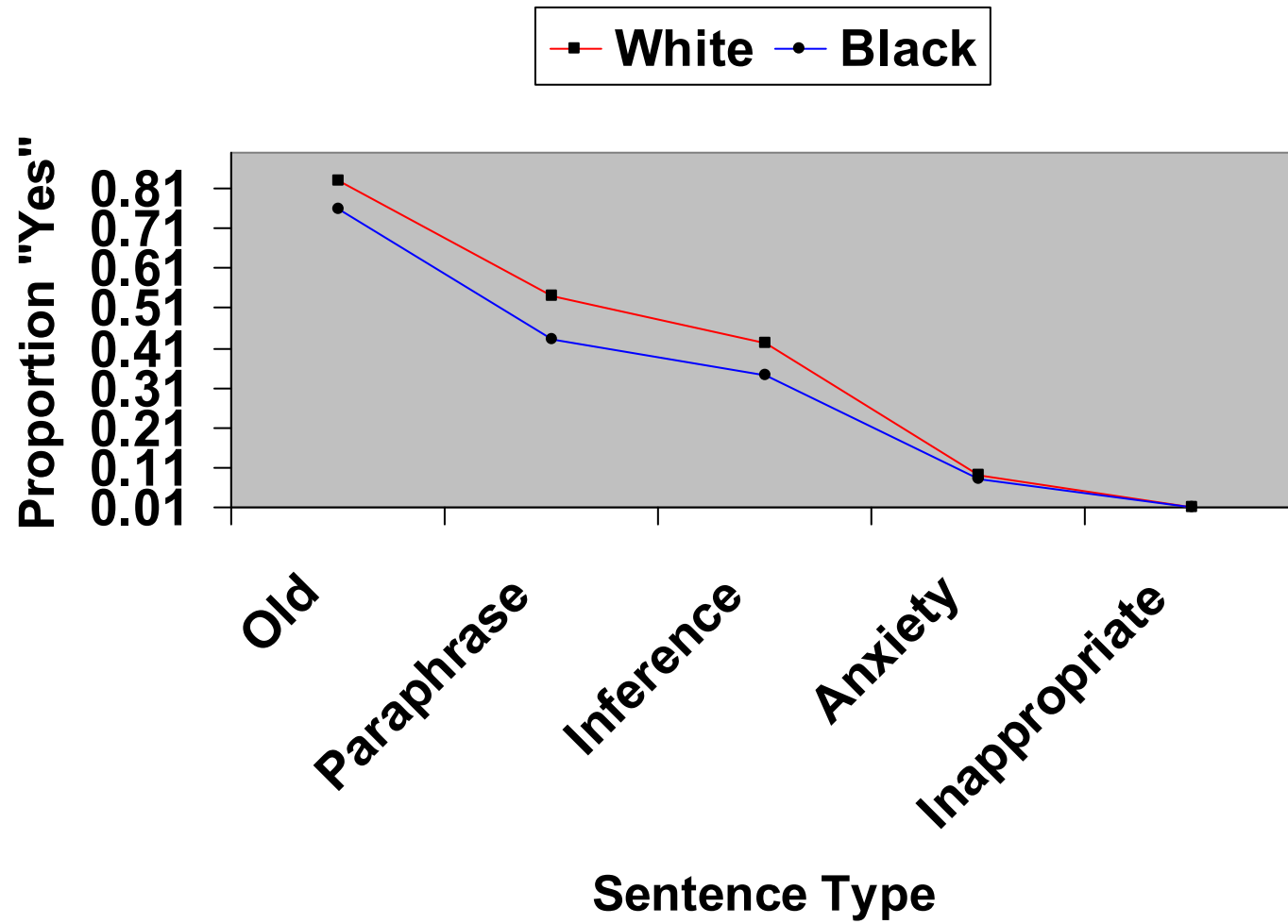*Figure 3*. Proprtion "yes" responses as a function of sentence type.

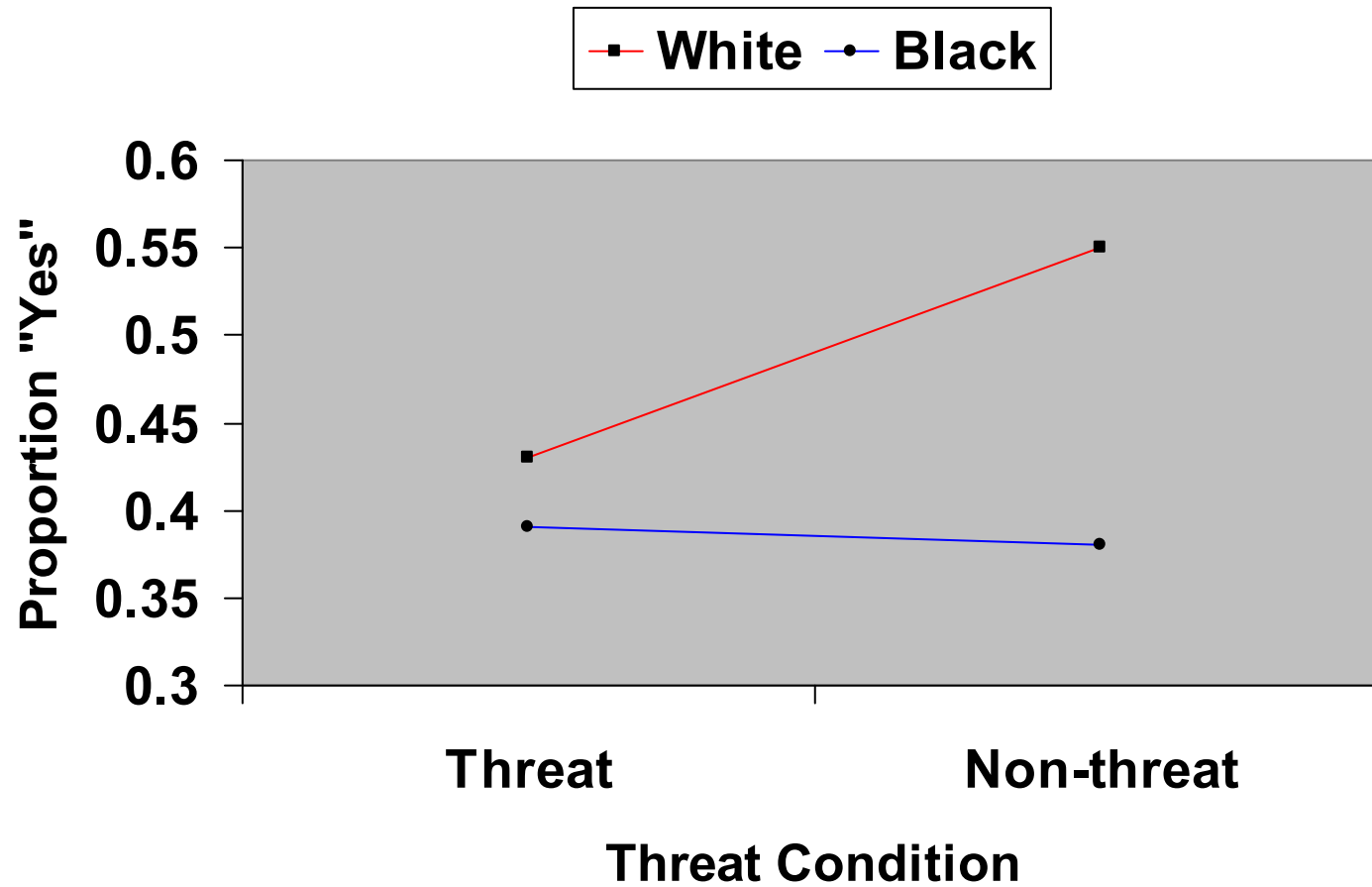*Figure 4*. Proportion "yes" responses as a function of race and sentence type.

Figure 5. Proprtion "yes" as a function of threat condition and race.

constant state of threat, due perhaps to the academic setting in which the research was conducted. The White students appear to have responded to the threat condition with a decreased willingness to endorse a sentence as having been seen before.

There was a marginal three-way interaction between text, threat, and participant race $F(2, 58) = 2.88$, $\rho = .06$. A test of the simple effects on the univariate data showed that the Black participants did not alter their rate of "yes" responses, regardless of threat condition. The combination of the threat condition and a more difficult, expository text resulted in the White participants performing similarly to their Black peers by reducing "yes" responses. There was no essential change generated in this interaction by the inclusion of the NDRT-T covariate $F(2, 57) = 2.41$, $\rho = .08$. Refer to Figure 6. This analysis suggests the Black participants may be operating under a continual state of threat, regardless of experimental manipulation. Meanwhile, the White students responded to the threat manipulation with a reduction in "yes" responses particularly when the text was expository or about a relatively unfamiliar topic.

In summary, the sentence types followed the typical pattern, with the anxiety sentences generating more "yes" responses than the inappropriate sentences, but fewer than the inference sentences. The White students made more "yes" responses overall than the Black students did. The inclusion of the threat condition resulted in a marginally significant reduction in "yes" responses for the White students. White students made more "yes responses" on paraphrase and inference sentences than their Black peers, and marginally significantly more on the old sentences. The White students responded to the threat condition with a reduction in "yes" responses for the most challenging of the three texts, while the Black students did not respond to the threat
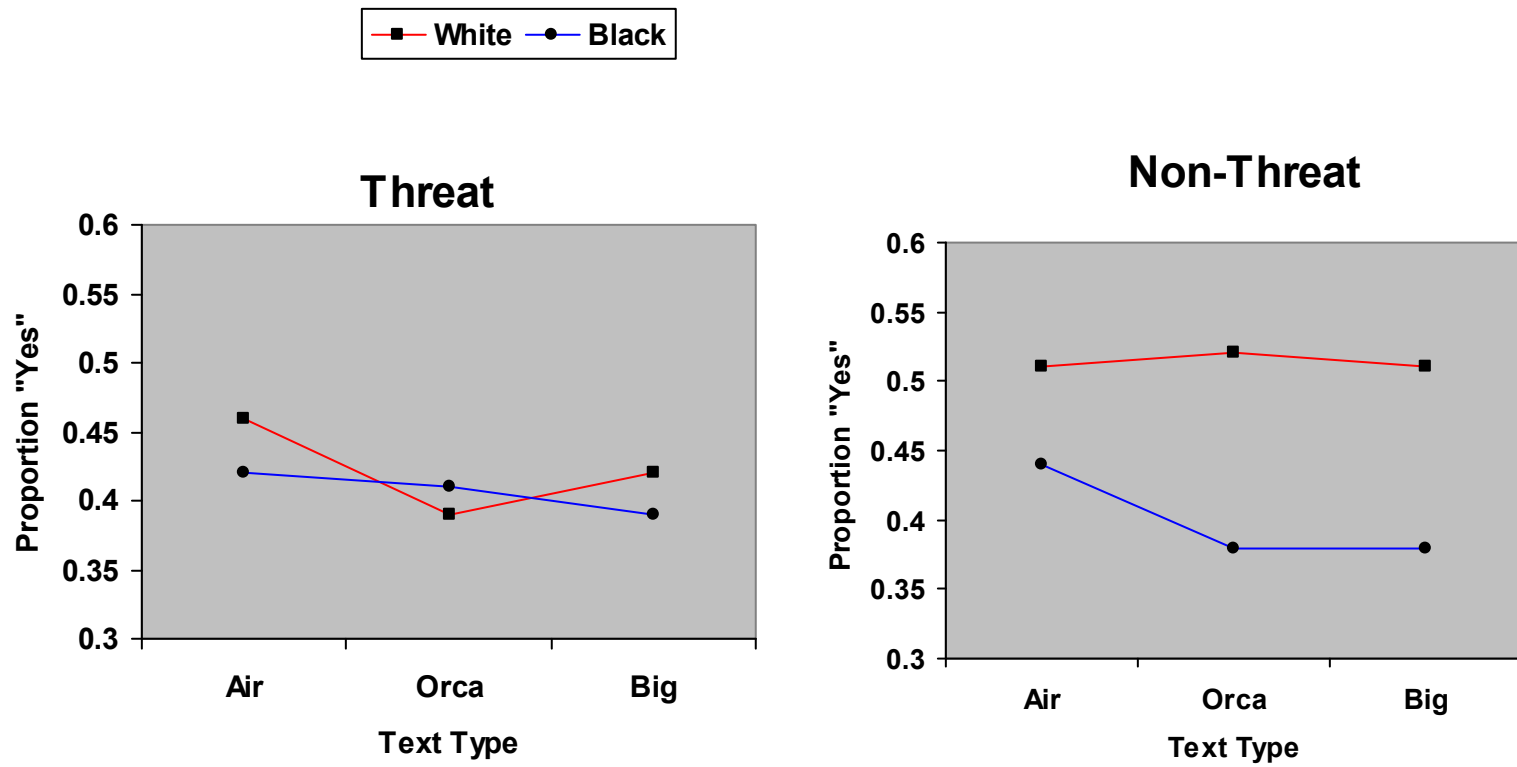
Figure 6. Proportion "yes" as a function of participant race, text type, and threat condition.

condition in any noticeable fashion. That is, White students operating under threat, and reading a challenging text, produced a similar pattern to that seen from the Black students under all conditions. The emerging pattern of results suggests that the Black students are perhaps responding to a continuous state of threat, while the White students are apparently responding to the combination of text difficulty and threat manipulation in a fashion that is generally seen among the Black students in the research literature.

Signal Detection Analyses

Overview

Signal detection is a technique that allows the investigator to identify characteristics of the participants' responses in terms of the strength of the target and/or the sensitivity of the receiver and the criterion for a response. Specifically, the signal detection analysis transforms "yes" responses into the relative intensity of a given stimuli (labeled d') and the point at which a particular participant sets their response criteria (labeled β). In the case of this experiment, d' represents the strength of the memory traces that contribute to a "yes" response on the recognition items. A higher d' indicates a stronger memory trace. The β scores represent the point at which a particular respondent makes a determination that a "yes" response is appropriate. This measure can be influenced by a number of factors including the value to the participant of different responses. The error rates necessary to calculate both d' and β were the response rate for the inappropriate sentences. The data were analyzed with the covariate of the NDRT-T score included, in order to control for the effects of preexisting reading skill.

d′ Analyses

Separate *d′* and *β* scores were calculated for each participant for each sentence type. The *d′* data were analyzed using a four way ANOVA (participant race x threat condition x sentence type x text type) with repeated measures on text type and sentence type. There was a significant main effect of race F(1, 58) = 6.48, ρ = .01. The White students (*M* = 2.19) scored significantly higher overall than did their Black peers (*M* = 1.72). This indicates that the White students had significantly stronger memory strengths across all situations.

There was also a marginally significant three way interaction of race by text by threat, F(2, 57) = 2.45, ρ = .09. A test of the simple main effects performed on the univariate analyses showed that the Black students had lower *d′* than White students under non-threat conditions. Under threat conditions, the Whites students had marginally significant or significantly higher *d′* for the Air Travel and the Big Changes texts but had equally low *d′* for the more difficult Orca text. With the more challenging text, under the threat condition, the White and Black students all had equally low memory strength. Refer to Figure 7.

In summary, the White participants in this sample had a stronger representation of the sentences, as measured by *d′,* than did their Black peers. The White students responded to the threat manipulation with a lower *d′* for the more challenging text than they produced under the non-threat conditions. This is consistent with recognition analyses, and supports the growing evidence that the Black students in this sample were operating under threat conditions on a constant basis, while the White students became

threatened when the task was presented as indicative of intellectual ability and the text was more difficult.

<div align="center">*Memory Trace Strength Analyses*</div>

The trace strengths for each of the three levels of text representation (surface, textbase, and situation model) were determined by transforming d' into trace strength in the following manner: for the surface level, subtracting the d' for paraphrase sentences from the d' for the old sentences for each participant computes the strength of the memory trace for the exact wording of the sentences. Subtracting the d' for the inferences from the d' for the paraphrases for each participant left the d' representing the textbase level of representation. Finally, the situation model d' was determined by subtracting the d' for the inappropriate sentences from the d' for the inferences for each participant.

These transformed data were analyzed using a four way ANOVA (participant race x threat condition x text type x memory trace) with repeated measures on text type and memory trace. There was a marginally significant main effect of trace $F(2, 57) = 2.67$, $\rho = .08$. A Bonferroni procedure indicated that the participants had a stronger memory trace for the surface level ($M = 1.18$) and the situation model ($M = 1.34$) than they did for the textbase ($M = 0.41$).

There was also a marginally significant three way interaction between text type, race, and trace $F(4, 55) = 2.3$, $\rho = .07$. A test of simple effects on the univariate analyses showed that, for the Orca text, Black respondents had a stronger memory trace for the surface level, while White respondents had a stronger memory trace for the situation model. There were no other significant differences between the Black and White
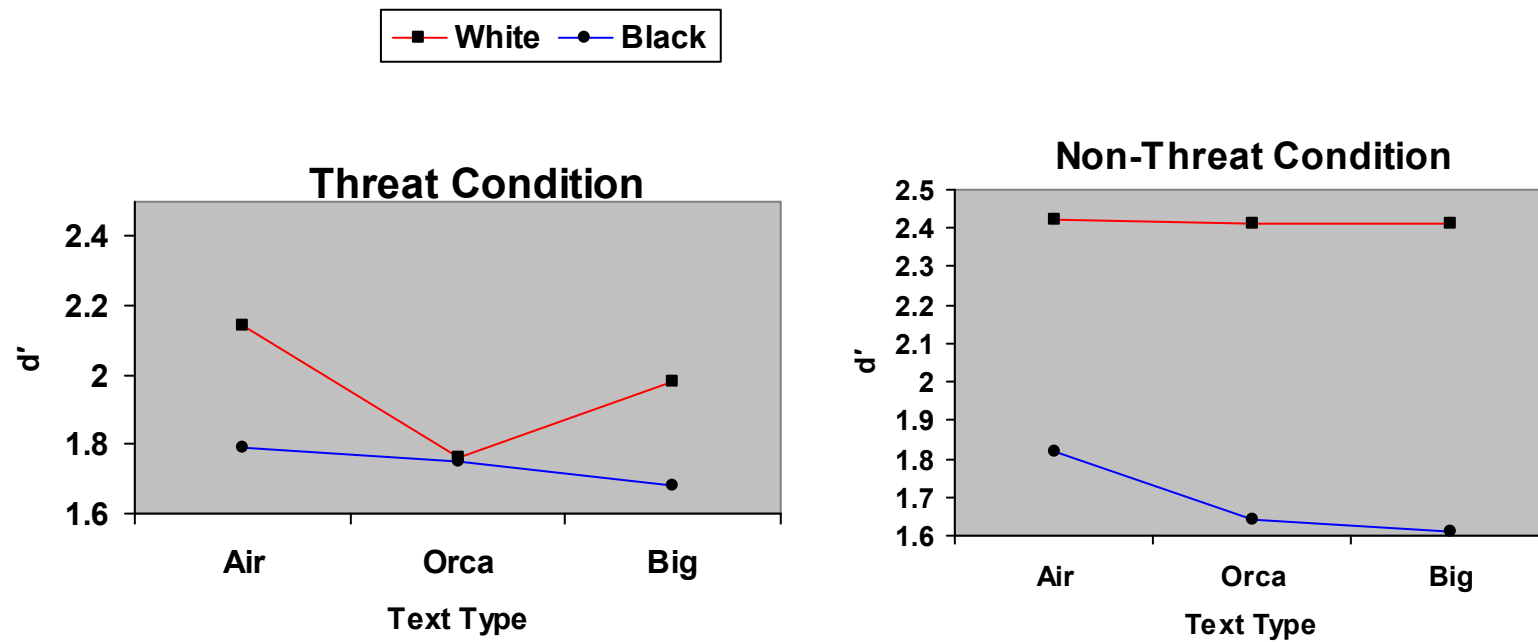
Figure 7. d' as a function of participant race, threat condition, and text type.

participants across the remaining texts. When the text was challenging, the Black students relied heavily on the precise wording of the text (surface level of representation), while their White peers relied on their own prior knowledge of the subject (situation model). Refer to Figure 8.

In summary, all participants showed stronger memory traces for the surface and situation model levels of the text than for the textbase level of representation. When a text was challenging, the participants showed racial differences in their response patterns, with the White participants having a stronger memory trace for the situation model, while their Black peers had a stronger memory trace for the surface level of the text. Text difficulty resulted in the participants selecting different strategies, with White students tending to focus on the situation model level of representation, while Black students tended to focus on the surface level. The implication would be either that White readers have more knowledge or experience from which to construct their situation models, or that Black readers were less skilled and strategically focused on the exact wording in order to maximize performance.

<center>*β Analyses*</center>

The *β* scores represent the criterion that respondents set that determines whether a "yes" response is appropriate for a given sentence. The reasons for criterion placement have received relatively little theoretical and empirical investigation compared to analyses of d'. But some research has suggested that people can shift criteria from test item to test item and generally exhibit more stringent criteria when the item is old or more familiar (Rhodes & Jacoby, 2005; Singer, Gagnon & Richards, 2002). However,
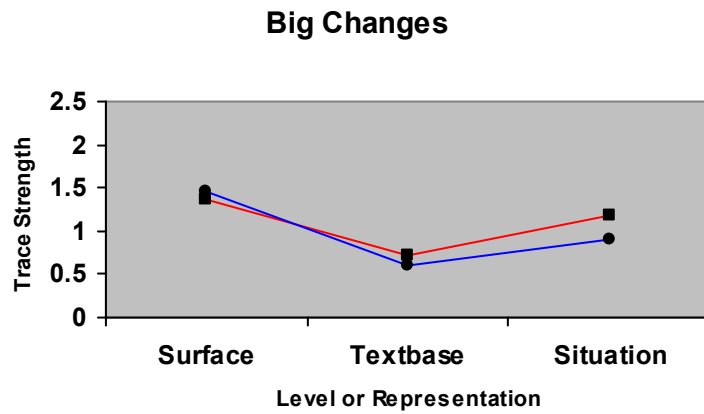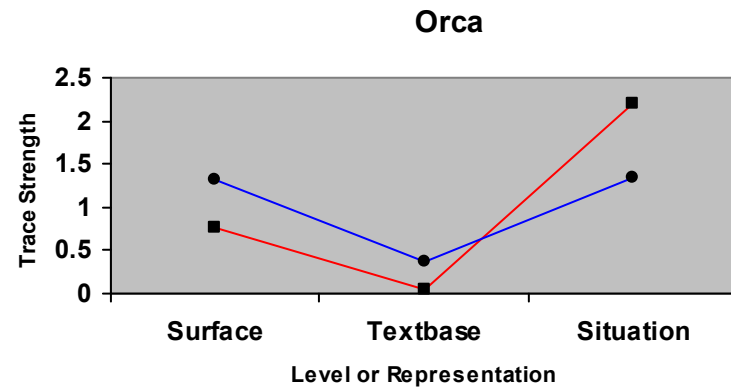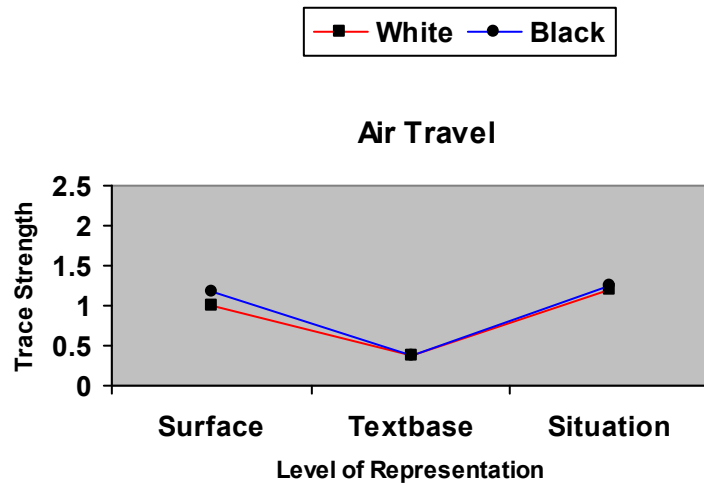
Figure 8. Trace strength as a function of text type, participant race, and level of text representation.

most of this research on criteria shifting comes from recognition tests of simple items from a list. Just how familiarity of a sentence represented at multiple levels from a complex text may or may not influence criterion shifting for a sentence remains to be fully explored. However, there are some consistencies between the simple items and text results reported here.

The β data were analyzed using a four way ANOVA (participant race x threat condition x sentence type x text type) with repeated measures on text type and sentence type. There was a significant interaction between sentence type and race F(3, 56) = 3.69, ρ = .02. A test of the simple effects using the univariate results indicated that the Black students set a more stringent criterion for the old sentences than their White peers, had a similar criterion for paraphrase sentences, and were significantly less stringent than the White students for inference and anxiety test sentences. This is consistent with earlier observations regarding the Black participants' use of more stringent criterion for familiar items, and suggests a reliance on exact wording. Refer to Figure 9

There was a marginally significant three-way interaction between text type, sentence type, and race F(6, 53) = 2.07, ρ = .07. A test of simple effects on the univariate analyses showed that Black participants set their most liberal criterion for inference sentences in the Big Changes text, which suggests these items were the least familiar. White participants set their criteria most liberally for the Orca text for both old and anxiety sentences. Refer to Figure 10.

There was also a marginally significant three-way interaction between the sentence type, participant race, and threat conditions F(3, 56) = 2.21, ρ = .09. A test of simple effects on the univariate analyses showed that there was no difference in the β for
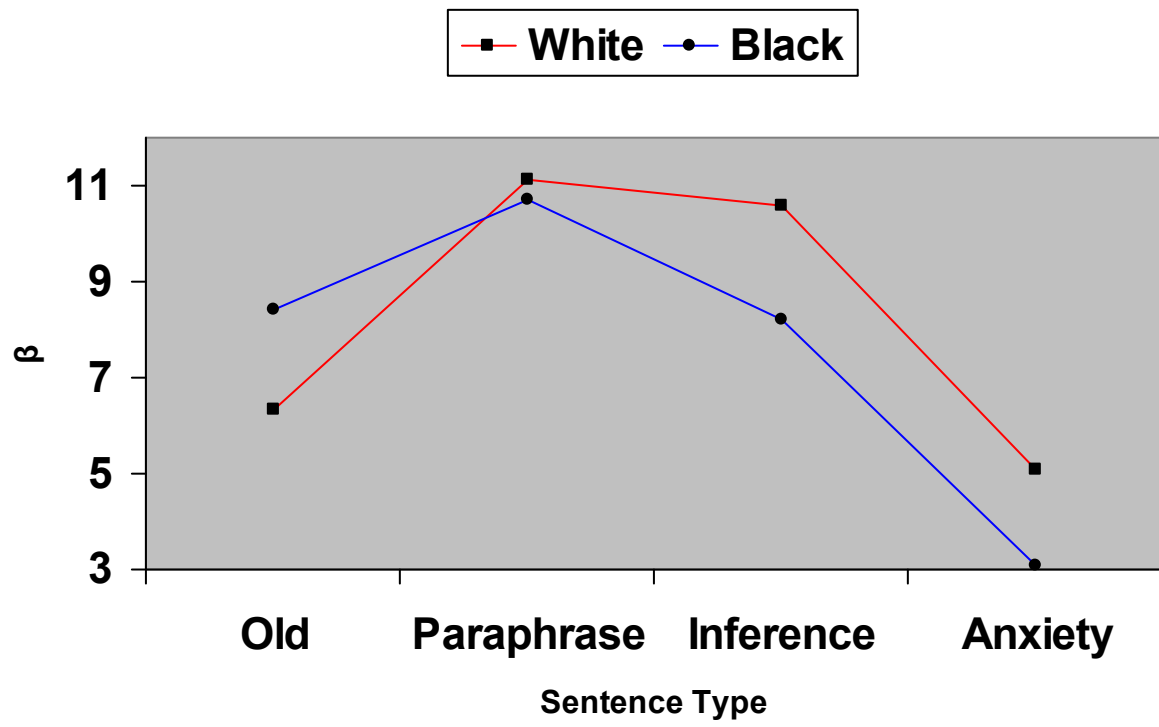
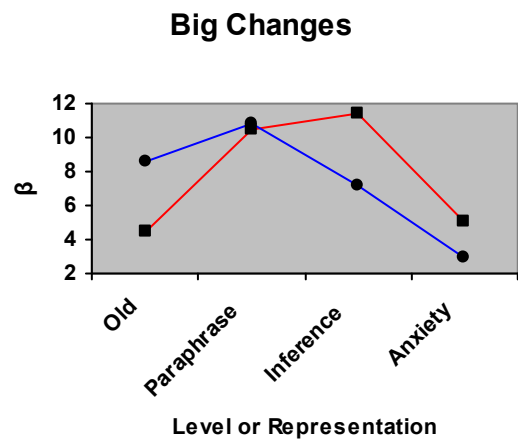*Fig. 9.* β as a function of sentence type and participant race.
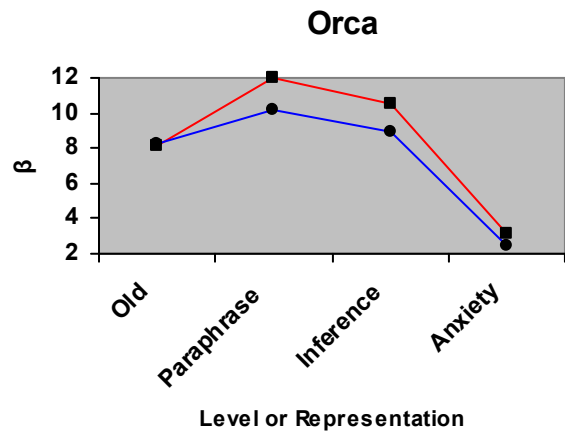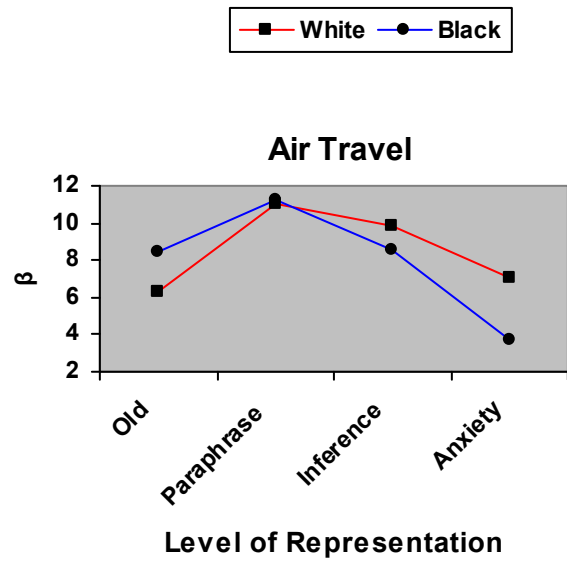
Figure 10. β as a function of sentence type, text type, and participant race.

Black students for any of the sentence types, regardless of threat condition. White participants operating under threat conditions set stricter criteria for old and paraphrase sentences, and less strict criteria for anxiety sentences. Thus, the White participants under threat conditions found anxiety sentences somewhat less familiar. Refer to Figure 11.

In summary, Black and White participants in this study appear to use different criterion placement for the test sentences, based on a variety of factors such as text difficulty and familiarity with the sentences. It does appear that participants shifted their criteria as needed in response to familiarity with the sentences. Black students set more stringent criteria for old sentences, focusing on identifying the precise wording of the sentences, regardless of threat condition. White students tended to adopt this same strategy only under the threat condition, with a more flexible approach used when not threatened. The β for the anxiety sentences provide no support for the hypothesis that, under threat conditions, anxiety intrudes and disrupts processing.

Figure 11. *β* as a function of sentence type, threat condition, and participant race.

CHAPTER 5

DISCUSSION

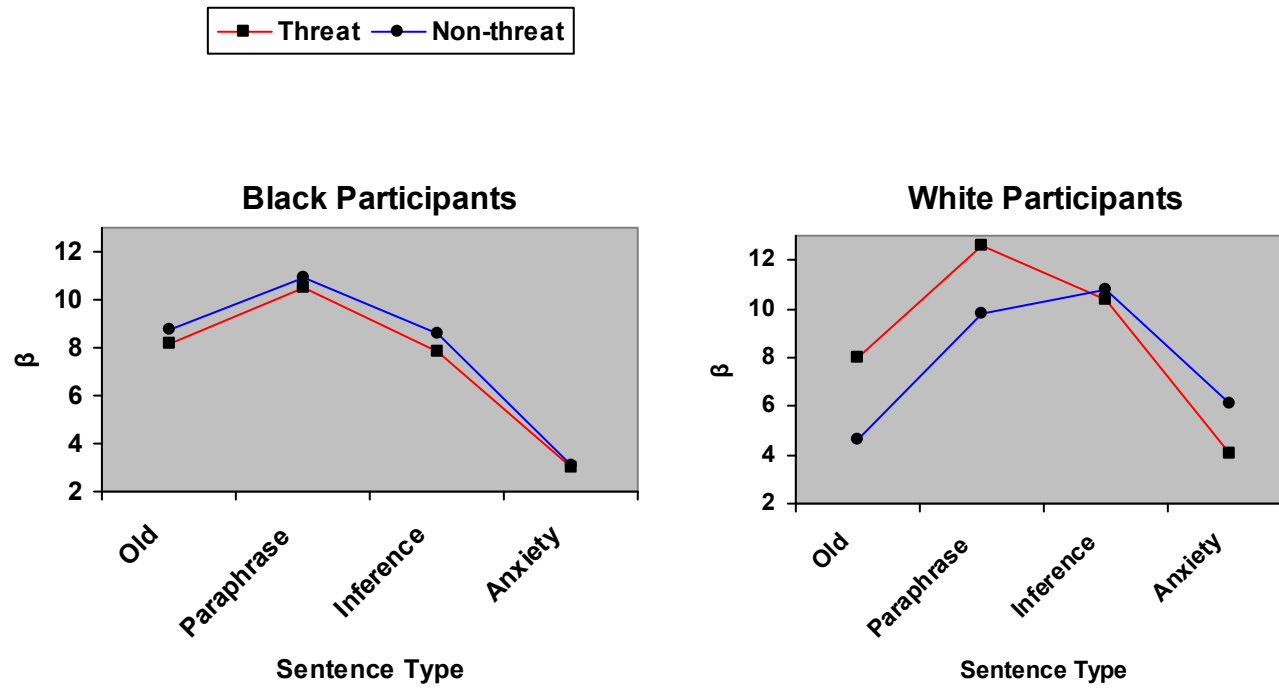Stereotype threat is a phenomenon that has received considerable attention in the recent literature. Beginning with Steele and Aronson (1995), researchers have explored various facets of stereotype threat, from the very existence of the phenomenon, to racial and gender differences, to the effect it has on individuals from various socioeconomic backgrounds. The research thus far has come in two waves. The initial wave focused on establishing the existence of a stereotype threat effect, while the second tested a number of factors influencing the effect, including several proposed hypotheses regarding the mechanism through which it acts. These proposals have largely been defined in vague and imprecise terms, such as a decline in processing capacity, or self-handicapping through expectation of failure.

The current project focused on four areas of concern. First, it was necessary to demonstrate once more the presence of a threat effect in the sample participating in this project. Secondly, the study attempted to demonstrate that there were no significant, pre-existing differences across race in terms of reading skills. The addition of an empirically derived covariate was included, in order to control for the effects of pre-existing differences in skill level for the reading task. Third, it sought evidence to support the presence of anxious interference with the task. Finally, the task assigned to participants has considerable ecological validity, and was selected to be one that is well understood. This allowed for the careful dissection of the data in order to examine precisely the source of potential differences in results.

Anxiety has played a role in many of the proposed causal mechanisms for stereotype threat (Schmader & Johns, 2003). The support for this proposal has been mixed, and the technique for measuring anxiety in participants has varied from self-report measures to assessment of physiological indicators for anxiety. Recent research into the nature of anxiety (Watkins, 2004; Barlow, 2004; Borkovec et al., 1998) supports a significant cognitive element to anxiety that fits well with the proposed mechanism investigated in this research project. That proposed mechanism of effect is the activation of stereotype relevant long-term working memory structures that interfere with the processing of task relevant information. The hypothesis suggests that the reading task is thus complicated by the activation of memory structures that contribute little to the understanding of the text. The decreased executive control, extensive memory networks regarding areas of anxiety, and ruminative quality of anxious thought contribute to the activation of this task-irrelevant information. That is, the participants operating under stereotype threat cannot help but activate these complex stored of knowledge pertaining to stereotypes and concern over performance. The addition of the anxiety sentences did not provide support for this anxiety activation hypothesis as a causal explanation, because the results showed that the only group to respond to the threat manipulation showed a lower familiarity for the anxiety sentences. Were anxiety a causal mechanism, it would be expected that individuals showing a reduced performance would be more familiar with sentences constructed to indicate anxious thought patterns. This could be the result of poorly developed anxiety sentences, the absence of anxiety in the readers, or a more general failure to understand the nature of anxiety's influence.

The NDRT was selected as a covariate to control for reading ability differences. The results of the NDRT indicated that the sample selected for this study had an average reading ability slightly lower than the normative sample for the NDRT (Brown, Fishco, & Hanna, 1993). This lower-than-average reading ability might explain some of the differences between these results and the results of prior research. The NDRT also showed that the White participants in this study had higher scores on Vocabulary and Comprehension, supporting the necessity of a covariate for this type of project. It is possible that stereotype threat may operate differently depending on the skill level of the reader. The general unwillingness of the Black participants to make inferences might be indicative of a role potentially being played by confidence differences. Because normal comprehension is dependent upon the ability to make inferences, it is possible that a lack of confidence in one's reading ability might impair comprehension through a reduction of such inferences.

The reading time data indicated that the White students were reading slightly more quickly than the Black students. Furthermore, the expository Orca text was read more slowly than the more narrative texts, regardless of the race of the reader. An interaction between text type and threat condition indicated that the Air Travel and Orca texts were read more slowly under the threat condition, regardless of the race of the reader. The combination of a difficult or unfamiliar text and the threat of having one's verbal intelligence assessed altered the reading behavior for the White participants. These individuals read more slowly under these conditions.

The recognition data indicated that the White participants made more "yes" responses than did their Black peers. There was a marginal threat/race interaction that

showed that the threat manipulation affected the performance of the White participants, but did not do so with the Black students, whose scores seemed to be at a consistently low level. That is, the Black students were more unwilling to endorse having seen a specific sentence before. This was especially true with the Orca text. The more difficult text, in combination with the threat effect, reduced the White participants' willingness to assert that they had seen a specific sentence before. This might well be an indication of different strategies on the parts of the White and the Black students to responding to the task. The expectation that long-term working memory would allow the Black students a greater access to stereotype relevant information was not supported. This would have been seen in the form of heightened responses on the part of the Black participants to anxiety-centered information on the stereotype-focused text ("Big Changes"), regardless of whether it was appropriate or not. This pattern was not noted.

The analysis of the d' data showed that White participants tended to have a stronger memory trace overall than their Black peers. When operating under threat conditions, especially on the more challenging texts, the White participants' trace strength dropped to the same level as the Black participants. Overall, the surface and situation model trace strengths were higher than the textbase for all participants. When the challenging Orca text was introduced, the racial differences emerged, and White participants were seen to have stronger situation models, while the Black participants had stronger surface level traces. It began to appear that the Black participants focused their attentions on identifying the exact wording of a sentence, while their White peers drew on prior knowledge of the subject matter. Both groups were equal regarding their performance on constructing a relatively weak textbase representation.

The β data further suggested the application of differing strategies. The participants utilized different criterion placement for the texts, depending on the text difficulty, threat, and past knowledge of the material. The Black students set more stringent criteria for old sentences, again suggesting that these individuals focused on identifying the precise wording of a particular item. The threat condition did not alter this strategy for the Black students, but it did force the White students to adopt the same strategy under the threat condition. When not threatened, the White students tended to have a more flexible approach, focusing on prior knowledge of the subject matter.

For both the β and the d' analyses, an error term was used to develop the relevant scores. The use of the inappropriate sentences as that error term resulted in a restricted range, which may have contributed to some of the patterns of results. With the universally very low response rates to these sentences, the β and d' data may have been better represented with the use of the anxiety inference sentences as the error rate.

The picture that emerges is one of a fundamental difference in the approach to the reading task for the Black and White students in this sample. The Black students drew on only one tool in their efforts to process the texts, focusing on exact wording. This use of only one strategy explain the pattern of increased time spent on the reading task (possibly indicating increased effort), coupled with a decreased performance on the task. A similar pattern was noted in Steele & Brown (1999). The White participants were better able to adapt their reading strategies to varying demands, and were able to shift to this surface level approach when threatened, but to use the situation model when not operating under threat. Seibt and Förster (2004) had also previously noted that individuals adapt their strategy for a task based on the presence of a stereotype threat, which was consistent with

93

the pattern of the Black participants approaching this experimental task differently than their White peers . Neither group made significant use of the textbase level of representation. This is supported by the overall lower than normative scores on the NDRT.

This project aimed to accomplish four tasks: demonstrate the stereotype threat effect, examine the role of anxiety in this effect, apply a covariate to stereotype threat research, and utilize a well-developed model of the task to identify where stereotype threat has its effect. The first task was accomplished, though not in the way predicted. One group of participants in this study did respond with altered performance to the threat manipulation, but it was not the Black students. This might be due to a floor effect, with the Black students in this sample performing at a level that was too low to allow for suppression with the addition of a threat manipulation. This raises the question of generalizability. It is possible that the sample of Black students participating in this project were especially unskilled readers, and thus were not representative of the wider population. It is also possible that, as described above, they could be operating under a continuous state of threat, given the academic setting in which the research was conducted. This is somewhat consistent with an idea put forth by Croizet & Claire (1998), who suggested that individuals from a lower SES background were functioning in a state of constant awareness of their stereotyped background.

By including anxiety based inference sentences in the task, it was hoped that a clearer picture of the role that anxiety plays in stereotype threat might be drawn. These sentences tended to attract little endorsement in the recognition task, suggesting that they

were not strongly activated in the processing of the texts. They do not seem to have played a large role in the strategies that participants used to approach the task.

The NDRT covariate proved to be a valuable addition to the project. It demonstrated a clear racial difference in reading ability for this sample. The Black students in this sample continually applied a strategy of focusing on the surface level of the text, while the White students resorted to this approach only when threatened. It is possible that the Black students in this sample were unable to utilize a different approach because they were in a constant state of threat. It is also possible however, that the initial results of the NDRT, in combination with the floor effect noted above, indicate that the Black students in this sample simply were poor readers, and had no other strategies to apply.

Finally, the application of the Construction Integration model for the texts was useful for several reasons. It provided the theoretical underpinnings for the inclusion of the various sentence types, which provided the basis for identifying the consistent pattern of Black students in this study focusing their attentions on the precise wording of the texts while White students adopted this strategy only under threat conditions. It also allowed for the analysis of how different types of texts (narrative versus expository) and different types of sentences were influenced by the threat effect. Without the Construction Integration model, this level of detail in the analyses would not have been possible.

This project leaves a great many questions yet unanswered, but it provides an approach to address these questions in future research. The limited sample size resulting from the small population of African American students resulted in a number of results

that only reached a marginal level of significance. It would be interesting to repeat the study, with some modifications, using a larger sample size. It would also perhaps be of use to determine if the sample of students used in this study differ significantly from the population at large. It was previously noted that this sample tended to fall below the average range of results on the NDRT, and there may well be other differences that were not immediately apparent. The issue of anxiety and its role in stereotype threat is as of yet unclear. Further study using different stimuli, developed to more precisely emulate an anxious rumination, may produce greater effects.

REFERENCES

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

Barlow, D. H. (2004). *Anxiety and its disorders: The nature and treatment of anxiety and panic.* New York, NY: Guilford Press.

Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science,* 12(3), 225-229.

Borkovec, T. D., Ray, W. J., & Stöber, J. (1998). Worry: A cognitive phenomenon intimately linked to affective, physiological, and interpersonal behavioral processes. *Cognitive Therapy and Research,* 22(6), 561-576.

Brown, J., Fishco, V., & Hanna, G. (1993). *The Nelson-Denny Reading Test*. Chicago, IL: Riverside Publishing.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology,* 33, 267-285.

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin,* 24(6), 588-594.

Dovidio, J. F., Kawakami, K. & Gaertner, S. L. (2000). Reducing contemporary prejudice: Combating explicit and implicit bias at the individual and intergroup levels. In S. Oskamp (Ed.), *Reducing Prejudice and Discrimination* (pp. 137-163). Hillsdale, NJ: Erlbaum.

Ericsson, A. K., & Delaney, P. F. (2003). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 257-297). Cambridge, England: Cambridge University Press.

Ericsson, K. A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review,* 102(2), 211-245.

Galambos, J. A. (1982). *Normative studies of six characteristics of our knowledge of common activities.* (Tech. Rep. No. 14). New Haven, CT:Yale University.

Hufnagel, B., Deming, G. L., Landato, J. M., & Hodari, A. K. (2004). The effect of stereotype threat on undergraduates in an introductory astronomy class. *Journal of Women and Minorities in Science and Engineering,* 10, 89-98.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review,* 95(2), 163-182.

Kintsch, W. (1998). The representation of knowledge in minds and machines. *International Journal of Psychology,* 33(6), 411-420.

Kintsch, W. & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review,* 92(1), 109-129.

Kintsch, W., Patel, V. L., & Ericsson K. A. (1999). The role of long-term working memory in text comprehension. *Psychologia,* 42, 186-198.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language,* 29(2), 133-159.

Mayer, D. M., & Hanges, P. J. (2003). Understanding the stereotype threat effect with

    "culture-free" tests: An examination of its mediators and measurement. *Human*

    Performance, 16(3), 207-230.

McFarland, L. A., Lev-Arey, D. M., & Ziegert, J. C. (2003). An examination of

    stereotype threat in a motivational context. *Human Performance,* 16(3), 181-205.

Nguyen, H. D., O'Neal, A., & Ryan, A. M. (2003). Relating test-taking attitudes and

    skills and stereotype threat effects to the racial gap in cognitive ability test

    performance. *Human Performance,* 16(3), 261-293.

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(1), 55-71.

Recht, D. R. & Leslie, L. (1988). Effects of prior knowledge on good and poor readers.

    Journal of Educational Psychology, 80, 16-20.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces

    working memory capacity. *Journal of Personality and Social Psychology,* 85(3),

    440-452.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a

    programmer's manual: Verbatim, propositional, and situational representations.

    Journal of Memory and Language, 25, 279-294.

Schneider, W., Korkel, J., & Weinert, F. E. (1989). Domain specific knowledge and

    memory performance: A comparison of high- and low-aptitude children. *Journal*

    of Educational Psychology, 81, 306-312.

Seibt, B., & Förster, J. (2004). Stereotype threat and performance: how self-stereotypes influence processing by inducing regulatory foci. *Journal of Personality and Social Psychology,* 87(1), 38-56.

Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology,* 75(5), 1191-1197.

Steele, C. M. (1999, August). Thin ice: "stereotype threat" and Black college students. *The Atlantic Monthly,* 44-54.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology,* 69(5), 797-811.

Steele, C. M., & Brown, J. B. (1999). When White men can't do math: Necessary and sufficient stereotype threat. *Journal of Experimental Social Psychology,* 35, 29-46.

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(6), 1213-1227.

Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory and Psychology,* 4(4), 525-547.

Walker, C. H. (1987). Relative importance of domain knowledge and over-all aptitude on acquisition of domain-related knowledge. *Cognition and Instruction,* 4, 25-42.

Watkins, E. (2004). Appraisals and strategies associated with rumination and worry. *Personality and Individual Differences,* 37, 679-694.

Zimny, S. T. (1987). *Recognition memory for sentences from a discourse.* Unpublished doctoral dissertation, University of Colorado, Boulder, CO.

Zwaan, R. A., & Truitt, T. P. (2000). Inhibition of smoking-related information in smokers and nonsmokers. *Experimental and Clinical Psychopharmacology,* 8(2), 192-197.

APPENDICES

Appendix A

"Practice Text"

You are currently reading the sample text. This text is designed to assist you in preparing to read the actual texts used in the experiment. As you can see, the process is quite simple. In a moment, the experimental texts will be presented in the same way as this practice text. Simply continue to read the texts in the same way as you have read this one. The experimenter would like to thank you for your participation in this study!

Appendix B

Air Travel Text and Test Sentences

"Are You Flying for Business or Pleasure?"

Kelly had flown many times in the past, but it had been quite a while since her last trip. Her company used to sponsor these trips fairly often for the salespeople who had proven their value. Kelly had certainly done so, and was considered to be at the top of her profession. But times had changed, and the company was not so eager to send its best employees into foreign lands anymore. The seminar in France this year however was too good to pass up, and the company felt that the opportunity to improve Kelly's skills and increase their business contacts was worth the expense. She herself had fought to get them to see things that way. Kelly was very highly motivated to improve her value to the company, and getting a chance to take in the sights that France had to offer was quite a perk. Kelly wondered how much down time she would have at the conference while she hurriedly packed her things.

Once at the airport, Kelly fell into the familiar routine of searching for the appropriate airline's terminal. Between the shops and restaurants, the airport had grown quite large and confusing. But Kelly was no novice here, and managed to find her goal relatively quickly. This proved to be a very good thing, because the security checks had gotten increasingly more complicated since the last time she had flown. Nevertheless, she knew that the elaborate security procedures and random bag searches had made flying safer, and that the delays in boarding were well worth the added comfort.

Waiting at the gate was always the least interesting part of travel for Kelly. She had flown often enough that she knew planes were often late, and the wait could be quite

dull. She passed her time thinking of the destination. She had attended several seminars in the past, throughout Europe and Asia. But Kelly had not yet been to Paris. Thinking back over all the things she knew or had heard about the country where she would be spending the next week, Kelly was both excited and a little nervous. She hoped that the little French she spoke would help to improve her enjoyment of the visit, but worried that she was not skilled enough in the language to get the full experience. To make herself a little more confident, she went into the gift shop to purchase a French – English dictionary and phrasebook.

Just as she left the shop, Kelly heard her flight number being called. She showed her boarding pass, and followed the slowly moving line into her plane. When Kelly sat down, she found her thoughts once more drifting back towards Paris. Although she was being sent to Paris to improve her work for her employers, Kelly knew from past experience that there would be plenty of opportunity to get away and enjoy the sights that the city had to offer. Kelly hoped that she could find enough free time to visit some of the museums and national monuments for which the French capital was famous. Even more than the sights however, Kelly was looking forward to the exotic foods for which France was most well known. Some of the foods seemed strange to Americans, but Kelly was adventurous and looked forward to sampling them.

When the flight attendant asked that everyone buckle their seatbelts in anticipation of take-off, Kelly was reminded of how safe flying had become. In fact, she knew that flying was generally considered to be safer than driving. As the plane taxied to the runway, Kelly scanned the cabin of the plane, wondering what brought the other passengers to the flight, and what plans they had upon arrival in France.

Old Sentences

- Kelly had flown many times in the past, but it had been quite a while since her last trip.

- Nevertheless, she knew that the elaborate security procedures and random bag searches had made flying safer, and that the delays in boarding were well worth the added comfort.

- Kelly hoped that she could find enough free time to visit some of the museums and national monuments for which the French capital was famous.

- Just as she left the shop, Kelly heard her flight number being called.

- When the flight attendant asked that everyone buckle their seatbelts in anticipation of take-off, Kelly was reminded of how safe flying had become.

Paraphrase Sentences

- In fact, she knew that driving was generally considered to be less safe than flying.

- While it had been quite a while since her last trip, Kelly had flown many times in the past.

- Kelly was both excited and a little nervous, thinking back over all the things she knew or had heard about the country where she would be spending the next week.

- She knew that the elaborate security procedures and random bag searches were worth the delays in boarding, having made flying much safer.

- Kelly hoped that she could find enough free time to visit some of the national monuments and museums for which the French capital was famous.

Inference model-correct

- Kelly was an important and valued member of the company for whom she worked.

- Kelly's past experiences in foreign countries had been positive, and made her feel

    confident that the current trip also be pleasant.

- The steps involved in traveling by airplane can be complicated, but were familiar to

    Kelly.

- Kelly is interested in learning new things, and having new experiences.

- The opportunity to see new places made traveling well worth any small inconveniences

    in Kelly's mind.

Inference model-stereotype

- Flying frightens a majority of those who must do so.

- The constant threat of terrorism is always a concern when flying.

- Even without concerns over possible terrorist activity, flying is an dangerous mode of

    travel.

- Visiting foreign lands is a scary prospect, since it is so difficult to adapt to their customs

    and ways.

- Kelly was uncomfortable and concerned for her safety throughout the course of her

    flight.

Appendix C

Orca Text and Test Sentences

"The Ocean's Social Giants"

Orcas, popularly known as killer whales, are not actually whales. These mammals inhabit every ocean on earth, and are the largest of the dolphin family. Orcas have no natural enemies. While some do kill great white sharks and whales, there is only one documented assault on a human. In this case the Orca grabbed a surfer, but quickly let go. Other Orcas have shown curiosity and friendliness towards humans, and are quite sociable giants.

Scientists have found that they can identify individual Orcas by their dorsal fins and markings. This knowledge has allowed the scientists to study the lives, deaths, and social interactions of individual Orcas, from the waters of Alaska to the southern California coast, for over 30 years now. They have made amazing observations.

Scientists study two communities of "Resident" Orcas in the northern and southern parts of the Pacific region. These two communities are different in many ways and do not interbreed, although they share some of the same ocean. There is no aggression between or within these communities. These Resident communities are comprised of clans of Orcas with common ancestors. Each Orca lives in a "pod" of 10 to 40 creatures, headed by the oldest female in the family. The pods contain several generations of that family, with individuals often living into their 90's. Their childhoods are lengthy, and they learn pod behavior from their parents and siblings. They largely eat salmon and other fish. The northern community is increasing in numbers, while the

southern residents are in decline. This perhaps reflects the changing quality of the fisheries in the two regions. Each community has behaviors that are not shared with the other community. For example, the northern Orcas like to rub their bodies along the pebbles of the shallow oceans. The southern community is more likely to erupt into the aerial displays that delight people at Sea World. Each community also communicates with a different complex array of sounds. Scientists have found that each pod has its own dialect, not unlike the distinctions in human language. The more distant one pod is from another genetically, the more different the dialects. Different communities do not seem to communicate with each other.

In addition to the Orca communities, scientists have discovered two other types of pods in this region. "Transients" live in smaller groups of 2 to 6, are less noisy than residents, and eat mammals such as seals. In fact, Transient Orcas have been known to starve to death when only offered fish as food. Their prey is able to distinguish between Resident sounds and Transient sounds, apparently being terrified by the latter. Transients have sophisticated and coordinated hunting techniques. They quietly stalk mammals, use deception, and wash animals into the water by flapping their tails, creating waves of water to push their prey into the ocean. They often kill their prey by ramming it and holding it underwater. Transients have been known to transmit information about a kill to other Transients over distances of 17 miles. DNA evidence shows that Transients and Residents have not interbred for over 10,000 years.

Less is known about the "Offshore" Orcas. These live in large groups, are only occasionally seen, stay far out in the ocean, and probably primarily eat shark.

As scientists learn more about these masters of the ocean, they are discovering a complex social structure, with close and enduring family ties, different communication dialects, and distinctive behavioral and eating patterns. They have even observed different personalities within a pod. Some Orcas are more curious than their family members and seem to enjoy interactions with humans more. Scientists speculate about what other kinds of animals Orcas resemble. Are they like lions, or wolves? An emerging view is that they are most like humans, with different societies, tribes, cultures, dialects, and behaviors.

Test Sentences

Old Sentences

- Orcas have no natural enemies.

- There is no aggression between or within these communities.

- As scientists learn more about these masters of the ocean, they are discovering a
    complex social structure, with close and enduring family ties, different
    communication dialects, and distinctive behavioral and eating patterns.

- Each community has behaviors that are not shared with the other community.

- Transients have sophisticated and coordinated hunting techniques.

Paraphrase Sentences

- These mammals are the largest member of the dolphin family, and are found in every
    ocean on earth.

- Scientists have found that dorsal fins and markings allow them to identify individual

    Orcas.

- DNA evidence shows that, for at least 10,000 years, Transients and Residents have not

    interbred.

- "Transients" eat mammals such as seals and dolphins, are less noisy than residents, and

    also live in smaller groups.

- Some Orcas are more curious and seem to enjoy interactions with humans more than

    others within the same family.

Inference model-correct

- Before scientists knew they could identify individual Orcas, they couldn't collect

    information about their social groups.

- Many of the distinct Orca behaviors are not inborn but learned during childhood.

- One of the main characteristics of Orca groups that dictate which other Orcas they

    interact with is their Orca language.

- Less is known about Offshore Orcas because it is difficult for scientists to find and

    follow them.

- Orca social life is nearly as complex and differentiated as human social life.

Inference model-stereotype

- Swimmers have to be cautious when Orcas are around, because the whales have often

    been noted to be aggressive.

- Because Orcas will eat nearly anything, surfers and swimmers must avoid the waters

    where these whales are present.

- The danger presented by killer whales is increased by the fact that they hunt in groups.

- Scientists are fearful that Orcas pose a serious threat to water-sports enthusiasts.

- The nickname "killer whale" was developed by anxious lifeguards, whose jobs often

    bring them close to these dangerous creatures.

Appendix D

Stereotype Text and Test Sentences

"Big Changes"

Marcus and Tyra spoke about their recent experiences as they walked to class in the morning. They had been friends for years, having grown up in the same apartment building. That friendship was a source of comfort when they headed out for college. They needed that familiarity, because so much had changed over the past few weeks. In the city, they had felt at home. Out here in the country, in the small town where their university was located, they felt like they were on a different planet. And this was a planet with which they were not entirely comfortable.

Back home, just about anything they could want was within walking distance. Even when they had to go further than a few blocks, the buses and subway systems were reliable and steady. At school though, the buses were unreliable, and there was no subway at all. Of course, as Marcus was fond of pointing out, there wasn't much reason to go very far from campus in the first place. The local entertainment was pretty limited. A few bars and one movie theater were all that the town offered. A drive of nearly two hours was necessary to get to anything resembling a city. Even the local television channels usually skimmed over the events which occurred in the city where the two friends grew up. Marcus hated that there wasn't even a decent sports team in the area.

And the local residents weren't all that friendly either. Walking down the streets left them both feeling almost like aliens. It was clear the locals were uncomfortable around them, and the feeling was often mutual. The hostility was pretty quiet, but they could both tell that kids from the city were not entirely welcome. Tyra talked about the

way that their style of dress was frowned upon. They both had noticed how the locals tended to frown when they heard the two friends spoke to each other. It was like the townspeople felt superior to those from less rural areas.

The students were somewhat better, but a lot of them seemed to be from similar small towns, and to have similar small-town attitudes. The two friends spoke about how uncomfortable it was to be around these students, who seemed to be judging the pair constantly. At least on campus, Marcus and Tyra were able to find a number of people who were like themselves. These new friends shared their interests and their background, with most of them having grown up in a city as well. Some of their new friends, having been in the area for a while, helped the pair pick out the more pleasant, welcoming places to eat. These friends also helped with adjusting in other, unexpected ways. Even relatively simple things had gotten more complicated. It was surprising how difficult it was to find a decent barber or hairstylist. None of the local shops seemed to be able to do a very good job. Tyra in particular had struggled to find a decent stylist with whom she felt comfortable.

But this school was one of the few options available in the state. Tuition wasn't terribly expensive, and the university offered decent financial assistance and scholarships. Most of the professors were helpful, though Marcus often got the impression that he was being judged more harshly than many of the other students. The classes were certainly much harder than they had been in high school. The level of work that was expected of them had drastically increased. They were spending more time on homework than ever before, and studying was a full-time job. As they approached their classroom, they both

focused their thoughts on the upcoming exam, and the questions they could expect to be asked.

Test Sentences

Old Sentences

- They had been friends for years, having grown up in the same apartment building.

- At least on campus, Marcus and Tyra were able to find a number of people who were like themselves.

- Walking down the streets left them both feeling almost like aliens.

- Tuition wasn't terribly expensive, and the university offered decent financial assistance and scholarships.

- Of course, as Marcus was fond of pointing out, there wasn't much reason to go very far from campus in the first place.

Paraphrase Sentences

- Out in the in the small country town where their university was located, they felt like they were on a different planet.

- A movie theater and a few small bars were all that the town offered for entertainment.

- It was clear the friends were uncomfortable around the locals, and the feeling was often mutual.

- These new friends shared their background and hobbies, and had mostly grown up in a city as well.

- They were spending more time studying than ever before, and homework was like a full-time job.

Inference model-correct

- Marcus and Tyra grew up in an urban area, far from the counryside.

- The two friends relied on each other for support during their difficult transition to college life.

- The locals seemed to always be suspicious around people different from themselves, including Marcus and Tyra.

- The student population was more diverse than that of the townspeople, which was a relief to the two friends.

- Marcus and Tyra had to spend a great deal of time studying, because the work was harder than it had been in high school.

Inference model-stereotype

- Both Marcus and Tyra were worried about mid-terms, which would be starting soon.

- The two friends became very nervous when the local police started to watch them very closely.

- Marcus was anxious about his chances of getting into graduate school after he completed his coursework.

- Tyra knew that the high school she attended was not as good as those her current classmates came from.

- The friends feared that they could not keep up with their classmates academically.

Appendix E

Instructions to Participants in the Control Condition

"Instructions"

Thank you for agreeing to participate in this study. The purpose of this experiment is to test a new type of text book. Manufacturers of text books are looking to utilize a computerized format, in order to do away with the traditional style of books, which use a great deal of paper, and cannot be altered once they are printed. In order to move on to the new, computerized format, the manufacturers must first be certain that reading from a computer screen is easy to do, can be accomplished reasonably quickly, and allows the student to retain as much information as the traditional, paper books with which you are familiar.

You will be asked to read a series of four short stories. The first story is a practice item, and is designed to familiarize you with the way in which the computer presents text. After reading each line, press the space bar to move onto the next line. When the practice item is completed, you will see a title screen for the first item of the experiment. Again, use the space bar to advance the story. When you finish each text, you will be asked to read a sentence, and determine whether or not that EXACT SENTENCE was part of the previous story. If it was, you will press the "F" key. If that EXACT SENTENCE was not a part of the story, you will press the "J" key. The instructions to respond to the sentences will be presented onscreen, so you will not have to remember which key means "yes" and which means "no." You will then be asked to proceed to the next story, and repeat the procedure. Do you have any questions before we begin?

Appendix F

Instructions to Participants in the Experimental Condition

"Instructions"

Thank you for agreeing to participate in this study. The purpose of this experiment is to test verbal intelligence, an ability which is strongly related to success in school. In order to accomplish this, the computer will record the amount of time it takes to read the following stories. A brief memory/comprehension measure will also follow each text.

You will be asked to read a series of four short stories. The first story is a practice item, and is designed to familiarize you with the way in which the computer presents text. After reading each line, press the space bar to move onto the next line. When the practice item is completed, you will see a title screen for the first item of the experiment. Again, use the space bar to advance the story. When you finish each text, you will be asked to read a sentence, and determine whether or not that EXACT SENTENCE was part of the previous story. If it was, you will press the "F" key. If that EXACT SENTENCE was not a part of the story, you will press the "J" key. The instructions to respond to the sentences will be presented onscreen, so you will not have to remember which key means "yes" and which means "no." You will then be asked to proceed to the next story, and repeat the procedure. Do you have any questions before we begin?

Demographic Questionnaire

<u>Demographics Questionnaire</u>

This is a brief questionnaire requesting some background information on the demographic composition of the current group of Psychology 101 students. Please take a moment to complete this survey by filling in the blanks or circling the appropriate response, and return it to the course instructor. Thank you for your participation.

Banner ID #_____   Today's Date_____

1) What is your year in school?

      Freshman   Sophomore   Junior   Senior

2) How do you identify yourself racially?

      African American

      Asian American

      Bi/Multi-racial

      Caucasian American

Latino American

Native American

Other:_____

3) What is your gender?

Female

Male

4) How would you describe the area where you spent the majority of your childhood?

Rural

Small Town

Small City

Medium City

Large City

Table 1

Summary of Memory Traces in Proportion "Yes" Recognition Responses

	Verbatim	Textbase	Model correct	Model stereotype
Old	X	X	X	
Paraphrase		X	X	
Inference model-correct			X	
Inference model-stereotype				X
Inappropriate				