The Dissertation Committee for Lisa Renee Grimm Narvaez certifies that this is the approved version of the following dissertation:

Stereotype Threat Reinterpreted as a Regulatory Fit

Committee:

Arthur B. Markman, Supervisor

William T. Maddox

Robert A. Josephs

David M. Schnyer

Rajagopal Raghunathan

Stereotype Threat Reinterpreted as a Regulatory Fit

by

Lisa Renee Grimm Narvaez, B.A., M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2007

Acknowledgements

I deeply appreciate advice on this project from Arthur B. Markman and W. Todd Maddox and the insightful comments from Robert A Josephs, Rajagopal Raghunathan, David M. Schnyer, and Brian J. Stankiewicz. I am grateful for the contribution of Grant Baldwin who helped at the start this project and patiently gave programming guidance, and to Darrell Worthy who provided modeling assistance. I thank Benjamin Narvaez for his encouragement, and Jonathan Rein, Jeff Laux, and Tanya Feinstein for their continuing friendship and thoughtful comments on this and related research, and Kyler Eastman, Micah Goldwater, Brian Glass, and Serge Blok for valuable conversations. Lastly, this dissertation research could not have been completed without Leland Lockhart who tirelessly recruited and ran participants, and my wonderful research assistants: Chaz Bryant, Yi-Ting Cheng, Tram Dinh, Jared Douthit, Kourtneigh Forster, Francesca Fraga, and Sarah Voges.

Stereotype Threat Reinterpreted as a Regulatory Fit

Publication No. ____________

Lisa Renee Grimm Narvaez, Ph.D.

The University of Texas at Austin, 2007

Supervisor: Arthur B. Markman

Starting with Steele and Aronson (1995), research documents the performance decrements resulting from the activation of a negative task-relevant stereotype. I suggest that negative stereotypes can generate better performance, as they produce a prevention focus (Higgins, 2000; Seibt & Förster, 2004), because a prevention focus leads to greater cognitive flexibility in a task where points are lost (Maddox, Markman, & Baldwin, 2006). My prior work, Experiments 1 and 2, done in collaboration with Arthur B. Markman, W. Todd Maddox, and Grant C. Baldwin, used a category learning task that requires the participant test different explicit rules to correctly categorize stimuli. Half of the participants gained points for correct responses while half of the participants lost points for correct responses. We primed a positive or a negative gender stereotype. The negative prime matches the losses environment while the positive prime matches the gains environment. The match states are assumed to increase dopamine release into frontal brain areas leading to increased cognitive flexibility and better task performance

whereas the mismatch states should not. Thus, we predict and obtain a 3-way interaction between Stereotype (Positive, Negative), Gender (Male, Female), and Reward structure (Gains, Losses) for accuracy and strategy.

Experiments 3 and 4 used a category learning task, which requires the implicit learning system to govern participant responses. This task had an information-integration category structure and involves the striatum (e.g., Maddox & Ashby, 2004). Importantly, cognitive flexibility will hurt performance using this category structure. I therefore predicted that regulatory match states, created by manipulating Stereotype and Reward structure, will produce worse performance than mismatch states. I did not completely reverse the effects described in Experiments 1 and 2 as predicted. I found evidence supporting my predictions using computational models to test for task strategy in Experiment 3 and found results consistent with the flexibility hypothesis in Experiment 4. Importantly, I believe that stereotype threat effects should not be conceptualized as a main effect with negative stereotypes producing worse performance than positive stereotypes, but instead as an interaction between the motivational state of the individual, task environment, and type of task performed.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Usually cognitive psychologists reside in their ivory tower brimming to the top with disembodied symbols. These symbols combine to become mental representations that are the building blocks of thought. Although it is clearly a worthy and very challenging pursuit to determine the structure of mental representations and the processes that act on them, I believe cognitive psychologists often fall short in their efforts to explain cognition because they neglect a critical piece of the puzzle –motivation. Without understanding why people act, what goals they are trying to meet, and how the motivational system is engaged to promote goal attainment, cognitive psychology will fail to provide an appropriate model of cognitive processing.

My research directly examines the motivational factors that drive behavior. These factors are always present. Consider the undergraduate research participant trying desperately to complete the research requirement in the last week of the semester. This individual is trying to avoid the very negative outcome of receiving an incomplete in a course. It would be naïve to believe that this individual enters a psychology experiment, turns off her motivational system, and performs a typical cognitive task without any motivational arousal.

In the service of understanding the role of the motivational system, my dissertation examines the influence of the motivational system on cognitive processing. Often ignored by cognitive psychologists, the motivational system critically influences cognition. For example, as demonstrated by Markman, Maddox, and colleagues (Markman, Baldwin, & Maddox, 2005), very subtle manipulations of regulatory focus

(Higgins, 1997b), a motivational variable, can produce drastically different patterns of responses depending on whether a classification task is framed in terms of gains or in terms of losses. While this result may seem unremarkable, the potential ramifications of this and related work are profound. As they argue, classic psychological research tends to require participants to acquire points, get correct answers, or, at the largest grain size, gain course credit. As such, individuals predisposed to deal with an environment of gains will perform better than an individual predisposed to deal with an environment of losses.

I use this framework to investigate a possible cause of stereotype threat effects. Stereotypes impact human cognition and are a pervasive part of human experience. Starting with Steele and Aronson (1995), research documents the negative impact on performance given the activation of a negative task-relevant stereotype. In their studies, Black participants underperformed White participants on tests of intellectual ability when the test was framed as diagnostic of their ability. These decrements occur in a range of domains from the academic sector to athletic performance and are known as *stereotype threat* effects (Aronson, Lustina, Good, Keough, & Steele, 1999; Stone, Lynch, Sjomeling, & Darley, 1999).

There have been many posited stereotype threat mechanisms. For example, stereotype threat effects may occur because of too much effort or too little effort, self-handicapping, anxiety, or low performance confidence (Smith, 2004). Seibt and Förster (2004) argue that activating stereotypes induces regulatory foci, which in turn influence performance. They demonstrate that a negative stereotype induces a prevention focus while a positive stereotype induces a promotion focus. Critically, they frame the influence of regulatory foci as an interaction of focus and processing requirements: a promotion focus induced by a positive stereotype leads to better performance and a prevention focus induced by a negative stereotype leads to worse performance if elaborative processing is required. Conversely, negative stereotypes lead to better performance than positive stereotypes when vigilant processing is required.

In contrast, I argue that stereotype threat effects occur because of the interaction between the induced regulatory focus, the reward structure of the task, and the type of task. Seibt and Förster's account fails to consider the effects of regulatory fit produced from focus and task reward structure. As demonstrated by Maddox, Markman, and colleagues (Maddox, Baldwin, & Markman, 2006; Maddox, Markman, & Baldwin, 2007; Markman, Baldwin et al., 2005; Markman, Maddox, & Baldwin, 2005; Markman, Maddox, & Worthy, 2006), the reward structure of the task interacts with the induced regulatory focus. Given a task that involves gains and non-gains, individuals with a promotion focus are experiencing a *regulatory match* while individuals with a prevention focus are experiencing a *regulatory mismatch.* Likewise, given a task that involves losses and non-losses, individuals with a promotion focus are experiencing a *regulatory mismatch* and prevention-focused participants are experiencing a *regulatory match.*

For example, assume that there exists a negative stereotype for women and a positive stereotype for men in mathematics. When confronted with a standard math test, prevention-focused women and promotion-focused men are trying to get test problems correct. That is, they are trying to gain points. Using the regulatory fit framework, women are in a mismatch and should perform poorly while men are in a match and should perform well. In contrast, altering the test goal to emphasize avoiding incorrect answers changes the test structure from gains to losses. This reverses the predicted gender effect because the fit pairings change; women are now in a match and men in a mismatch.

Further, I argue that the stereotype threat effects in the literature are likely the result of prevention-focused participants completing gains/non-gains tasks that require flexible processing. This argument relies on the presence of a three-way interaction the regulatory focus of the individual, the reward structure of the task, and the type of task. Individuals can be in a regulatory match or mismatch based on how their primed or chronic focus corresponds to the reward structure of the task. In addition, the influence of the match or mismatch will vary by task. The match states are assumed to increase dopamine release into frontal brain areas leading to increased cognitive flexibility whereas the mismatch states should not. In my tasks, cognitive flexibility is defined as a persistence to use the explicit processing system in favor of the implicit processing system. This increased cognitive flexibility should lead to better performance on tasks that require flexibility, like an explicit rule-based categorization task, and worse performance on tasks where flexibility is detrimental, like an implicit categorization task.

To explore these relationships, my dissertation experiments manipulate stereotypes, the task reward structure, and the type of task. In Experiments 1 and 2, I use a category learning task that requires the participant test different explicit rules to correctly categorize stimuli. This task is assumed to require cognitive flexibility. In Experiments 3 and 4, I use a task that requires the implicit learning system and is assumed to be disrupted if flexible processing is employed. In all experiments, half of the participants gain points for correct responses while half of the participants lose points for correct responses. I prime a positive or a negative gender stereotype. The negative prime matches the losses environment while the positive prime matches the gains environment. These matches should lead to better performance in Experiments 1 and 2 and worse performance in Experiments 3 and 4 because of the increased cognitive flexibility afforded by a match.

To make this argument, in Chapter 2, I describe stereotype threat and previous explanations for this phenomenon. In Chapter 3, I review the literature on regulatory focus and demonstrate the cognitive implications of regulatory fit.

In Chapter 4, I describe experiments that reinterpret stereotype threat effects as a four-way regulatory fit interaction. I used a category learning task that requires the participant test different explicit rules to correctly categorize stimuli or a task that penalizes participants for persisting with explicit testing in favor of the implicit learning system. This latter task used an information-integration category structure (e.g., Maddox & Ashby, 2004). Importantly, cognitive flexibility will hurt performance using this category structure because of a persistence to explicitly test classification rules. Half of the participants gained points for correct responses while half lost points for correct responses. I primed a positive or a negative gender stereotype. In Experiments 1 and 2, I predicted and obtained a 3-way interaction between Stereotype (Positive, Negative), Gender (Male, Female), and Reward structure (Gains, Losses) for accuracy and strategy. In Experiments 3 and 4, I found data consistent with a 3-way interaction for strategy but not accuracy. My results in Experiment 4 are stronger than my results in Experiment 3.

In Chapter 5, I reexamine several stereotype threat studies discussed in Chapter 2. I consider how the regulatory fit framework would account for the data and what other predictions the framework affords. I highlight how to accommodate other explanations for stereotype threat given my data in Chapter 4. Further, I discuss some possible neural mechanisms responsible for regulatory fit effects and some related individual difference variables to be considered in future work. I end with a discussion of some practical implications of my approach and present data from a follow-up experiment using a chronic stereotype (e.g., women are bad at math) and GRE math problems. Participants gained or lost points. I found that women performed better in the losses task than in the gains task and men performed better in the gains task than in the losses task. Thus, I demonstrate experimentally that regulatory fit influences standardized test performance.

Chapter 2: Stereotype Threat

DEFINITION OF STEREOTYPE THREAT

Stereotypes are an omnipresent part of human psychological experience. Even the well-educated rely on stereotypes to form impressions of social groups. For example, Lawrence Summers, former President of Harvard, claimed that women are innately deficient at mathematics and therefore are less-able scientists. What psychological consequences do women suffer as a result of this negative stereotype? How does this experience generalize to any member of a negatively-stereotyped group?

Starting with Steele and Aronson (1995), research documents the performance decrements resulting from the activation of a negative task-relevant stereotype. These decrements occur in a range of domains from the academic sector to athletic performance and are known as *stereotype threat* effects (Aronson et al., 1999; Spencer, Steele, & Quinn, 1999; Steele, 1997; Stone et al., 1999). Stereotype threat effects are extremely common in the literature (Wheeler & Petty, 2001). In a 2001 review, Wheeler and Petty found that 15 of 16 studies manipulating a negative stereotype found behavior consistent with the negative stereotype. These performance decrements are even possible for groups typically not stigmatized, like White men (Aronson et al., 1999; Leyens, Desert, Croizet, & Darcis, 2000; Stone et al., 1999).

Not confined to laboratory settings or specific populations as previously thought (Whaley, 1998), stereotype threat effects can be found in real-world contexts (Cohen, Garcia, Apfel, & Master, 2006; Keller & Dauenheimer, 2003; Steele, James, & Barnett, 2002). Steele, James, and Barnett (2002) demonstrated that women in male-dominated fields, such as math and engineering, are more likely than those in female-dominated fields to think about changing their major. They propose that this difference suggests women are avoiding the possibility of confirming a negative stereotype by switching into fields like the social sciences that do not have negative stereotypes for women (see Davies, Spencer, Quinn, & Gerhardstein, 2002 for a laboratory demonstration).

Similarly, individuals experiencing stereotype threat will avoid confirming group membership (McFarland, Lev-Arey, & Ziegert, 2003; Steele & Aronson, 1995) or will highlight unique self-characteristics to avoid the threat state (Croizet, Desert, Dutrevis, & Leyens, 2001). In fact, emphasizing intellectual strengths prior to a test framed as diagnostic of intellectual abilities increased the performance of women experiencing stereotype threat (Croizet et al., 2001).

STEREOTYPE THREAT EFFECTS

In the first experiments on this topic, Steele and Aronson (1995) investigated the influence of a negative racial stereotype on intellectual performance. In their studies, Black participants underperformed White participants on tests of intellectual ability when the test was framed as diagnostic of their ability. In Experiment 1, Steele and Aronson gave their participants a 30-minute test from the Verbal GRE. In the stereotype threat condition the test was described as diagnostic of intellectual ability. This was thought to make the negative stereotype about Black participants' intellect salient making them worried about confirming the stereotype. In two control conditions, a non-diagnostic condition and a challenge condition, the test was described as examining psychological factors involved in verbal problem solving or as difficult for even individuals with excellent verbal skills, respectively. Black participants underperformed relative to White participants in the diagnostic condition only.

This type of paradigm can be applied generally when groups have task-relevant negative stereotypes. Using a clever design, Stone, Lynch, Sjomeling, and Darley (1999) examined the performance of Black and White participants in a golf task. Their participants completed a golf course and the number of strokes required to finish the course was recorded. They manipulated the framing of the task between subjects. The task was framed as either diagnostic of “sports intelligence” or “natural athletic ability” or “general sports performance.” Based on culturally-known stereotypes, the sports intelligence condition should prime a negative stereotype for the Black participants while the natural ability condition should prime a negative stereotype for the White participants. The general sports performance condition acted as a control condition. Stone et al. found that Black participants performed worse than the control condition when the golf task was framed as diagnostic of “sports intelligence” but better than the control if the task was framed as diagnostic of “natural athletic ability.” In contrast, White participants performed worse than control when the task was framed as diagnostic of “natural athletic ability.”

In another domain, Aronson, Lustina, Good, and Keough (1999) examined math performance of White men. Participants in the stereotype threat condition were told that the purpose of the study was to examine the mathematical superiority of Asians and were asked to read a series of articles on Asians excelling at math. In the control condition, participants were told nothing. Aronson et al. found that men in the stereotype threat condition scored worse on a math test as compared to men in the control group.

EXPLANATIONS FOR STEREOTYPE THREAT

Researchers have manipulated stereotype threat in a number of ways. The most subtle manipulation involves participants merely noting their race on a test form or as part of a demographic questionnaire prior to the test (Steele & Aronson, 1995). Other researchers rely on framing the test as diagnostic of ability, where the ability is thought to prime a negative stereotype for a particular group. The strongest manipulation of stereotype threat involves telling participants that another group, specifically the participants' out-group, out-performs their in-group.

So why does stereotype threat occur? A given stereotype could be self-relevant or other-relevant and could be interpreted as positive or negative (Wheeler & Petty, 2001). Fundamentally, the psychological environment needs to afford stereotype-consistent behavior. That is, the activated stereotype needs to be self-relevant (Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Davies et al., 2002) and the environment needs to allow for stereotype confirmation in that the stereotype should be applicable (Ben-Zeev, Fein, & Inzlicht, 2005; Spencer et al., 1999). For example, Inzlicht and Ben-Zeev (2003) argue that women in mixed-gender environments show more stereotype threat consistent behavior than women in same-gender settings.

Based on the importance and real-world applications of stereotype threat (Keller & Dauenheimer, 2003), there have been many posited mechanisms. For example, stereotype threat effects may occur because of too much effort or too little effort, self-handicapping, anxiety, or low performance confidence (Cadinu et al., 2003; Smith, 2004). For example, Cadinu et al. (2003) argue that stereotype threat effects occur because of lower performance expectancies. That is, the lower the expected level of performance, the lower the actual performance. As another example, Schmader, Johns, and Barquissau (2004) provide behavioral data differentiating individuals based on stereotype endorsement. In two studies, they compared a control group of women with women who endorsed the stereotype that men are better at math than women. Stereotype endorsement led to decreased confidence in learning new material, lower domain self-esteem, less desire to continue on in related careers, and poorer performance on a math test.

Brown and Josephs (1999) demonstrate that math performance differences can be attributed to task-specific concerns. They predicted and obtained support for the claim that women are concerned with confirming a negative stereotype while men are concerned with confirming a positive stereotype. In a clever design, Brown and Josephs framed their test as diagnostic of weak ability or strong ability. Women underperformed in the weak ability condition relative to women in the strong ability condition and men showed the reverse effect.

Ideomotor theorists argue that stereotype threat effects occur because of a connection between the represented stereotype and the corresponding stereotypic behavior (Bargh, Chen, & Burrows, 1996; Wheeler & Petty, 2001). That is, activation of the stereotype activates stereotype-consistent behavior. For example, Bargh et al. demonstrate that priming an elderly stereotype caused participants to walk more slowly down a hallway than unprimed participants.

While interesting, this theory may be unable to explain *stereotype lift* effects documented by Walton and Cohen (2003). Stereotype lift occurs when individuals in the non-stereotyped group show an increase in their performance relative to control participants. If the stereotype threat mechanism involved direct activation of stereotypic behaviors given the activation of a stereotype, the stereotype does not need to be self-relevant to influence performance. Given that both negative and positive stereotypes tend to be primed, all participants should have all stereotypic behaviors activated. At a minimum, ideomotor theories need to explain the selection of stereotype-consistent behaviors and as such seem to gain little ground as compared to other relevance-related theories.

There are a couple different motivational accounts of stereotype threat. First, stereotype threat is conceptualized as activation and inhibition of specific stereotypes based on active goals (Fein, von Hippel, & Spencer, 1999; Sinclair & Kunda, 1999). For example, Sinclair and Kunda (1999) gave participants positive or negative task feedback from a Black 'doctor'. Given positive feedback, participants more quickly identified words associated with a doctor stereotype but given negative feedback participants more quickly identified words associated with a Black stereotype. This demonstrates that a desire to view an individual in a specific way can interact with the activated stereotypic information. However, as Fein et al. (1999) note, based on the wealth of other data it seems unlikely that this mechanism accounts for the effects. People often seem unable to inhibit negative stereotypes, even given an active desire, and thus show the corresponding performance decrements.

Second, arousal theorists (see Brehm & Self, 1989 for a general discussion on the role of arousal) believe that stereotype threat increases system arousal affecting performance on difficult but not on easy tasks (Ben-Zeev et al., 2005; O'Brien & Crandall, 2003). From a stereotype threat framework, Wheeler and Petty (2001) argue that stereotypes are more relevant for difficult tasks as compared to easy tasks. However, arousal theorists argue that tasks do not need to be stereotype-relevant because an increase in arousal should influence general task performance. For example, Ben-Zeev et al. tested women on an easy writing task in which they wrote their name in cursive for 20 seconds and on a difficult writing task in which they wrote their name backwards for 20 seconds. Participants primed with a negative math stereotype underperformed control participants in the difficult task but outperformed the control in the easy task.

Further evidence comes from work on eradicating stereotype threat by providing obvious situational attributions for performance decrements (Brown & Josephs, 1999; Johns, Schmader, & Martens, 2005) or reducing anxiety (Cohen et al., 2006). For example, Brown and Josephs (1999) examined the influence of providing an external handicap. In Study 2, half of the participants were told that they could not complete practice math problems prior to a test because of computer failure. Women in this condition performed better as compared to control women. Using a more direct manipulation, Johns et al. (2005) taught a group of women about stereotype threat and found that math performance increased for this group. In a similar study, Cohen, Garcia, Apfel, and Master (2006) improved performance by African American students by reducing stress using self-affirmation techniques.

A more parsimonious account of these arousal effects comes from the working memory literature (Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Schmader & Johns, 2003). Schmader and Johns (2003) recently argued that stereotype threat effects are mediated by working memory capacity. Beilock, Jellison, Rydell, McConnell, and Carr (2006) extend this idea and demonstrate the working memory impairment is caused by explicit monitoring of performance for tasks that have been proceduralized (also see Cadinu et al., 2003 for an earlier discussion of the role of divided attention). They used golf experts and a putting task. For these participants, putting should be an automatized skill. Beilock et al. induced stereotype threat by telling half of their male participants that women tend to perform better on the putting task. As compared to a control group, these men performed worse. Interestingly, this performance decrement was eliminated in Experiments 2 and 3 by giving participants a dual task to perform. This result is nicely consistent with Beilock and colleagues previous work on "choking under pressure" (Beilock & Carr, 2005; Beilock, Kulp, Holt, & Carr, 2004) in which induced pressure resulted in explicit monitoring of performance, and previous demonstrations of the role of negative thinking under stereotype threat (Cadinu, Maass, Rosabianca, & Kiesner, 2005).

Most relevant to my Experiments, Seibt and Förster (2004) argue that activating stereotypes induces regulatory foci, which in turn influence performance. Seibt and Förster demonstrate that a negative stereotype induces a prevention focus while a positive stereotype induces a promotion focus. In order to evaluate this claim, I turn now to an overview of Regulatory Focus Theory.

Chapter 3: Regulatory-Focus Theory

DEFINITION OF REGULATORY FOCUS

Regulatory focus is a construct from social psychology developed by Higgins and his colleagues (Higgins, 1987, 1997a). Higgins suggests motivational states differ across individuals. Although philosophers and psychologists previously recognized that people seek pleasure and avoid pain, Higgins contributed by asking the source of these hedonic motives. How exactly are people acting in concert with the hedonic or pleasure principle? What are the mechanisms?

Higgins proposes that regulatory focus is a motivational mechanism that influences people's sensitivity to potential gains and losses in their environment (Higgins, 1987, 1997a). The motivation literature has long made a distinction between *approach* states (those that are desirable) and *avoidance* states (those that are undesirable) (see Carver & Scheier, 1990; Markman & Brendl, 2000; Miller, 1959 for further discussion). These approach and avoidance mechanisms underlie the hedonic principle and are *self-regulatory strategies* that focus people toward desired end states. Importantly, individuals can be striving to achieve different end states, which represent different needs: nurturance or security (Higgins, 1997a; Shah & Higgins, 1997; Shah, Higgins, & Friedman, 1998). For nurturance needs, the individual is seeking some positive state and for security needs, the individual is trying to avoid some negative state. Therefore, approach and avoidance strategies are associated with approaching desirable end states and avoiding undesirable end states, respectively.

Orthogonal to this distinction, Higgins (1987; 1997a) argues that individuals may differ in their relative attention to gains or losses in the environment. A focus on the presence or absence of gains is called a *promotion focus*, and a focus on the presence or absence of losses is called a *prevention focus*. For example, an individual with a promotion focus might be concerned with earning an A on an exam by earning enough points while an individual with a prevention focus might be concerned with preventing a B on an exam by losing too many points. In this example, a promotion-focused individual who successfully earns an A has the same outcome as a prevention-focused individual who successfully avoids earning a B. According to Higgins, when an individual decides to pursue a goal, they do so with a particular motivational orientation determined by stable personality characteristics and this focus guides their processing.

However, while people differ in the chronic accessibility of these foci, often situations that have salient potential gains or losses may induce a regulatory focus that overrides a person's chronic focus (Higgins, 2000b; Shah et al., 1998). For example, comments from a parent or experimental task instructions that frame the situation using gains or losses (e.g., Shah, Higgins, & Friedman, 1998) may be sufficient to induce a regulatory focus. Furthermore, some of the hallmarks of the regulatory foci can be used to induce a specific focus by activating security or nurturance needs (e.g., Friedman & Forster, 2001) or priming a self standard (e.g., Freitas, Liberman, & Higgins, 2002; Higgins, Roney, Crowe, & Hymes, 1994). For example, Higgins et al. (1994) demonstrate that asking individuals to think about their goals or ideals places them into a promotion focus whereas asking individuals to think about their responsibilities places them into a prevention focus.

As first outlined in self-discrepancy theory (Higgins, 1987), ideals represent possible positive end states approached when in a promotion focus while responsibilities represent possible negative end states to be avoided when in a prevention focus (Higgins, 1997b). These goal differences have been exploited to create measures of chronic and situationally-induced foci. Theoretically, the presence of a promotion focus makes the discrepancies between the ideal self and the actual self more accessible while a prevention focus makes the discrepancies between the ought self and the actual self more accessible (Higgins, Shah, & Friedman, 1997). Measures of chronic focus, such as the Selves Questionnaire (Higgins, Klein, & Strauman, 1985) or the Regulatory Focus Questionnaire (Higgins et al., 2001), can be used to determine the discrepancies between actual, ideal, and ought selves, while other procedures rely on the speed of accepting or rejecting self statements (Shah et al., 1998) or the endorsement of specific self statements (Lockwood, Jordan, & Kunda, 2002), such as "I often worry that I will fail to accomplish my academic goals."

EFFECTS OF REGULATORY FOCI

The implications of maintaining either a promotion or a prevention focus are numerous. Early research on regulatory focus examined differences in emotions and cognitive processes associated with promotion and prevention foci (e.g., Forster & Higgins, 2005; Higgins, 1997a). There are different affective experiences which result from successful or unsuccessful approach and avoidance (Higgins, 1997a). A promotion focus induces an attempt to approach positive end states. If an end state is achieved, the individual will feel happiness whereas failure will lead to sadness. In contrast, a prevention focus induces an attempt to avoid negative end states. If an end state is successfully avoided, the individual is likely to feel relief whereas failures to avoid the end state will likely result in anxiety.

Förster and Higgins (2005) argue that a promotion focus supports more global processing while a prevention focus supports more local processing. Evidence for this claim comes from embedded figures tests (Forster & Higgins, 2005), tests of creative performance (Friedman & Forster, 2001), preferences for stability and change (Liberman, Idson, Camacho, & Higgins, 1999b), hypothesis generation (Liberman, Molden, Idson, & Higgins, 2001), and probability estimates for conjunctive and disjunctive events (Brockner, Paruchuri, Idson, & Higgins, 2002). For example, Friedman and Förster (2001) motivated this prediction by assuming that security related concerns associated with a prevention focus historically required the individual to focus more on specific aspects of their local surroundings. A promotion focus does not require this attention to detail. They suggest that this fundamental difference evolved into different processing styles induced by regulatory foci. Being in a particular focus promotes a scanning of the environment to find things which are consistent with goal strivings to increase the likelihood of goal attainment. A prevention focus supports attention to more concrete details while a promotion focus supports attention to more ideal and more abstract elements.

In Experiments 1 and 2, their participants completed a maze task that involved either a mouse trying to get to a piece of Swiss cheese (promotion) or a mouse trying to avoid being eaten by an owl (prevention). These mazes served to induce either a promotion or a prevention focus. In Experiment 1, promotion-primed participants identified more embedded images in a task that required them to extract simple images masked with visual noise as compared to prevention-primed participants. In a better test of creative performance, in Experiment 2, promotion-primed participants generated more creative uses for a brick as compared to prevention-primed participants. These results were replicated with a different test of creative performance using chronic focus in Experiment 5.

Based on their results from Experiments 3 and 4, they believe enhanced creativity is the result of a more risky response bias. Higher risk should be associated with guaranteeing hits and ensuring against errors of omission whereas a more conservative strategy would be associated with a focus on correct rejections and errors of commission. In Experiment 3, participants were told that if they performed well they would get to do a pleasant task (promotion focus) or were told that by not performing poorly they would not have to do an unpleasant task (prevention focus). As predicted, promotion-focused individuals were more biased to say “yes” in a recognition memory task whereas prevention-focused individuals were more biased to say “no”.

The results of Experiment 4 are more intriguing. They use a word-fragment completion task and have participants complete the task twice. Promotion-primed participants generated more novel solutions to the second task than prevention-primed participants. They argue that promotion helps mitigate the effects of retrieval inhibition and suggest that promotion should also improve performance in functional fixedness and tip-of-the-tongue states. In essence, a promotion focus promotes a more *elaborative* processing style that reduces perseveration on already generated solutions.

Similarly, in Förster and Higgins (2005), participants completed an embedded figures task in which larger letters were composed of smaller component ones. For example, a large letter H would be constructed with small Ts. Participants responded to both global and local letters on different trials and their chronic levels of promotion and prevention were measured. Consistent with their predictions, Förster and Higgins found that promotion-focused participants responded more quickly to global letters whereas prevention-focused participants responded more quickly to local letters.

REGULATORY FIT

More recent work on regulatory focus examines how a person's regulatory focus typically interacts with salient aspects of the task to determine the cognitive and evaluative processes that are brought to bear on performance (Higgins, 2005; Higgins, Idson, Freitas, Spiegel, & Molden, 2003; Idson, Liberman, & Higgins, 2004; Keller & Bless, 2006; Lee & Aaker, 2004; Maddox et al., 2006; Maddox et al., 2007; Markman, Baldwin et al., 2005; Shah et al., 1998). For example, Higgins and colleagues found that the value people give items in the environment depends on the fit between a person's regulatory focus and aspects of the items being evaluated (Camacho, Higgins, & Luger, 2003; Higgins, 2000a; Shah et al., 1998). This correspondence between focus and the environment is known as *regulatory fit.*

A regulatory fit exists when there is a match between the focus and a corresponding state (Higgins, 1997b). A regulatory match could be present when the desired goal matches the possible outcome state, when there is a match between the focus and the strategies used to pursue a goal (Higgins, 2000a), and when there is a match between the focus and the strategy afforded by the environment (Keller & Bless, 2006; Shah et al., 1998). It is also possible to get a match between situational focus or chronic focus and specific motor actions related to approaching positive stimuli or avoiding negative stimuli (Cacioppo, Priester, & Berntson, 1993; Forster, Higgins, & Idson, 1998).

Another form of fit exists when a person's regulatory focus matches the reward structure of the task they are performing (Keller & Bless, 2006; Maddox et al., 2006; Shah et al., 1998). A promotion focus increases people's sensitivity to gains and nongains, and so there is a regulatory match between individuals with a promotion focus and tasks in which people gain rewards (e.g., points in a task), but a regulatory mismatch for those participants when they must avoid punishments (e.g., losing points). In contrast, a prevention focus increases people's sensitivity to losses and so there is a regulatory match between individuals with a prevention focus and tasks for which they must avoid losses, but a regulatory mismatch for those participants in tasks for which they must achieve gains.

So what are some of the cognitive and evaluative processes that vary with regulatory fit? Given a match, Higgins argues individuals experience their goal strivings more strongly and feel more positively about their reactions. He argues that a regulatory fit enhances task engagement, which increases the perceived value of the task (Higgins, 2000a). Simply, match states feel better than mismatch states (Aaker & Lee, 2006; Cesario, Grant, & Higgins, 2004; Kruglanski, 2006; Sassenberg, Jonas, Shah, & Brazy, 2007). For example, Higgins, Idson, Freitas, Spiegel, and Molden (2003) argue that a fit between strategy and focus makes people feel good about the decisions that they make (see also Avnet & Higgins, 2006; Idson et al., 2004). They asked participants to think about a choice between a mug and a pen. Half of the participants were instructed to think about what they would gain from the choice and the other half considered what they would lose from the choice. Participants then assigned a price to the mug. Participants assigned a higher price when the task instructions matched their chronic focus. That is, prevention-focused individuals assigned a higher price given the losses instructions as compared to the gains instructions whereas the opposite was true for promotion-focused individuals.

This result has been replicated by Förster and Higgins (2005) by priming participants with a local or global processing style. In this study, global-primed participants assigned a higher value to the mug given the gains task as compared to the losses task whereas the local-primed participants assigned a higher value to the mug given the losses task as compared to the gains task. Furthermore, there was a main effect of processing in that the global-primed participants assigned higher values than the local-primed participants.

Individuals also make qualitatively different decisions given a regulatory match (Lee & Aaker, 2004). For example, in Experiments 1 and 5, Lee and Aaker showed participants messages that were either gain-framed or loss-framed after situationally manipulating their focus and argued that a fit between the message frame and the situational focus resulted in the message appearing more persuasive. The situational focus manipulation involved reading a passage that described grape juice as an energy drink (promotion) or as a way to prevent cancer and heart disease (prevention). The message frame consisted of a positive or negative tagline associated with the passage.

For example, the promotion passage taglines were “Get Energized” and “Don’t Miss Out on Getting Energized” for the gains and losses, respectively.

Lee and Aaker interpret their effects as the result of a fit “feeling right” and therefore processing feels easier and more fluent. Lee and Aaker (2004, Experiment 4A) demonstrate that people rate the information in a passage as being easier to process when they are in a regulatory match produced by situational focus and message framing as compared to those in a regulatory mismatch. This effect persists using a more implicit measure of processing fluency in their Experiment 4B.

In addition, individuals seek role-models that match their regulatory focus concerns (Lockwood, Jordan, & Kunda, 2002), such that promotion individuals are more motivated by positive role models and prevention individuals are more motivated by negative role models who inspire fear by representing the to-be-avoided self. Lockwood et al. (2002) demonstrate that individuals primed with a promotion focus respond with higher levels of motivation when presented with a positive role model who activates approach strategies as compared to a negative role model who activates avoidance strategies. However, the opposite is true for prevention-primed participants. They argue the greatest motivational benefits are possible when individuals are presented with role models that match their underlying motivational focus.

Other studies have demonstrated improved task performance from regulatory fit (Keller & Bless, 2006; Shah et al., 1998; Spiegel, Grant-Pillow, & Higgins, 2004). Spiegel et al. (2004) required participants to read a health message about eating fruits and vegetables framed as either promotion or prevention. Half of the participants listed benefits of eating fruits and vegetables and half listed costs. Spiegel et al. demonstrated that the match conditions (promotion/benefits and prevention/costs) ate more fruits and vegetables over the following week than the mismatch conditions (promotion/costs and prevention/benefits).

Keller and Bless (2006) demonstrated that matching the chronic focus of the individual and the situational focus improves cognitive performance for both mathematical and spatial tests in a real-world context. They used secondary school students in grades 6 and 8 and measured chronic focus using a modified version of the Selves Questionnaire (Higgins et al., 1985) in Experiment 1 and the scale developed by Lockwood et al. (2002) in Experiment 2. In Experiment 1, Keller and Bless manipulated situational focus by telling promotion participants that the test was designed to identify people with exceptional ability but not weak ability. In contrast, the prevention participants were told that the test was designed with the reverse goal in mind. As predicted, they found that participants got more correct on the math test when their chronic focus matched the situational manipulation as compared to when their chronic focus did not match the situational manipulation. In Experiment 2, a situational-promotion focus was created by telling participants that they would gain one point for each correct response but gain no points for incorrect responses. For the situational-prevention focus condition, participants received a point for a correct response but lost a point for each incorrect or uncompleted response. Like Experiment 1, the match conditions produced better test performance as compared to the mismatch conditions. Of note, the situational-prevention focus condition is not a true losses task (for a better demonstration, see Maddox, Baldwin, Markman, 2006).

As another example, Shah, Higgins, and Friedman (1998) manipulated the task reward structure in both Experiments 1 and 2. Promotion focus participants were told that they could earn an extra dollar for good performance (e.g., 90% correct) while prevention focus participants started with the extra dollar and were told that they could lose the extra dollar or not based on their performance (e.g., miss no more than 10%). They also measured chronic focus using a modification of the Selves Questionnaire (Higgins et al., 1985) which recorded the accessibility of the attributes using response time. Their participants completed an anagram task that required them to unscramble letters to form words.

In Experiment 1, they found that the ideal self ratings better predicted anagram task performance given the promotion framing as compared to prevention framing whereas the ought self ratings better predicted performance as compared to ideal ratings given the prevention framing. Experiment 2 replicated Experiment 1 with a slight change to the anagram task. Some anagrams had the possibility of gaining points while others had the possibility of losing points. First, with respect to the original design, Experiment 2 replicated Experiment 1. Second, ideal ratings better predicted performance in the promotion and gains task as compared to the other tasks (e.g., prevention and gains, prevention and losses, and promotion and losses) and ought ratings better predicted performance in the prevention and losses task as compared to the other tasks. Shah et al. argue this result supports the claim that different foci are consistent with different means of goal attainment. The promotion participants focused more on the dual gains frame (extra money and points possible) whereas the prevention participants focused more on the dual losses frame (loss of money and points possible).

A number of recent studies have demonstrated that participants with a regulatory fit are able to perform more flexibly in cognitive tasks than are participants with a regulatory mismatch, regardless of whether that fit comes from having a promotion focus and a task with a gains reward structure or a prevention focus with a losses reward structure (Grimm, Markman, Maddox, & Baldwin, in press; Maddox et al., 2006; Maddox et al., 2007; Markman, Baldwin et al., 2005; Markman, Maddox et al., 2005; Markman et al., 2006). I describe one example of this effect of regulatory fit in detail, because the experiments I present in this dissertation are based on this example (Maddox et al., 2006).

Some participants were given a situational promotion focus by giving them the opportunity to obtain a raffle ticket for a drawing to win $50 if their performance exceeded a criterion. Other participants were given a situational prevention focus by giving them a raffle ticket for this drawing and telling them that they could keep the ticket as long as their performance exceeded the criterion, otherwise, they would lose it. This is a variation of a manipulation first presented by Higgins and colleagues (Higgins, 1997a; Liberman, Idson, Camacho, & Higgins, 1999a). The reward structure of the classification task was manipulated between subjects as well. Participants given a gains reward structure received points for every response, but got more points for correct responses than for incorrect responses. Participants given a losses reward structure lost points for every response, but lost fewer points for correct responses than for incorrect responses.

In Experiments 1a and 1b, participants were given a perceptual classification task in which they had to learn to classify lines that varied in their length, position, and orientation. The task required learning a subtle classification rule involving the length and orientation dimensions. A simple rule involving only the highly salient position dimension would yield good performance, but not sufficiently good performance to achieve the performance criterion.

The lines can be correctly classified using an explicit verbalizable rule. Simply stated, lines are in Category A if the length is long and the orientation is steep and otherwise the lines are in Category B (see Figure 3.1).

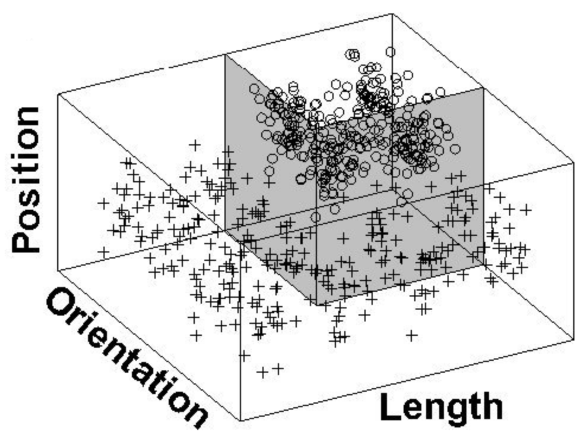


Figure 3.1. Stimulus space with correct conjunctive rule on length and orientation dimensions represented.

Using this rule, a participant can get 100% accuracy on the task. However, given that there are easier verbalizable rules using any of the dimensions independently, a participant needs to move from using one of these unidimensional rules, each of which yields at most 83% accuracy on the task, to the more complex conjunctive rule using length and orientation. Thus, this task requires flexibility because it forces participants to search the possible strategy space to find a less obvious but more effective strategy for classifying the items.

In Experiment 1A, participants gained points on every trial. Based on the sensitivity produced by a promotion focus to gains and non-gains in the environment, the promotion-primed participants are experiencing a *regulatory match* in this task environment whereas the prevention-primed participants are experiencing a *regulatory mismatch*. In Experiment 1B, participants lost points on every trial. Again, as suggested by work on regulatory focus, the prevention-primed participants are now experiencing a regulatory match while promotion-primed participants are experiencing a regulatory mismatch. Participants with a regulatory match (e.g., promotion focus with gains or prevention focus with losses) performed better and were more likely to achieve the performance criterion than were participants with a regulatory mismatch (e.g., prevention focus with gains or promotion focus with losses).

One advantage of using this classification task is that it is possible to fit mathematical models to the data to describe the performance of individual participants on a block-by-block basis. Maddox, Baldwin, and Markman (2006) found that early in learning, people's performance was best characterized as using a simple rule along one dimension. Later, they learned to classify on the basis of the correct two-dimensional rule. Participants with a regulatory fit found that two-dimensional rule earlier in the task than did those with a regulatory mismatch.

In Experiment 3, the classification task used an information-integration category structure (e.g., Maddox & Ashby, 2004) and a losses reward structure. Participants were primed with a situational focus using the same manipulation described above. The stimulus dimensions used for this task were length, orientation and position, as in Experiments 1 and 2, but correct classification required participants learn a rule that is not easily verbalizable (see Figure 3.2).

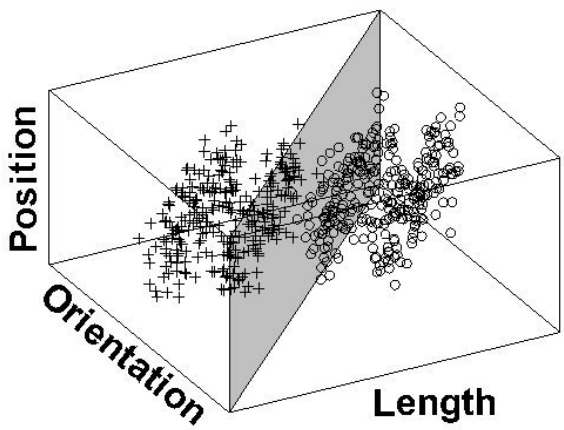


Figure 3.2. Stimulus space with correct information-integration rule on length and orientation dimensions represented.

The optimal information-integration rule allows for perfect performance on the task while a unidimensional rule at best allows for 83% accuracy. They argued that task performance would be hurt by flexible processing. Performing well on this task requires that the participant abandon the explicit rule-based testing system in favor of the implicit learning system. As such, they predicted and obtained support for the claim that participants experiencing a regulatory match would perform more poorly on the task than participants experiencing a regulatory mismatch.

Chapter 4: Experiments

There are several reasons to believe stereotype threat and regulatory focus are related phenomena. Work has explicitly linked stereotype threat effects with regulatory focus (Forster, Higgins, & Strack, 2000; Keller & Dauenheimer, 2003; Seibt & Forster, 2004) by examining the role of regulatory focus in the processing of stereotypic information (Forster et al., 2000) and by studying the mediation of stereotype threat by emotions induced by regulatory focus states (Keller & Dauenheimer, 2003). Further, a study on regulatory focus (Keller & Bless, 2006) and a study examining stereotype threat (Brown & Josephs, 1999) use the same manipulation. Brown and Josephs manipulate stereotype threat by framing a test as diagnostic of weak or strong ability. They argue that the weak ability condition corresponds to the negative stereotype women desire to avoid confirming and the strong ability condition corresponds to the positive stereotype that men desire to confirm. Likewise, Keller and Bless manipulate situational focus using the same test framing. However, they argue that the weak ability condition primed a situational-prevention focus and the strong ability condition primed a situational-promotion focus.

Förster et al. (2000) examined the influence of regulatory focus on the processing of stereotypic information. Their correlational study revealed that the higher participants' scores in modern sexism and in prevention the more they recalled stereotype-incongruent information. In contrast, the higher participants' score in promotion the more they recalled stereotype-congruent information. Theoretically, these relationships were predicted because the more vigilant processing style associated with a prevention focus causes individuals to seek disconfirming information. It is less clear why higher promotion resulted in more recall of congruent information.

In another important demonstration, Keller and Dauenheimer (2003) recognized the interaction between stereotype threat and situationally-induced regulatory focus. They predicted that affective reactions to stereotype threat are mediated by situationally-induced focus. To induce a promotion focus, they told all students to try to solve as many math problems as possible. Half of the students were given a gender stereotype manipulation: "The following math test…has been shown to produce gender differences in the past." Keller and Dauenheimer assumed this primed girls with a negative stereotype and boys with a positive stereotype. While working on the math problems, students rated the degree to which they were experiencing emotions related to promotion and prevention foci: tense, nervous, anxious, depressed, uncertain, agitated, calm, self-conscious, quiet and unconcerned were prevention emotions, and disappointed, frustrated, sad, contented, enthusiastic, light-hearted, happy, and balanced were the promotion emotions. Keller and Dauenheimer found that girls underperformed in the stereotype threat condition as compared to boys, and boys and girls in the control condition. Examining only the girls, dejection (a combination of disappointed, frustrated, and sad) served as a mediator of performance. They argued that this result shows situational promotion influencing the emotions experienced in stereotype threat.

In 2004, Seibt and Förster advanced an insightful proposal that differences in regulatory focus are in fact the cause of stereotype threat effects. In a series of experiments, they demonstrate that priming individuals with a negative stereotype induces a prevention focus while priming individuals with a positive stereotype induces a promotion focus. They argue that the induced regulatory focus produces differential performance on tasks for which elaborative processing is required: a positive stereotype supports more global/elaborative processing because of the induced promotion focus while a negative stereotype supports more local processing because of the induced prevention focus. In Experiment 2, they test for this performance difference by examining speed/accuracy tradeoffs. They assume that individuals given a positive stereotype should be faster and make more errors. In contrast, individuals given a negative stereotype should be slower and make fewer errors due to the use of a more vigilant processing style. As predicted, they found faster performance for the positive-stereotype group relative to the control and slower performance for the negative-stereotype group relative to the control. In addition, they found the opposite pattern for errors.

In my view, Seibt and Förster (2004) correctly argue that activating stereotypes induces regulatory foci, which in turn influence performance. They demonstrate that a negative stereotype induces a prevention focus while a positive stereotype induces a promotion focus. Based on Friedman and Förster (2001), they assume that a prevention focus causes individuals to pay more attention to details and be more vigilant whereas a promotion focus causes individuals to engage in more abstract creative processing.

The critical question is why did Seibt and Förster find these speed/accuracy differences between their focus groups? Is it true that a prevention focus always results in more detail-oriented processing and higher accuracy on tasks that require less elaborative processing? I believe that the answer depends on the reward structure of the task.

As argued by Markman, Maddox and colleagues, understanding the influence of the reward structure of the task is a critical part of untangling the effects of regulatory focus. Most laboratory experiments, explicitly or implicitly, involve a gains and non-gains environment. Participants are trying to get the most answers correct or are trying to get the most points. In the most abstract sense, undergraduate participants may be trying to earn course credit by participating while others may be trying to get paid for participation. Much of the prior work on regulatory focus and all of the work cited above examining the influence of regulatory focus on stereotype threat fails to consider the importance of the interaction of regulatory focus with this type of task environment.

While making an excellent observation linking stereotypes with induced-regulatory foci, Seibt and Förster (2004) fail to consider the influence of task reward structure and therefore miss a critical part of the explanation. Let us assume that their participants were functioning in a gains and non-gains environment. This would result in their promotion-focused participants being in a regulatory match. That is, their focus matches the task environment. In contrast, their prevention-focused participants would be in a regulatory mismatch.

I suggest that the differences in regulatory fit actually are responsible for producing their effects. Based on the work by Maddox, Baldwin, and Markman (2006), I argue that a regulatory fit produces flexible cognitive performance. Critically, flexible abstract processing is a hallmark of a regulatory match, not of a promotion focus, just as detailed local processing is a hallmark of a regulatory mismatch, not of a prevention focus.

Table 4.1 summarizes the interaction between regulatory focus and task reward structure. My argument is that previous demonstrations of stereotype threat have assessed the left-hand column of this table. Typical cognitive tasks involve an explicit or implicit gain structure. Participants are trying to achieve correct answers to questions and are typically rewarded for being correct. Participants who have a negative task-relevant stereotype have a prevention focus, and thus are in a regulatory mismatch. Further, the majority of tasks, like math problems and verbal reasoning, tend to require flexible processing. Because the tasks are difficult, this mismatch leads to poorer performance than is observed in participants who do not have a negative task-relevant stereotype. This latter group either has a positive task-relevant stereotype, in which case they likely have a promotion focus, or else they have no task-relevant stereotype in which case their performance will be driven in part by their chronically accessible regulatory focus.

My analysis suggests that if I assessed the performance of participants in a loss condition (the rightmost column of Table 4.1), then the effects of having a negative task-relevant stereotype should reverse. That is, participants with a negative task-relevant stereotype should actually do better when there is a loss reward structure than should those participants with a positive task-relevant stereotype.

	Gains	Losses
Positive stereotype („Promotion")	Match	Mismatch
Negative Stereotype („Prevention")	Mismatch	Match

Table 4.1. Schematic representation of regulatory matches and mismatches

I present the following experiments to support my claim that the processing strategies differentially employed by induced promotion and prevention are the result of the interaction between task reward structure and regulatory focus instead of a main effect of promotion and prevention. Importantly, this suggests that classic stereotype threat effects are the result of a regulatory mismatch. The activation of a negative stereotype produces prevention-primed participants who are likely acting in a gains and non-gains environment, which results in a regulatory mismatch. These participants would then underperform on tasks requiring flexible processing.

EXPERIMENTAL OVERVIEW

According to the COmpetion between Verbal and Implicit Systems (COVIS) model of multiple memory systems (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Waldron, 1999), one memory system involves an explicit rule-based processor that provides the means to test different explicit hypotheses while another competing system is an implicit procedural learning system. The explicit system is believed to take place in frontal brain regions which are used for flexible processing while the procedural system is believed to be instantiated in sub-cortical areas, such as the tail of the caudate nucleus (Maddox & Ashby, 2004). Experiments 1 and 2 use a rule-based classification task, which requires the participant use rule-based strategies to explicitly test different rules to correctly categorize the stimuli. Arthur B. Markman, W. Todd Maddox, and Grant C. Baldwin collaborated on these experiments. The task requires cognitive flexibility because participants need to discard easy to find unidimensional classification rules in favor of a more complicated conjunctive classification rule that requires the use of two stimulus dimensions. Experiments 3 and 4 use an information-integration category structure that needs to be learned by the procedural learning system. For this category structure, persistence with flexible rule testing using the explicit system would hurt performance.

EXPERIMENTS 1 AND 2: RULE-BASED TASK

In these studies, we selected the rule-based classification task used by Maddox, Baldwin, & Markman (2006) described above. Because this task involves the classification of lines, we could create arbitrary stereotypes and present them to participants. Thus, in Experiment 1, we told male and female participants that this task is one for which women have previously been demonstrated to do better than men. In Experiment 2, we presented participants with the opposite story, so participants were told that men perform better than women on this task. In both studies, the negative task-relevant stereotype was expected to create a prevention focus, and the positive task-relevant stereotype was expected to create a promotion focus. Participants were then given the classification task with a gains or a losses reward structure. Thus, participants with a negative task-relevant stereotype have a regulatory fit when the task has a losses reward structure, and so they should perform better than when the task has a gains reward structure and they have a mismatch. In contrast, participants with a positive task-relevant stereotype should have a regulatory fit for the gains reward structure, and thus should perform better than when they perform the task with a losses reward structure and have a regulatory mismatch. After the presentation of the two experiments, I fit classification models to participants' performance to examine how rule-use changes over the course of the studies. The structure of Experiments 1 and 2 appears in Table 4.2.

Experiment	Stereotype	Gender	Gains	Losses
1	Positive ("Promotion")	Women	Match	Mismatch
	Negative ("Prevention")	Men	Mismatch	Match
2	Positive ("Promotion")	Men	Match	Mismatch
	Negative ("Prevention")	Women	Match	Mismatch

Table 4.2. Structure of Experiments 1 and 2

Based on the work by Maddox, Markman, and colleagues, the match condition participants should engage in more flexible processing leading to better performance in the classification task as compared to the mismatch conditions. Critically, I will show that a promotion focus induced by a positive stereotype will not always lead to more flexible processing. Instead, an induced promotion focus will only lead to flexible processing in the gains task. Likewise, I will demonstrate that an induced prevention focus will lead to more flexible processing in the losses task. I will replicate stereotype threat effects but only in the gains task. I predict a negative stereotype will lead to poorer performance because of the presence of a regulatory mismatch. In sum, given that this task requires flexible processing, the match conditions should outperform the mismatch conditions.

These predictions are also supported by some prior work on stereotype threat (Quinn & Spencer, 2001; Shih, Pittinsky, & Ambady, 1999; Walton & Cohen, 2003; Wraga, Helt, Duncan, & Jacobs, 2006). First, Wraga et al. (2006), Walton and Cohen (2003), and Shih et al. (1999) present evidence for improved performance by groups with positive stereotypes. Walton and Cohen label this phenomena *stereotype lift*. In a meta-analytic review of 43 studies, they found improved performance by the non-negatively stereotyped group in the stereotype-relevant condition as compared to the stereotype-irrelevant or control condition.

Second, Quinn and Spencer (2001) find reduced strategy use given stereotype threat. In their study, women and men completed a series of math problems from the SAT while verbalizing their thought processes. Women were primed with a negative stereotype. Quinn and Spencer coded the number of problem solving strategies used by participants. They found women in the stereotype threat condition failed to find any strategy 14% of the time as compared to 2% in the control condition. This finding maps directly to our claim that participants in a regulatory mismatch (i.e., negative stereotype in a gains task) will display less flexible processing as compared to participants in a regulatory match (i.e., positive/neutral stereotype in a gains task).

Experiment 1

In Experiment 1, we arbitrarily chose to give women a positive stereotype and men a negative stereotype; in Experiment 2 we do the opposite. We assume that priming a positive stereotype will induce a promotion focus while priming a negative stereotype will induce a prevention focus. We also manipulated the task reward structure; half of the participants completed the task gaining points for correct responses and gaining zero points for errors while half lost fewer points for correct responses than for errors. Using the concept of a regulatory fit, we predict that women will perform better in the gains version of the task as compared to the losses version. In contrast, men will perform better in the losses version of the task as compared to the gains version. Furthermore, we predict that we will replicate the classic stereotype threat work and demonstrate that women will perform better than men in the gains version of the task. Based on our theoretical framework, we also predict that men will perform better than women in the losses version of the task.

Method

Participants

Eighty undergraduate students (40 men and 40 women) at the University of Texas at Austin were given $8 for their participation. Half of the men and half of the women were randomly assigned to the gains and losses reward structures.

Design

This Experiment had a 2 (Gender: Male, Female) × 2 (Reward Structure: Gains, Losses) design. Reward Structure was manipulated between subjects.

Stimuli and Stimulus Presentation

Participants viewed stimuli on a computer screen and were asked to classify a set of items into one of two categories. The stimuli to be categorized were lines that varied across items in their length, orientation, and position within a box on the screen. The stimulus structure is shown in Figure 3.1. For Category A, there were 24 stimuli sampled from each of 12 bivariate normal distributions on length and orientation resulting in 288 stimuli. For Category B, there were 72 stimuli sampled from 4 bivariate normal distributions on length and orientation resulting in 288 stimuli. The position dimension was sampled independently of length and orientation for each category: Category A used a univariate normal distribution with a mean of 253 pixels and a standard deviation of 75 and Category B used a univariate normal distribution with a mean of 397 pixels and a standard deviation of 75.[1] The lines were presented inside of a black 650 x 650 pixel box, centered vertically, and were randomly ordered for each participant in each block. There were 48 trials in each block and 12 blocks.

The stimuli were generated such that using the position on the screen or the orientation of the line or the length of the line to classify the stimuli will result in 83% accuracy for a block of trials. For example, Figure 4.1 shows the stimulus space and the set of items. Each of the three possible dimensions (length, orientation, and position) is represented; each point is a specific line stimulus. This stimulus space is being divided by a plane representing a decision criterion set using position. A subject using this decision bound would classify all stimuli falling above the bound into Category A (represented by circles) and all stimuli falling below the bound into Category B (represented by plus-signs). The unidimensional rules are fairly easy to verbalize and are salient to participants (Maddox, Baldwin, & Markman, 2006). However, in this example, using a position decision criterion only allows for 83% correct classification.

---

[1] By independently sampling position, we were able to make position especially salient to insure that our participants would start with a simple unidimensional rule.

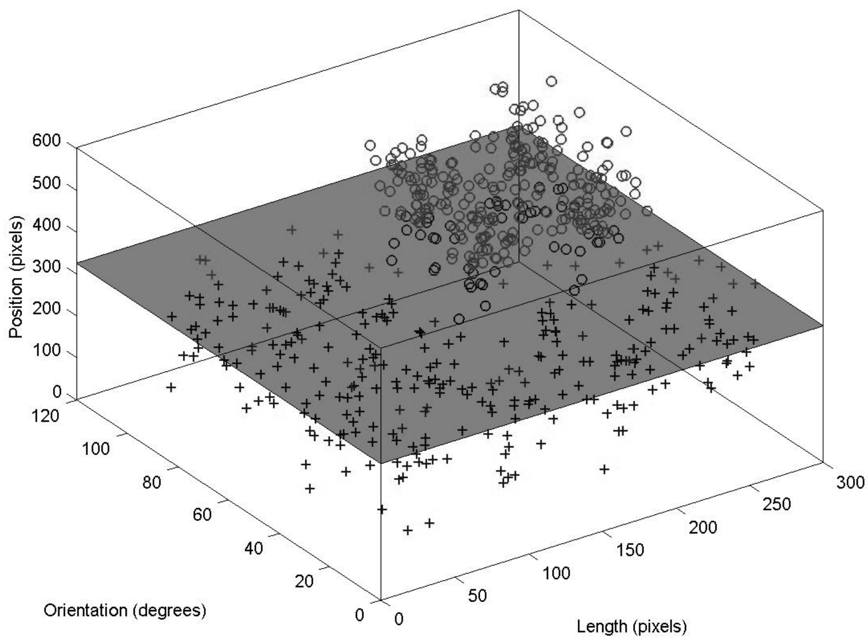


Figure 4.1. Stimulus space used in Experiments 1 and 2 with a unidimensional rule on position represented.

There is an optimal decision bound for this task that, if used, will yield 100% accuracy on the task. This decision criterion requires a rule that takes into account both length and orientation. This rule is: If the length is long and the orientation is steep, then respond Category A; otherwise, respond Category B (please see Figure 3.1 for a graphical representation of this rule). In order for participants to perform well in the task, they need to abandon the use of unidimensional rules in favor of the more complex conjunctive one. This switch requires cognitive flexibility.[2]

Materials

We used the Regulatory Focus Questionnaire (RFQ: Higgins et al., 2001) as a measure of chronic promotion and chronic prevention focus. This questionnaire asks participants to rate the frequency of specific events in their lives. Participants completed the Beck Anxiety Inventory (BAI: Beck, Epstein, Brown, & Steer, 1988) the Penn State Worry Questionnaire (PSWQ: Meyer, Miller, & Metzger, 1990). The BAI requires the participant to report how much they have been bothered by a range of symptoms in the last week, such as “terrified”, “nervous”, and “faint”. The PSWQ requires that the participants rate how characteristic displayed items are of them. Participants also completed the Positive and Negative Affect Schedule (PANAS: Watson, Clark, & Tellegen, 1988) which is a 20 adjective checklist that asks participants to report current emotional states (please see Appendix for scales).

Procedure

Participants were tested in individual cubicles. Participants first completed the RFQ, the PSWQ, and the BAI. At the beginning of the classification task, participants were told that their job was to learn to classify items into two categories. To induce a stereotype our participants read: “This is an experiment testing sex differences in spatial

[2] It is important to note that it is possible to use a conjunctive rule on length and orientation and not have perfect task performance. Participants may set a rule using both dimensions but will not do so with a high level of precision. This form of the rule is known as a sub-optimal rule on length and orientation.

abilities. Previous research has shown that women perform better than men on tests of spatial ability." Thus, women in this task have a positive task-relevant stereotype and men have a negative task-relevant stereotype.

In the gains task, participants were told that women tended to earn more than 86 points per block, which is equivalent to a 90% correct criterion (correct on 43 of 48 trials), and men tended to earn fewer. In the losses task, participants were told that women tended to lose less than 58 points per block, which is again equivalent to a 90% criterion (correct on 43 of 48 trials), and men tended to lose more.

Participants were able to track their progress using a vertically oriented "point meter". The point meter was located on the right side of the screen and was 750 x 50 pixels. The 0 point was marked on the meter as was the 90% criterion line. Every time a participant correctly categorized an item, they heard a "ching" sound, like that from a cash register, and the word "Correct" appeared on the screen. When participants were incorrect, they heard a buzzer and the word "Incorrect" appeared. For participants in the gains task, the point meter started at 0, located at the bottom of the point meter, and participants gained 2 points for each correct response and gained 0 points for an incorrect response. Also, the 90% criterion line was labeled "86 points". For participants in the losses task, the point meter started at 0 but 0 was located at the top of the point meter and the bonus criterion was labeled "-58 points". During this task, participants lost 1 point for a correct response and 3 points for an incorrect response. Samples of the gains and losses task screens are in Figure 4.2.

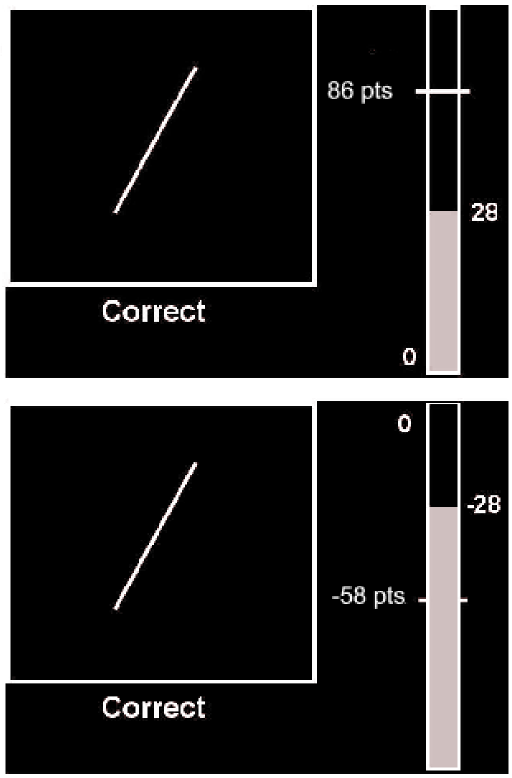


Figure 4.2. Sample screens for the gains and losses tasks

A series of rating scales were used as manipulation checks. These ratings were given just before participants began the classification task. Participants rated: “how well do you think you will perform in this task on a scale of 1 to 9, where 1 = very bad and 9 = very good? How much do you like the task? (1 = not at all, 9 = very much) and How motivated are you to do well on the task (1 to 9)”. Next, participants took the PANAS to get a measure of the positive and negative affect prior to completing the classification task.

Each participant completed 12 blocks of trials with 48 trials. For each trial, the stimulus was displayed until the participant responded "A" or "B". Following feedback, the stimulus display disappeared for 250ms for the inter-trial-interval. The point meter always remained visible.

After the classification task, participants completed a final set of questionnaires. Participants completed the PANAS to get a measure of positive and negative affect after the classification task. Participants were also asked to rate how well they believed they performed overall, how well they performed relative to men, and how well they performed relative to women.

Results

To test our hypotheses, we performed two different sets of analyses. First, we analyzed the accuracy data to determine how the interaction of reward structure and regulatory focus influenced a basic performance metric. Second, we used quantitative models to give us further insight into the strategies used by participants in the service of completing the task. By identifying the strategies likely implemented by participants, we are able to make claims about the types of processes used during the perceptual classification learning task. These model results are presented as supplementary support after the results from Experiment 2. We also present the results from the pre- and post-test questionnaires.

Questionnaire Results

All of the significant comparisons in the questionnaire results appear in Table 4.3. In addition, we found a significant interaction for the Negative Affect subscale of the PANAS collected after the manipulation. The data were analyzed using an analysis of variance (ANOVA) with Gender (Male, Female) and Reward Structure (Gains, Losses). This analysis revealed a marginally significant two-way interaction between Gender and Reward Structure, $F(1,76) = 3.56$, $MSE = 26.0$, $p = .06$. Men in the losses and gains tasks averaged 11.9 and 12.4, respectively. Women in the losses and gains tasks averaged 15.4 and 11.6, respectively and this difference was marginally significant, $t(38) = 1.88$, $p = .06$.

	Women	Men	Test
RFQ: Prevention subscale	17.7 (3.6)	15.9 (3.3)	$t(78) = 2.35$, $p < .05$
PSWQ	51.9 (13.8)	45.6 (11.1)	$t(78) = 2.26$, $p < .05$
BAI	34.2 (9.8)	30.6 (7.6)	$t(78) =1.87$, $p = .07$
Performance	6.1 (1.9)	6.8 (1.4)	$t(78) = 1.89$, $p = .07$
Performance: relative to men	6.3 (1.6)	7.1 (1.5)	$t(78) = 2.32$, $p < .05$

Table 4.3. Significant comparisons in the Questionnaire data in Experiment 1 with means (standard deviations in parentheses).

We used Analysis of Covariance (ANCOVA) to determine if the significant effects found in the questionnaire data could account for our effects. The collected scales and ratings were not significant covariates, and as such, do not provide a possible alternative explanation for our effects.[3] The bivariate correlations appear in Table 4.4.

[3] This is not surprising because main effects are not likely to account for our interactions in the accuracy data. The one significant interaction came from the Negative subscale of PANAS. This too did not serve as a covariate $F(1,78) = 1.42$, $MSE = .005$, $p = .24$.

	Total Accuracy	Promotion	Prevention	PSWQ	BAI	Expectation	Liking	Motivation	Positive PANAS (pre)	Negative PANAS (pre)	Positive PANAS (post)	Negative PANAS (post)	Performance	Performance versus men	Performance versus women
Total Accuracy	1	0	-0.084	-0.054	0.039	0.141	0.131	0.137	0.017	0.068	0.123	0.03	.553(**)	.492(**)	.446(**)
Promotion		1	.228(*)	-0.19	-0.149	-0.004	0	0.142	.227(*)	-.232(*)	.271(*)	-0.188	0.127	0.056	-0.012
Prevention			1	-0.085	-0.213	-0.042	-0.019	-0.039	-0.121	-0.081	0.049	-.292(**)	0.06	-0.028	-0.09
PSWQ				1	.501(**)	-.262(*)	0.078	0.004	-0.034	.495(**)	-0.021	.377(**)	-0.177	-0.169	-.291(**)
BAI					1	-.307(**)	0.081	0.057	0.044	.585(**)	0.048	.395(**)	-0.054	-0.031	-.229(*)
Expectation						1	.374(**)	.235(*)	0.127	-0.147	0.07	-0.076	.328(**)	.409(**)	.439(**)
Liking							1	.661(**)	.313(**)	.269(*)	0.214	-0.04	.249(*)	.405(**)	.272(*)
Motivation								1	.562(**)	0.207	.460(**)	-0.025	.377(**)	.420(**)	.342(**)
Positive PANAS (pre)									1	-0.095	.623(**)	-0.114	.237(*)	.229(*)	.240(*)
Negative PANAS (pre)										1	-0.032	.594(**)	0.139	0.131	-0.102
Positive PANAS (post)											1	-0.041	.289(**)	.226(*)	0.21
Negative PANAS (post)												1	-0.186	-0.16	-0.154
Performance													1	.790(**)	.729(**)
Performance versus men														1	.629(**)
Performance versus women															1

Note: * $p < .05$, ** $p < .01$.

Table 4.4. Correlations between scales and total accuracy in Experiment 1. Significant correlations appear in grey.

Accuracy Analysis

First, we computed the average accuracy for each participant in each block of trials. Second, we calculated the proportion of participants reaching criterion (90% accuracy) for each block. For our task, the criterion was established as part of the stereotype threat manipulation. Our participants were told that women tended to get the amount of points corresponding to 90% correct on the task, or more, whereas men did not. Third, we determined the first block that each participant met or exceeded this criterion.

Average accuracy for each participant for each block. The data were analyzed using a repeated measures ANOVA with Gender (Male, Female) and Reward Structure

(Gains, Losses) between participants and Blocks of trials within participants. This analysis revealed a significant two-way interaction between Gender and Reward Structure, $F(1,76) = 4.93$, $MSE = .005$, $p < .05$ (see Figure 4.3).

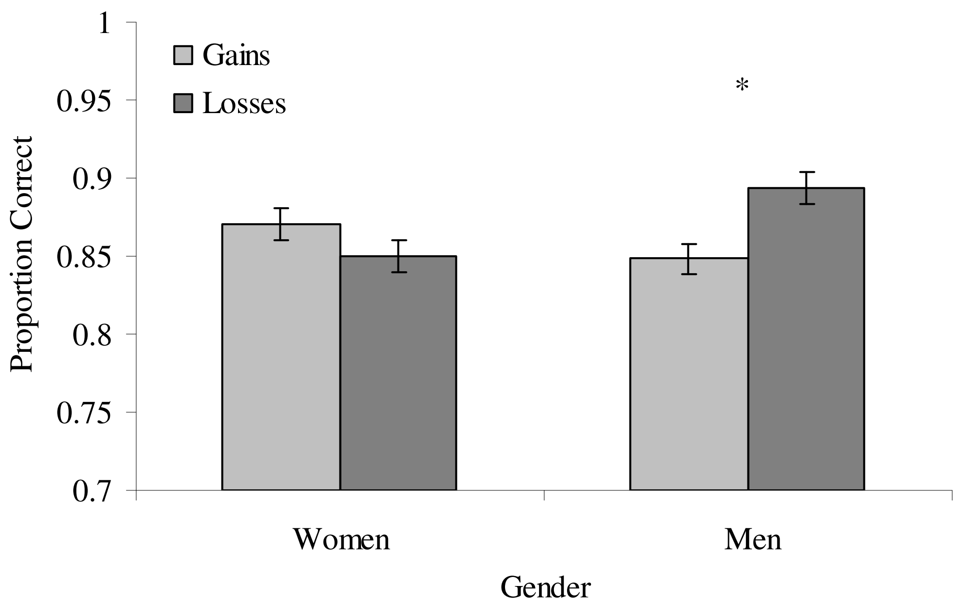


Figure 4.3. Proportion correct for women and men in the gains and losses tasks in Experiment 1

To examine this interaction, we compared the average accuracy scores within each gender for gains and losses. As predicted, men (negative stereotype) who performed the losses task performed significantly better ($M = .89$) than men who performed the gains task ($M = .84$) $t(38) = 1.92$, $p < .05$. As shown in Figure 4.4, moreover, a binomial sign test reveals that men in the losses task performed better than men in the gains task and were more accurate in all 12 experimental blocks, $p < .05$.

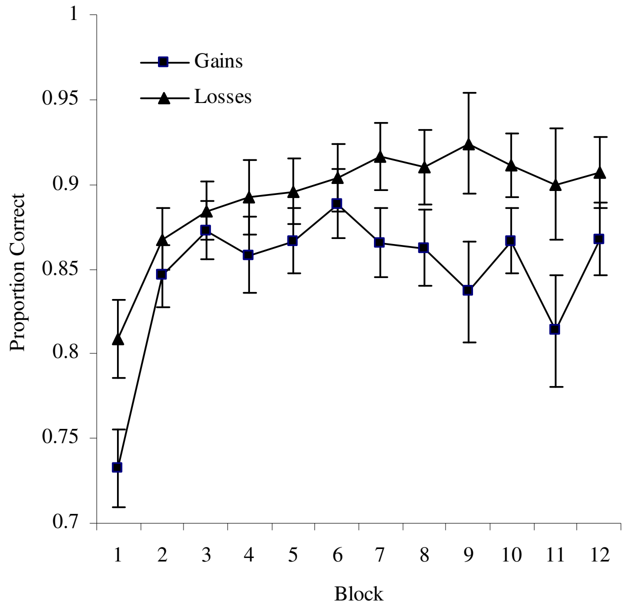


Figure 4.4. Proportion correct across blocks for men in the gains and losses tasks in Experiment 1

Comparing women in the gains (M = .87) and losses conditions (M = .85), revealed a non-significant difference, p = .13. However, as predicted and depicted in Figure 4.5, using a binomial sign test, women (positive stereotype) in the gains task outperformed women in the losses task and were more accurate on 10 of the 12 blocks, $p < .05$.

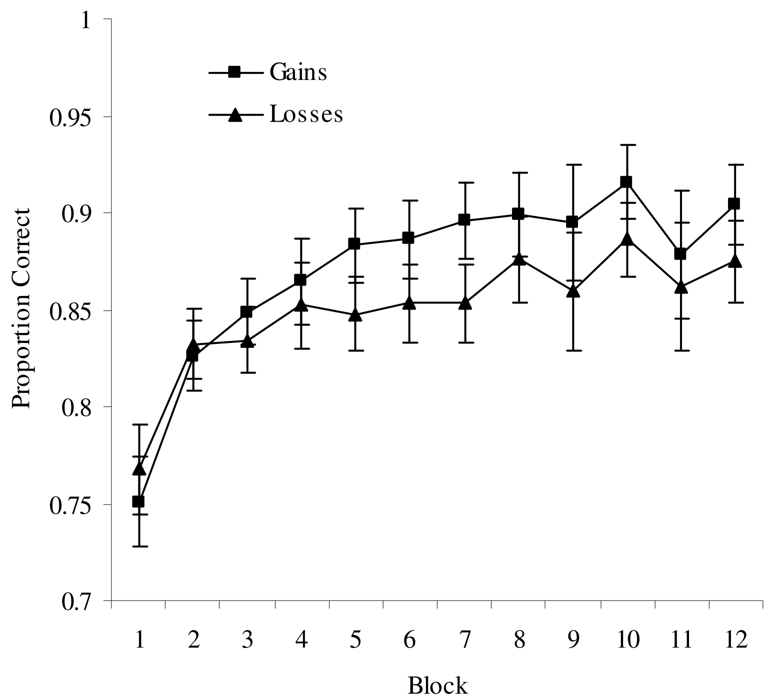


Figure 4.5. Proportion correct across blocks for women in the gains and losses tasks in Experiment 1

Additionally, we examined the interaction in a manner consistent with the extant work on stereotype threat. Instead of looking within men and within women to test the effects of being in a regulatory fit versus being in a regulatory mismatch, we secondarily examined the interaction of Gender and Reward Structure by testing men versus women for the gains and losses structures separately. For gains, there was not a significant main effect of sex. However, a binomial sign test reveals that women outperformed men, $p < .05$. They performed better on 9 of the 12 blocks of trials and performed equally well on one block. For losses, there was a main effect of sex, $F(1,38) = 4.73$, $MSE = .04$, $p < .05$. Men outperformed women on every block. A binomial test reveals this male performance advantage was significant, $p < .05$.

Proportion of participants reaching criterion for each block. We used binomial tests to assess whether the proportion of men who met or exceeded the criterion was larger in the losses task as compared to the gains task. As shown in Figure 4.6, the binomial tests for blocks 1, 6, 7, 9, and 11 revealed the loss advantage, $p < .05$, and the test for block 10 was marginally significant in the same direction, $p = .06$. A binomial sign test across blocks revealed that the male losses condition outperformed the male gains condition, $p < .05$, with a higher proportion of the participants meeting or exceeding the criterion on 11 of the 12 blocks (the 1 remaining block was a tie). Collapsing across blocks, 50% of the participants in the losses task exceeded the criterion as compared to 25% of the participants in the gains task. Using a binomial test, this difference is statistically reliable, $p < .05$.

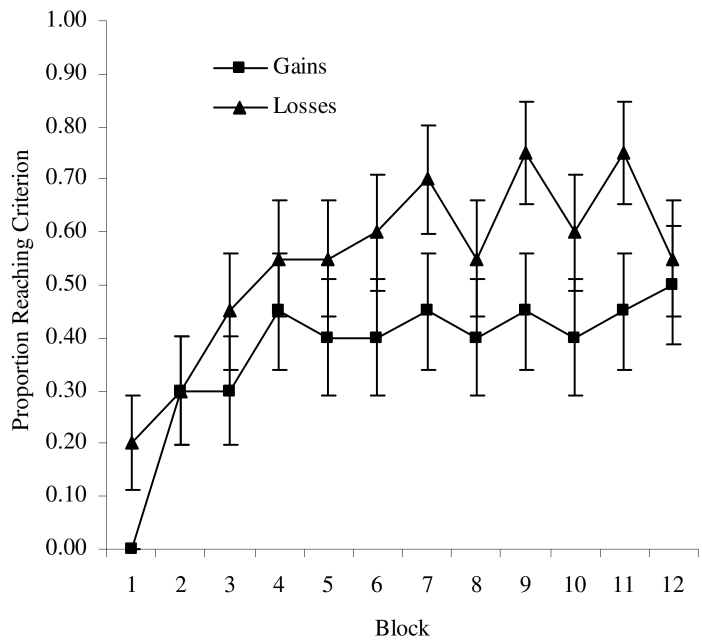


Figure 4.6. Proportion reaching criterion across blocks for men in the gains and losses tasks in Experiment 1

As shown in Figure 4.7, for women in the gains task as compared to women in the losses task, the binomial tests for blocks 5, 6, 7, 12 revealed better performance by women in the gains task, $p < .05$. A binomial sign test across blocks revealed that the women in the gains condition outperformed the women in the losses condition, $p < .05$, with a higher proportion of the participants meeting or exceeding the criterion on 10 of the 12 blocks: 1 remaining block was a tie and in the first block 0% of the gains participants met or exceeded the criterion as compared to 5% of the losses participants. This difference is not statistically reliable. Collapsing across blocks, 30% of the participants in the gains task exceeded the criterion as compared to 25% of the participants in the losses task. Using a binomial test, this difference is not statistically reliable.

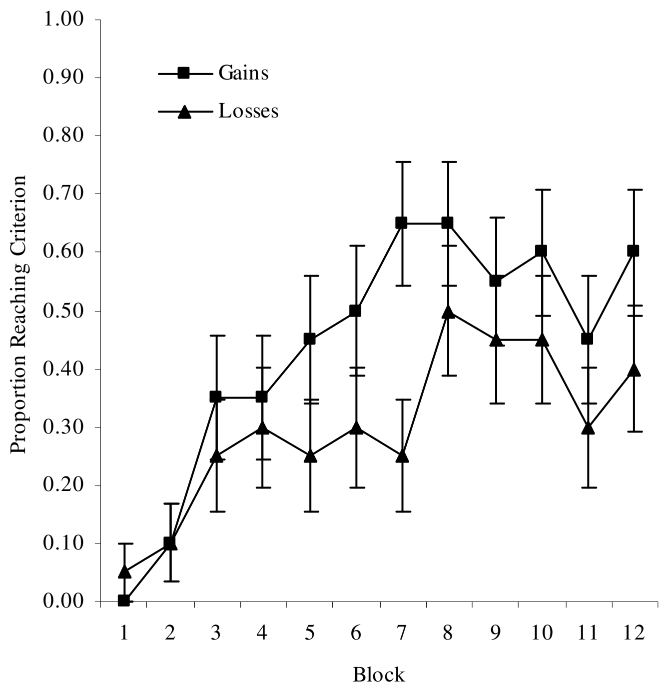


Figure 4.7. Proportion reaching criterion across blocks for women in the gains and losses tasks in Experiment 1

First block exceeding criterion. Any participant who failed to meet the criterion during the experiment was coded as a 13. This was done because this is the minimum value possible for a participant who had not met the criterion during the course of the 12 block experiment. Men in the losses condition exceeded the criterion sooner (after 3.65 blocks on average) as compared to men in the gains condition (after 5.2 blocks on average). This difference was marginally significant, $t(38) = 1.51$, $p = .07$. Women in the gains condition exceeded the criterion sooner (after 4.9 blocks on average) as compared to women in the losses condition (after 6.85 blocks on average), $t(38) = 1.91$, $p < .05$.

Discussion

Using an arbitrarily determined stereotype, we found that women and men responded differently to the gains and losses reward structures using task accuracy and the proportion of participants reaching the task criterion. In this Experiment, we expected stereotype threat-consistent effects for the gains structure. We found results consistent with our interpretation of the stereotype threat literature. Women (given a positive stereotype) outperformed men (given a negative stereotype) in the gains version of the task. Furthermore, we found effects consistent with our notion of regulatory fit. In the gains version of the task, women are experiencing a regulatory match and men are experiencing a regulatory mismatch. The opposite is true for the losses reward structure. Again, as predicted, men outperformed women in the losses version of the task. Men have the regulatory match in the losses task. We also find regulatory fit effects looking within gender. Primed with a negative stereotype, the men in our study performed better on the losses version of the task as compared to the gains version. In contrast, priming a positive stereotype led women perform better in the gains version of the task as compared to the losses version. Importantly, we randomly decided to induce a positive stereotype in women and a negative stereotype in men. However, if we are correct in assuming that the results are due to the interaction between the task environment and induced focus we should be able to completely reverse the effects for men and women by reversing the valence of the stereotypes given to each gender. This is the aim of Experiment 2.

Experiment 2

In Experiment 2, we reverse the stereotype valence given to each gender. Now, we give men a positive stereotype, and women a negative stereotype. We predict that men will perform better in the gains version of the task as compared to the losses version and that women will perform better in the losses version of the task as compared to the gains version. Furthermore, we predict men will perform better than women in the gains version and women will perform better than men in the losses version. To anticipate, the pattern of data supports our hypothesis and demonstrates that participants in a regulatory match perform better on the task as compared to participants in a regulatory mismatch.

Method

Participants

Eighty undergraduate students (40 men and 40 women) at the University of Texas at Austin were given $8 for their participation. Half of the men and half of the women were randomly assigned to the gains and losses reward structures. The questionnaire data for 24 participants (12 men and 12 women) was lost because of a computer error.

Design

This experiment used a 2 (Gender: Male, Female) × 2 (Reward Structure: Gains, Losses) design. Reward Structure was manipulated between subjects.

Stimuli and Stimulus Presentation

We used the same stimuli as Experiment 1. Like Experiment 1, there were 48 trials in each block and 12 blocks.

Materials

We used the Regulatory Focus Questionnaire (RFQ; Higgins et al., 2001), the Beck Anxiety Inventory (BAI; Beck et al. 1988), the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990), and the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988).

Procedure

Participants were tested in individual cubicles. At the beginning of the classification task, participants were told that their job was to learn to classify items into two categories. To induce a stereotype our participants read that "men perform better than women on tests of spatial ability."

In the gains task, participants were told that the men tended to earn more than 86 points per block, which is equivalent to a 90% correct criterion (correct on 43 of 48 trials), and women tended to earn fewer. In the losses task, participants were told that men tended to lose less than 58 points per block, which is again equivalent to a 90% criterion (correct on 43 of 48 trials), and women tended to lose more.

The remainder of the task procedure was identical to Experiment 1.

Results

Again, to test our hypotheses, we analyzed accuracy data and used quantitative models to give us insight into the strategies used by participants during the experimental task. These model results are presented as supplementary support in the next section. We also present the results from the pre- and post-test questionnaires.

Questionnaire Results

All of the significant comparisons in the questionnaire results appear in Table 4.5. Correlations appear in Table 4.6. Prior to the experimental manipulation, our participants completed the RFQ, the PSWQ, and the BAI. There were neither significant interactions of reward structure and gender nor were there effects of reward structure. However, there were effects of gender. Prior to the experimental manipulation, like Experiment 1, women scored higher than men on the Prevention subscale of the RFQ and on the PSWQ. After the manipulation, we collected the Positive and Negative subscales of the PANAS and ratings of liking, motivation, and expected performance. There were no significant differences between our groups on these measures.

	Women	**Men**	**Test**
RFQ: Prevention subscale	17.5 (4.4)	15.2 (3.8)	$t(54) = 2.04, p < .05$
PSWQ	53.7 (13.5)	45.8 (14.8)	$t(54) = 2.09, p < .05$

Table 4.5. Significant comparisons in the Questionnaire data in Experiment 2 with means (standard deviations in parentheses).

After the manipulation and classification task, we found a significant interaction for the Positive Affect subscale of the PANAS. The data were analyzed using an ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses). This analysis revealed a significant two-way interaction between Gender and Reward Structure, $F(1,52) = 5.22$, $MSE = 85.3$, $p < .05$. Women in the losses and gains tasks averaged 15.4 and 23.5, respectively and this difference was significant, $t(26) = 2.8$, $p < .05$. Men in the losses and gains tasks averaged 24.5 and 21.3, respectively.

	Total Accuracy	Promotion	Prevention	PSWQ	BAI	Expectation	Liking	Motivation	Positive PANAS (pre)	Negative PANAS (pre)	Positive PANAS (post)	Negative PANAS (post)	Performance	Performance versus men	Performance versus women
Total Accuracy	1	-0.026	-0.23	0.186	-0.088	0.089	0.012	0.143	0.06	0.041	-0.051	-0.154	.411(**)	.378(**)	0.205
Promotion		1	0.004	-0.156	-0.052	0.081	-0.126	0.033	0.04	-0.084	-0.082	.276(*)	-.328(*)	-0.147	-.269(*)
Prevention			1	-0.1	-0.019	0.091	.284(*)	0.077	0.027	0.035	0.225	-0.024	0.224	0.099	.315(*)
PSWQ				1	.366(**)	0.04	-0.047	0.127	-0.015	.321(*)	-0.089	0.191	0.107	0.201	0.031
BAI					1	-0.068	-0.068	-0.131	-0.01	0.235	-0.175	0.042	-0.161	-0.145	-0.08
Expectation						1	.678(**)	.705(**)	.600(**)	0.179	.370(**)	0.164	.273(*)	0.14	.399(**)
Liking							1	.763(**)	.712(**)	0.256	.471(**)	0.045	.284(*)	0.139	.388(**)
Motivation								1	.705(**)	0.166	.451(**)	0.032	.321(*)	0.219	.395(**)
Positive PANAS (pre)									1	.291(*)	.577(**)	0.195	0.21	0.092	.306(*)
Negative PANAS (pre)										1	0.132	.569(**)	0.095	0.069	0.247
Positive PANAS (post)											1	0.074	.355(**)	0.15	.335(*)
Negative PANAS (post)												1	-0.238	-0.162	-0.061
Performance													1	.761(**)	.772(**)
Performance versus men														1	.589(**)
Performance versus women															1

Note: * $p < .05$, ** $p < .01$.

Table 4.6. Correlations between scales and total accuracy in Experiment 2. Significant correlations appear in grey.

We used ANCOVAs to determine if the significant effects found in the questionnaire data could account for our effects. We started by considering our pre-manipulation questionnaire data: the Promotion and Prevention subscales of the RFQ, the BAI, and the PSWQ. The covariate, Prevention, was significantly related to task accuracy $F(1,48) = 4.12$, $MSE = .006$, $p < .05$. After controlling for Prevention scores, there was a main effect of reward structure, $F(1,48) = 5.0$, $MSE = .006$, $p < .05$. Participants in the gains structure ($M = .86$) performed better than participants in the losses structure ($M = .81$). We secondarily calculated correlations between accuracy and the potential covariates collected after the manipulation. We found that ratings of performance and ratings of performance relative to men were correlated with task accuracy, $r = .41$ and .38, respectively. These ratings were also highly correlated with each other, $r = .76$. When entered independently as covariates, each is significantly related to task accuracy [performance, $F(1,48) = 8.3$, $MSE = .005$, $p < .05$, and relative to men, $F(1,48) = 7.2$, $MSE = .006$, $p < .05$] and remove the effect of our experimental groups. While true covariates, these variables are removing the group effect based on the strong (and unsurprising) relationship that exists between accuracy and subjective performance. Furthermore, this relationship is an artifact of our experimental manipulation.

Accuracy Analysis

As for Experiment 1, we computed the average accuracy for each participant in each block of trials, the proportion of participants reaching criterion (90% accuracy) for each block, and the first block that each participant met or exceeded this criterion. Again, for our task, the criterion was established as part of the stereotype threat manipulation.

Our participants were told that men tended to get the amount of points corresponding to 90% correct on the task, or more, whereas women did not.

Average accuracy for each participant for each block. The data were analyzed using a repeated measures ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses) between participants and Blocks of trials within participants. This analysis revealed a marginally significant two-way interaction between Gender and Reward Structure, $F(1,76) = 3.32$, $MSE = .007$, $p = .07$ (see Figure 4.8).[4]

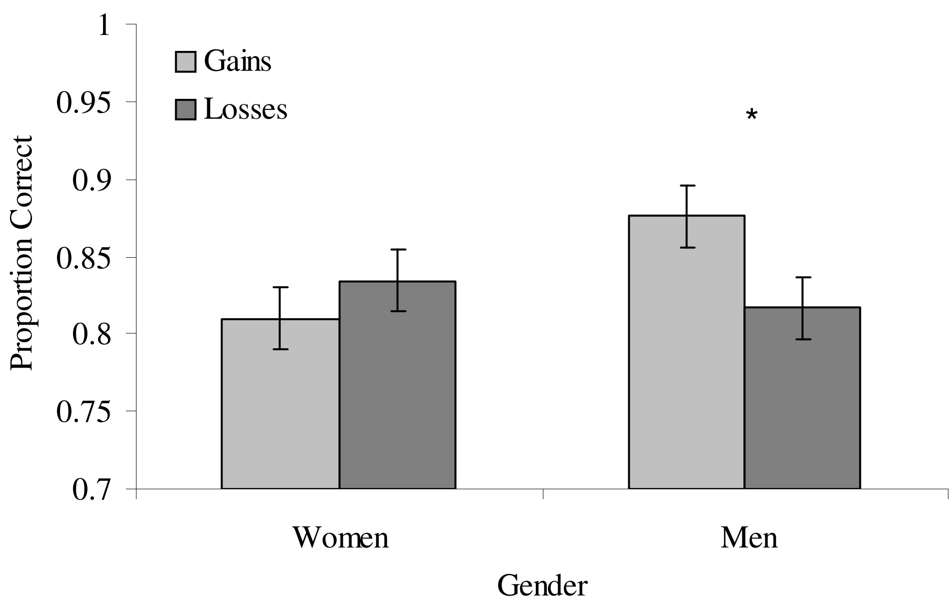


Figure 4.8. Proportion correct for women and men in the gains and losses tasks in Experiment 2

[4] In addition, this analysis revealed a significant two-way interaction between Block and Reward Structure, $F(11,836) = 2.68$, $MSE = .006$, $p < .05$. To examine this interaction, we compared the average accuracy scores within each block for gains and losses. Participants who performed the gains task performed significantly better than participants who performed the losses task in block 6 [t (78) = 2.13, $p < .05$].

To examine this interaction, we compared the average accuracy scores within each gender for gains and losses. As predicted, men (positive stereotype) who performed the gains task performed significantly better (M = .86) than did men who performed the losses task (M = .82) $t(38) = 1.75$, $p < .05$. As shown in Figure 4.9, moreover, a binomial sign test reveals that men in the gains task performed better than men in the losses task, $p < .05$, and in fact obtained higher performance accuracy in 11 of the 12 experimental blocks. The one block where the losses condition performed better than the gains condition was block 1.

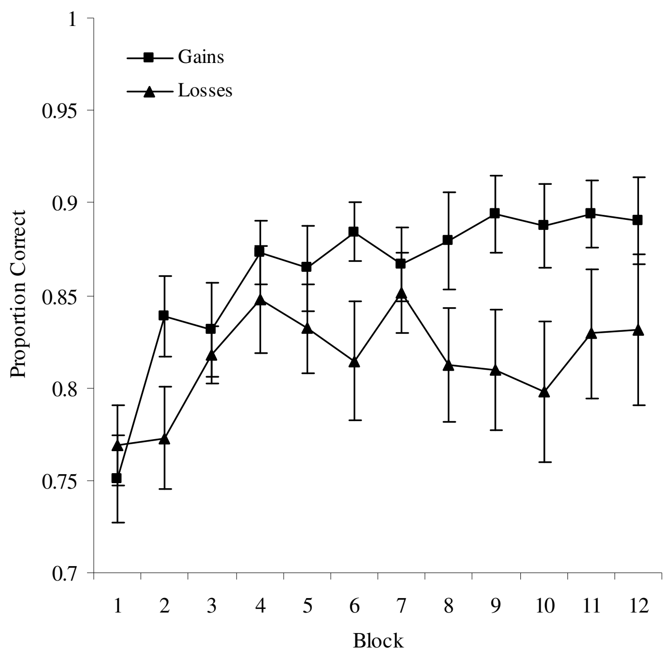


Figure 4.9. Proportion correct across blocks for men in the gains and losses tasks in Experiment 2

For women, there was not a statistically reliable difference for performance on the gains (M = .83) and losses (M = .84) tasks. As depicted in Figure 4.10, women (negative stereotype) in the losses task outperformed women in the gains task and obtained higher accuracy on 8 of the 12 blocks although this difference was not significant using a binomial sign test.

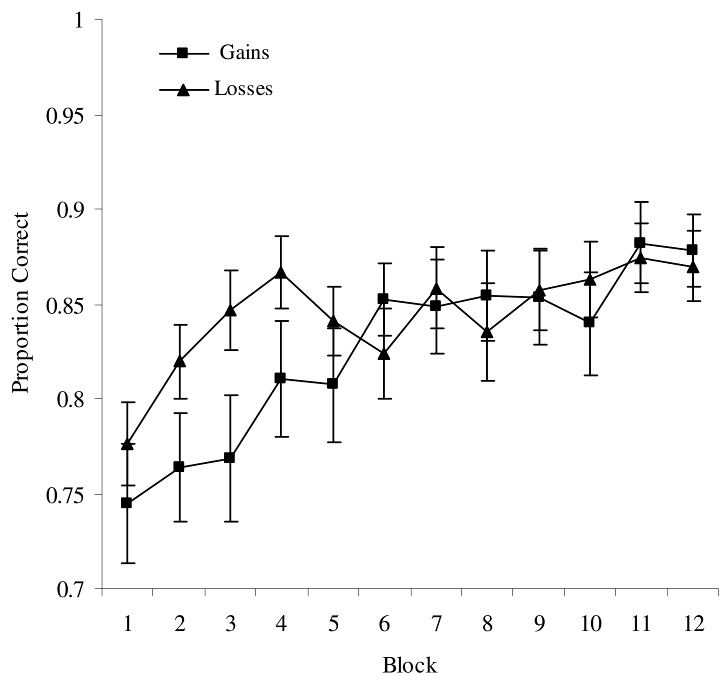


Figure 4.10. Proportion correct across blocks for women in the gains and losses tasks in Experiment 2

Additionally, we examined the interaction of Gender and Reward Structure by testing men versus women for the gains and losses structures separately. This is the method that is consistent with the work on stereotype threat. For gains, there was not a significant main effect of gender. However, a binomial sign test reveals that men outperformed women, $p < .05$. They performed better on all 12 blocks of trials. For losses, there was not a significant main effect of gender. However, women outperformed men on every block. A binomial test reveals this female performance advantage was significant, $p < .05$.

Proportion of participants reaching criterion for each block. We used binomial tests to assess whether the proportion of men who met or exceeded the criterion was larger in the gains task as compared to the losses task. As shown in Figure 4.11, the binomial tests for blocks 2, 5, 6, 8, 9, and 10 revealed this advantage of the gains over the losses task, $p < .05$, and the test for block 3 was marginally significant in the same direction, $p = .06$. A binomial sign test across blocks revealed a marginally significant effect in that the male gains condition outperformed the male losses condition, $p = .07$, with a higher proportion of the participants meeting or exceeding the criterion on 9 of the 12 blocks (2 remaining blocks were a tie and men in the losses condition outperformed men in the gains condition in block 1). Collapsing across blocks, 40% of the participants in the gains task exceeded the criterion as compared to 20% of the participants in the losses task. Using a binomial test, this difference is statistically reliable, $p < .05$.

As shown in Figure 4.12, for women in the losses task as compared to women in the gains task, the binomial test for block 3 revealed the advantage for the losses task, $p < .05$. A binomial sign test across blocks revealed that women in the losses condition outperformed women in the gains condition in 5 of the 12 blocks. This difference is not statistically reliable. Collapsing across blocks, 25% of the participants in the gains task exceeded the criterion as compared to 20% of the participants in the losses task. Using a binomial test, this difference is not statistically reliable.

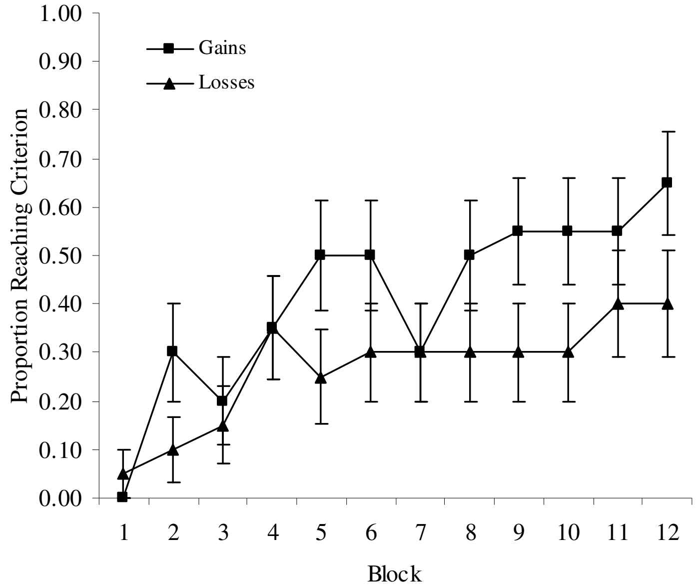


Figure 4.11. Proportion reaching criterion across blocks for men in the gains and losses tasks in Experiment 2

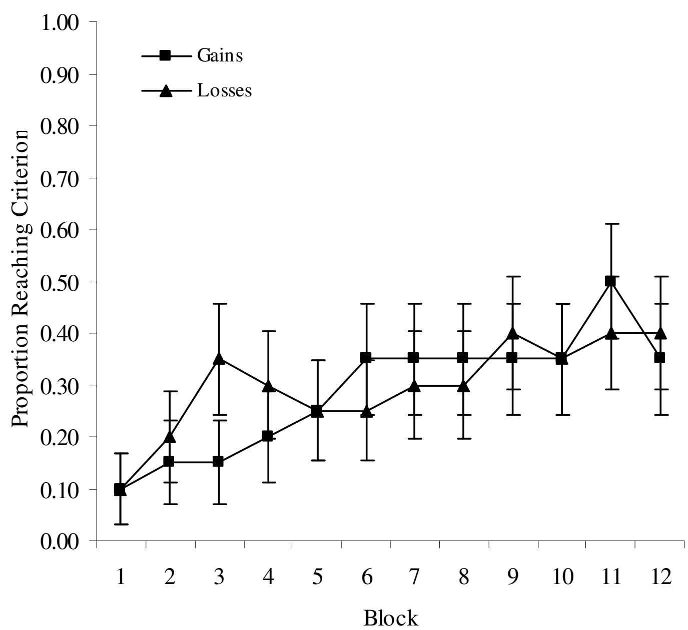


Figure 4.12. Proportion reaching criterion across blocks for women in the gains and losses tasks in Experiment 2

First block exceeding criterion. Again, any participant who failed to meet the criterion during the experiment was coded as a 13. Men in the gains condition exceeded the criterion sooner (after 4.8 blocks on average) as compared to men in the losses condition (after 7.15 blocks on average), $t(38) = 1.91$, $p < .05$. Women in the losses condition exceeded the criterion sooner (after 6.25 blocks on average) as compared to women in the gains condition (after 6.8 blocks on average), but this difference was not statistically reliable.

Discussion

By switching the valence of the stereotypes applied to each gender, we reversed the effects for each gender found in Experiment 1. We found results consistent with our interpretation of the stereotype threat literature. Men outperformed women in the gains version of the task. Furthermore, women outperformed men in the losses version of the task. This performance difference occurs because men and women are experiencing regulatory fits in the gains and losses versions, respectively. Also consistent with regulatory fit, primed with a positive stereotype, the men in our study performed better on the gains version of the task as compared to the losses version. Admittedly, our effects in Experiment 2 for women are not as large as those in Experiment 1. This cross-experimental difference is examined in more detail in the General Discussion section for Experiments 1 and 2 in light of some covariation analyses.

Model-based Analyses

In Experiment 1, we found that women performed better in the gains task as compared to women in the losses task while men in the losses task performed better than men in the gains task. In contrast, in Experiment 2, we found that men performed better in the gains task than men in the losses task. The results for women were more equivocal. Women in the losses task obtained higher accuracy scores than women in the gains task in 8 of the 12 blocks but this difference was not statistically reliable.

An advantage of using this classification task is that we have computational models that allow us to characterize participants' responses on a block-by-block basis. Models allow us to determine the types of strategies used by participants during classification learning instead of being forced to infer strategies from accuracy data. In this section we present data from our model-based analyses.

Following Maddox, Baldwin, & Markman (2006), we hypothesize that participants start with simple unidimensional rules to classify the stimuli and then switch to the more complex conjunctive rule on length and orientation that can provide a means to exceed the 90% accuracy criterion. We believe that participants experiencing a regulatory match will be more likely to abandon the simple rules in favor of the more complex conjunctive rule.

To test this hypothesis, we fit a series of decision-bound models to the data for each participant for each block (Ashby & Maddox, 1993; Maddox & Ashby, 1993). The unidimensional model on position assumes that the participant used a criterion on position and put all of the lines to the left in one category and all of the lines to the right in the other category. The unidimensional model on orientation assumes that the participant's criterion involved one response for shallow lines and another response for steep lines. The unidimensional model on length assumes one response for short lines and another response for long lines. Each of these unidimensional models uses two free parameters: one decision criterion and one noise parameter. The conjunctive model assumes that the participant used length and orientation. We fit two different conjunctive models. First, we fit an optimal model which assumes the participant used the optimal criterion on both length and orientation. This model only has one free noise parameter. Second, we fit a suboptimal model which assumes that the participant used criteria on both length and orientation but these criteria were not optimal. Therefore, this model has three free parameters: one for the length criterion, one for the orientation criterion, and one noise parameter.

The model parameters were estimated using maximum likelihood (Ashby, 1992). We found the best fitting model using: $AIC = 2r - 2lnL$ (Akaike, 1974; Takane & Shibayama, 1992) where r is the number of parameters in the model and lnL is the log likelihood of the model given the data. This criterion allows us to assess the goodness-of-fit of models that differ in the number of free parameters, and select the model that provides the most parsimonious account of the data (i.e., the model with the smallest AIC value).

Before summarizing the conjunctive model results, it is important to determine that the models provided a good account of the data. The suboptimal conjunctive model accounted for an average of 91% and 89% of the total category responses in Experiments 1 and 2, respectively (see Table 4.7).

	Experiment	
Model	**1**	**2**
Optimal Conjunctive	0.87	0.83
Suboptimal Conjunctive	0.91	0.89
Length	0.63	0.61
Orient	0.63	0.63
Position	0.83	0.82

Table 4.7. Percent of total category responses accounted for by each model type

For both experiments, the unidimensional rules on length and orientation were rarely used by participants. The unidimensional length and orientation models best fit the data 5% and 17% of the time, respectively. As such, they will not be discussed further. In contrast, the unidimensional position rule best fit 30 % of the data or more for each of the experimental groups. The conjunctive model fit over 60 % of the data in the final block of trials for all groups in Experiment 1 and over 45 % of the data in Experiment 2. The average accuracy for each participant using the conjunctive rule was approximately 90 % for Experiments 1 and 2.

Experiment 1: Women primed with positive stereotype. Figure 4.13 displays the proportion of data sets best fit by a conjunctive rule model for men in the gains and losses tasks across block. Because men in the losses condition are in a regulatory match relative to men in the gains condition, we predict that a larger proportion of men/losses data sets will be best fit by a conjunctive rule model. This pattern held in 10 of the 12 blocks of trials (significant based on a sign test), and was significant (based on binomial tests) in blocks 7, 8, 9, 10, and 11 $p < .05$.

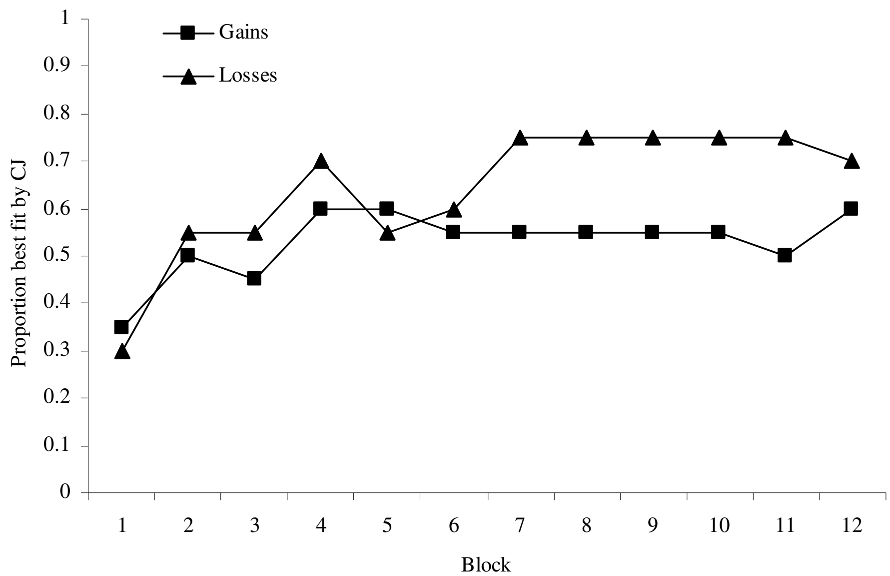


Figure 4.13. Proportion best fit by the correct conjunctive rule for men in the gains and losses tasks in Experiment 1

The opposite pattern was predicted for women. Specifically, women in the gains condition are in a regulatory match and should be more likely to use a conjunctive rule then women in the losses condition who are in a regulatory mismatch. This pattern held in 10 of the 12 blocks of trials (significant based on a sign test), and was significant (based on binomial tests) in blocks 2, 3, 4, 5, 6, 7, 8, 9, 10, and 11, $p < .05$ (see Figure 4.14).

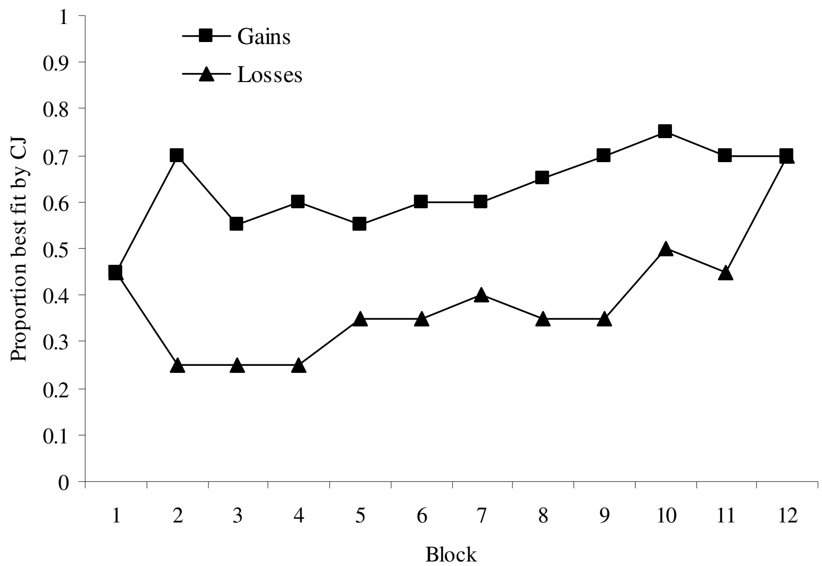
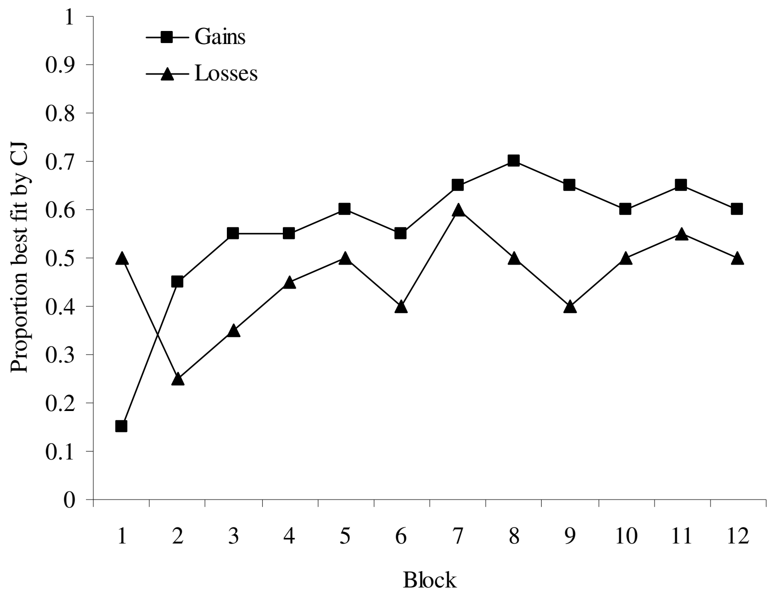


Figure 4.14. Proportion best fit by the correct conjunctive rule for women in the gains and losses tasks in Experiment 1

Experiment 2: Men primed with a positive stereotype. As predicted, for men, as shown in Figure 4.15, the binomial tests for blocks 2, 3, and 9 revealed the gains advantage, $p < .05$, while block 1 showed a loss advantage, $p < .05$. A binomial sign test across blocks revealed that the data in the men/gains condition was better fit by the conjunctive rule more frequently than the data in the men losses condition, $p < .05$, with a higher proportion of the participants likely using the conjunctive rule in 11 of the 12 blocks.



<u>Figure 4.15</u>. Proportion best fit by the correct conjunctive rule for men in the gains and losses tasks in Experiment 2

For women, a binomial test for block 11 revealed more conjunctive rule use likely in the losses task, $p < .05$, while block 6 showed more women in the gains task likely using the rule, $p < .05$ (see Figure 4.16). A binomial sign test across blocks revealed that the women/losses condition was not more likely to use the conjunctive rule than the women/gains condition.

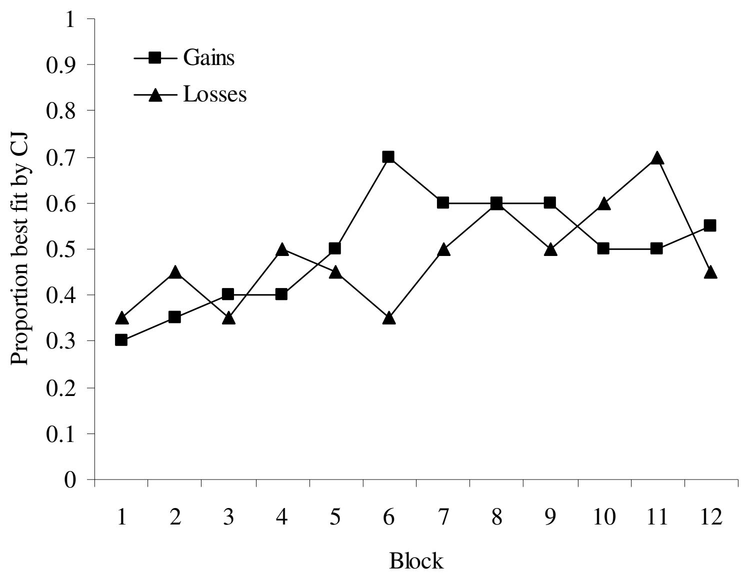


Figure 4.16. Proportion best fit by the correct conjunctive rule for women in the gains and losses tasks in Experiment 2

General Discussion

In Experiments 1 and 2 we found results consistent with the work on stereotype threat looking within the losses and gains tasks. In Experiment 1, women were primed with a task-relevant positive stereotype and men were primed with a task-relevant negative stereotype. Women outperformed men in 9 of the 12 blocks in the gains task. In Experiment 2, we switched the valence of the stereotypes applied to gender and got a predicted performance reversal: men outperformed women in all 12 blocks of trials in the gains task.

We also got effects that support our notion of regulatory fit. Women in Experiment 1 and men in Experiment 2 were experiencing a regulatory match in the gains task whereas the other groups were experiencing a regulatory mismatch. Using the proposed motivation-learning framework (Maddox, Markman, & Baldwin, 2006), we predicted and obtained the predicted results for the losses task. In Experiment 1, men outperformed women in all 12 blocks and in Experiment 2, women outperformed men in all 12 blocks of trials. This result was predicted because participants primed with a negative stereotype are experiencing a regulatory match in the losses task.

We also predicted and obtained results that support regulatory fit looking within gender for the gains and losses tasks. Participants performed better in a task that matched their situationally-primed focus than in a task that mismatched their primed focus.

Our participants completed a classification task in which they learned to classify lines that varied in their length, orientation, and position. Participants could achieve perfect task performance if they learned to classify the lines using a conjunctive rule on both the length and orientation dimensions. To meet the learning criterion, participants needed to switch from using the easier and more obvious unidimensional rules to the more complex conjunctive rule. This rule switching requires cognitive flexibility.

Based on the prior work by Maddox, Baldwin, and Markman (2006), we predicted that individuals experiencing a regulatory match would perform better in the task than participants in a regulatory mismatch. We argued like Seibt and Förster (2004) that priming a negative stereotype induced a prevention focus while priming a positive stereotype induced a promotion focus. After stereotype priming (valence manipulated across experiments), our participants completed a gains version of our task where they gained points for correct responses or a losses version of the task where they lost points for correct responses. For the gains task, we predicted that participants primed with a positive stereotype would be experiencing a regulatory match while participants primed with a negative stereotype would be experiencing a regulatory mismatch. We predicted the opposite would be true for the losses task.

In Experiment 1, we gave women a positive stereotype and men a negative stereotype. As predicted, we found that women performed better in the gains task as compared to women in the losses task. Using task accuracy and the proportion to reach the criterion, women in the gains task performed better in 10 of the 12 experimental blocks. Using the first block that a participant met or exceeded the criterion, women in the gains condition exceeded the criterion sooner than women in the losses condition. Men in the losses task outperformed men in the gains task. They performed better in all 12 blocks of trials. They also reached the performance criterion more often in 11 of the 12 blocks and reached or exceeded the performance criterion sooner.

In Experiment 2, we gave men a positive stereotype and women a negative stereotype. We found men who performed the gains task performed significantly better than men who performed the losses task. They scored higher on 11 of the 12 experimental blocks and more of them reached the criterion in 9 of the 12 blocks. In addition, men in the gains task met or exceed the performance criterion sooner than men in the losses task. For women, our results are not statistically reliable.

The modeling results support the task analyses. In Experiment 1, the female data in the gains task more likely came from conjunctive rule use as compared to the data in the losses task. The reverse was true for men: the data in the losses task was more consistent with conjunctive rule use than the data in the gains task. As predicted, in Experiment 2, the male data for the gains task was more consistent with conjunctive rule use than the data for the losses task. For women, the modeling did not reveal likely differences in conjunctive rule application during classification learning.

Across both experiments, we find data in support of our claim that a regulatory match produces more flexible processing than a regulatory mismatch. Our data from Experiment 1 show the complete cross-over interaction with women in gains performing better than in losses and men in losses performing better than in gains. In Experiment 2, we find men in gains performing better than in losses and find limited support for our prediction that women in losses would perform better than women in gains.

It is possible that women in Experiment 2 were unaffected by the induction of a task-relevant negative stereotype. We believe this explanation is likely given the results of our covariation analyses in Experiment 2. Prior to adding covariates, we found that men in gains (M = .86) performed better than men in losses (M = .82). After adding the Prevention scores, the adjusted male data did not change significantly: average accuracy for gains is .86 and for losses .81. In contrast, the data for women did change significantly. Prior to accounting for the covariate, women in gains (M = .83) did not differ from women in losses (M = .84). However using the Prevention scores, women in gains (M = .86) performed better than women in losses (M = .82). This suggests that women were influenced by their chronic level of focus instead of by our presented situational focus. That is, women did not adopt a situational-prevention focus. Given that women in gains performed better than women in losses, it is possible that women used self-protective strategies to maintain or generate a mild promotion focus.

EXPERIMENTS 3 AND 4: INFORMATION INTEGRATION TASK

Experiments 3 and 4 use an information-integration category structure (e.g., Maddox & Ashby, 2004), and a classification task that requires that participants learn to classify lines into two categories. As shown in Figure 3.2, the stimulus dimensions are the same as those in Experiments 1 and 2, but in this case a plane cuts the stimulus space on the diagonal creating a rule that is difficult to verbalize. It is seemingly nonsensical to describe a rule in which a stimulus goes in a category because it is longer than it is steep because the dimensions are measured by different units.

As demonstrated previously by Maddox, Baldwin, and Markman (2006), this structure requires participants abandon the explicit rule-based system in favor of the implicit learning system. For this reason, flexible processing hurts performance. Participants behaving flexibly are more likely to spend time explicitly testing rules.

I present this category structure to participants in two Experiments. In both Experiments, half of the participants gain points for correct responses and half lose points for correct responses. In Experiment 3, I prime a positive stereotype for women and a negative stereotype for men. In Experiment 4, I reverse the gender assigned to positive and negative stereotypes. To increase the size of the effects found in Experiment 1 and 2, I use a more powerful stereotype manipulation. Participants have the instructions and threat manipulation auditorally and visually presented. Furthermore, the stereotype manipulation includes "facts" to support the threat claims.

As in Experiments 1 and 2, I predict that a positive stereotype will induce a situational promotion focus and a negative stereotype will induce a situational prevention focus. Further, I predict that individuals experiencing a regulatory match (e.g., positive stereotype with gains and negative stereotype with losses) will perform worse than individuals experiencing a regulatory mismatch (e.g., positive with losses and negative with gains). That is, the match participants will continue to test rules that cannot yield good enough performance. In sum, I expect to completely reverse the effects found in Experiments 1 and 2. In the gains version, these results would replicate Seibt and Förster (2004) Experiment 2.

These results would provide critical insight into stereotype threat mechanisms. Across all four Experiments, I would demonstrate that the influence of a negative stereotype depends both on the reward structure of the task (creating a regulatory match or mismatch) and the degree of flexibility required for good task performance. To anticipate, the pattern of data in these Experiments is not entirely supportive of my hypotheses. I do not find consistent results in the accuracy data. However, I find some evidence for predicted strategy use in Experiment 3 and the complete interaction for strategy as predicted in model-based analyses in Experiment 4.

Experiment 3

In Experiment 3, I give women a positive stereotype and men a negative stereotype. I predict that women will perform better in the losses version of the task as compared to the gains version and that men will perform better in the gains version of the task as compared to the losses version. Furthermore, I predict women will perform better than men in the losses version and men will perform better than women in the gains version. This pattern of data is consistent with my hypotheses because it would demonstrate that participants in a regulatory mismatch perform better on the task as compared to participants in a regulatory match. The match participants are predicted to persist in using the explicit testing system which will hurt their performance on the information integration task.

Method

Participants

Eighty undergraduate students (40 men and 40 women) at the University of Texas at Austin were given $8 for their participation. Half of the men and half of the women were randomly assigned to the gains and losses reward structures.

Design

This experiment used a 2 (Gender: Male, Female) × 2 (Reward Structure: Gains, Losses) design. Reward Structure was manipulated between subjects.

Stimuli and Stimulus Presentation

I used the information-integration stimuli shown in Figure 3.2. Other than the change in stimuli to form categories consistent with an information integration rule, stimuli selection and presentation were identical to Experiments 1 and 2. There were 48 trials in each block and 12 blocks.

Materials

I used the Regulatory Focus Questionnaire (RFQ; Higgins et al., 2001), the Beck Anxiety Inventory (BAI; Beck et al. 1988), the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990), and the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988).

Procedure

Participants were tested in individual cubicles and all instructions were presented on the computer screen and auditorally using headphones. At the beginning of the classification task, participants were told that their job was to learn to classify items into two categories. To induce a stereotype participants read that women perform better than men on tests of spatial ability. For example, participants in the losses task were presented with:

> This is an experiment testing sex differences in spatial abilities. Recent research (Simon & Small, 2005) found that college students differ in their ability to think spatially. Some students are better able to mentally separate objects into pieces. These students easily separate objects correctly and finish quickly. In a large review of 286 studies, Voyer, Voyer, and Bryden (1995) determined that women perform better than men across a wide-range of spatial tasks.
>
> Researchers have speculated on the aspects of our evolutionary environment that would have led to these differences. For example, in early hunter-gatherer societies women were responsible for gathering berries and vegetables for consumption, and so they had to develop accurate spatial maps of their

environment. This association between gender-roles and the early evolutionary environment may have led to differences in male and female brain development as well as differences in the brain's response to estrogen and testosterone.

We are interested in further examining this sex difference.

In this experiment, you will lose only one point for each correct response and lose three points for each incorrect response. In this task, women tend to lose no more than 58 points per block of trials and men tend to lose more than 58 points per block of trials.

Please try your best in this task. Before continuing, please indicate whether you are male or female. If you are male, press the "M" key. If you are female, press the "F" key.

In the gains task, participants were told that women tended to earn more than 86 points per block, which is equivalent to a 90% correct criterion (correct on 43 of 48 trials), and men tended to earn fewer. In the losses task, participants were told that women tended to lose less than 58 points per block, which is again equivalent to a 90% criterion (correct on 43 of 48 trials), and men tended to lose more.

The remainder of the task procedure was identical to Experiments 1 and 2.

Results

To test the hypotheses, I analyzed accuracy data and used quantitative models to give me insight into the strategies used by participants during the experimental task. I also present the results from the pre- and post-test questionnaires. The model fits are presented in a supplementary section after the results from Experiment 4.

Questionnaire Results

All of the significant group comparisons in the questionnaire data appear in Table 4.8. Correlations appear in Table 4.9. Prior to the experimental manipulation, participants completed the RFQ, the PSWQ, and the BAI. There were neither significant interactions of reward structure and gender nor were there main effects of reward structure. However, there were main effects of gender. As in Experiments 1 and 2, women scored higher than men on the Prevention subscale of the RFQ and on the PSWQ.

After the manipulation, I collected the Positive and Negative subscales of PANAS and ratings of liking, motivation, and expected performance. The data were analyzed using an ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses). This analysis revealed a marginally significant two-way interaction between Gender and Reward Structure for motivation, $F(1,76) = 3.7$, $MSE = 2.65$, $p = .06$. Women in the gains and losses tasks averaged 7.7 and 6.7, respectively, and this difference was marginally significant, $t(38) = 1.8$, $p = .07$. Men in the gains and losses tasks averaged 6.7 and 7.2, respectively.

	Women	Men	Test
RFQ: Prevention subscale	18.4 (3.3)	16.5 (4.0)	$t(78) = 2.37$, $p < .05$
PSWQ	53.3 (13.0)	41.2 (12.4)	$t(78) = 4.27$, $p < .07$

Table 4.8. Significant comparisons in the Questionnaire data in Experiment 3 with means (standard deviations in parentheses).

After the manipulation and classification task, I collected the PANAS and ratings of perceived performance. I found marginally significant interactions for the performance ratings when the comparison group was not specified and when the comparison group was women. The data were analyzed using an ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses). This analysis revealed a marginally significant two-way interaction between Gender and Reward Structure for no specified comparison group, $F(1,76) = 2.88$, $MSE = 2.7$, $p = .09$. Women in the gains and losses tasks averaged 5.3 and 4.5, respectively. Men in the gains and losses tasks averaged 5.9 and 6.3, respectively. There was also a main effect of gender, $F(1,76) = 10.19$, $MSE = 2.7$, $p < .05$. Women rated their performance ($M = 4.9$) lower than men ($M = 6.1$).

A second analysis revealed a marginally significant two-way interaction between Gender and Reward Structure when women was the comparison group, $F(1,76) = 3.3$, $MSE = 3.2$, $p = .07$. Women in the gains and losses tasks averaged 5.2 and 3.6, respectively, and this difference was significant, $t(38) = 2.5$, $p < .05$. Men in the gains and losses tasks averaged 4.7 and 4.8, respectively.

ANCOVAs revealed that none of the significant effects found in the questionnaire data produced significant main effects or interactions in the accuracy data after being accounted for.

	Total Accuracy	Promotion	Prevention	PSWQ	BAI	Expectation	Liking	Motivation	Positive PANAS (pre)	Negative PANAS (pre)	Positive PANAS (post)	Negative PANAS (post)	Performance	Performance versus men	Performance versus women
Total Accuracy	1	0.085	0.12	-0.177	-0.033	0.198	-0.058	-0.002	0.043	-0.135	0.084	-.237(*)	.265(*)	.243(*)	0.201
Promotion		1	0.126	-0.181	-0.068	0.114	-0.069	0.019	0.016	0.005	-0.094	-0.061	0.09	0.068	0.079
Prevention			1	0.015	-0.089	0.079	-0.028	0.079	0.021	-0.219	0.134	-0.188	0.015	0.121	-0.066
PSWQ				1	.447(**)	-.357(**)	-0.129	-0.009	-0.093	.368(**)	-0.019	.284(*)	-.298(**)	-.252(*)	-.274(*)
BAI					1	-0.016	-0.02	-0.05	0.021	.414(**)	-0.094	.262(*)	-0.136	-.227(*)	-0.088
Expectation						1	.457(**)	.385(**)	.251(*)	-0.198	0.137	-0.003	.394(**)	.369(**)	.389(**)
Liking							1	.590(**)	.387(**)	-0.07	.255(*)	-0.043	.239(*)	0.119	0.107
Motivation								1	.511(**)	-0.122	.320(**)	0.07	0.189	0.021	0.117
Positive PANAS (pre)									1	-0.019	.588(**)	0.052	.274(*)	0.212	0.192
Negative PANAS (pre)										1	0.012	.457(**)	-0.125	-0.079	-0.036
Positive PANAS (post)											1	-0.059	.408(**)	.246(*)	.280(*)
Negative PANAS (post)												1	-0.163	-0.094	-0.062
Performance													1	.517(**)	.620(**)
Performance versus men														1	.651(**)
Performance versus women															1

Note: * $p < .05$, ** $p < .01$.

Table 4.9. Correlations between scales and total accuracy in Experiment 3. Significant correlations appear in grey.

Accuracy Analysis

As for Experiments 1 and 2, I computed the average accuracy for each participant in each block of trials, the proportion of participants reaching criterion (90% accuracy) for each block, and the first block that each participant met or exceeded this criterion. Again, the criterion was established as part of the stereotype threat manipulation. Participants were told that women tended to get the amount of points corresponding to 90% correct on the task, or more, whereas men did not.

Average accuracy for each participant for each block. The data were analyzed using a repeated measures ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses) between participants and Blocks of trials within participants. There were neither significant interactions nor main effects. The data averaged across trials is shown in Figure 4.17. Within women and men, there was not a statistically reliable difference for performance on the gains (M = .82 and M = .82, respectively) and losses (M = .80 and M = .81, respectively) tasks.

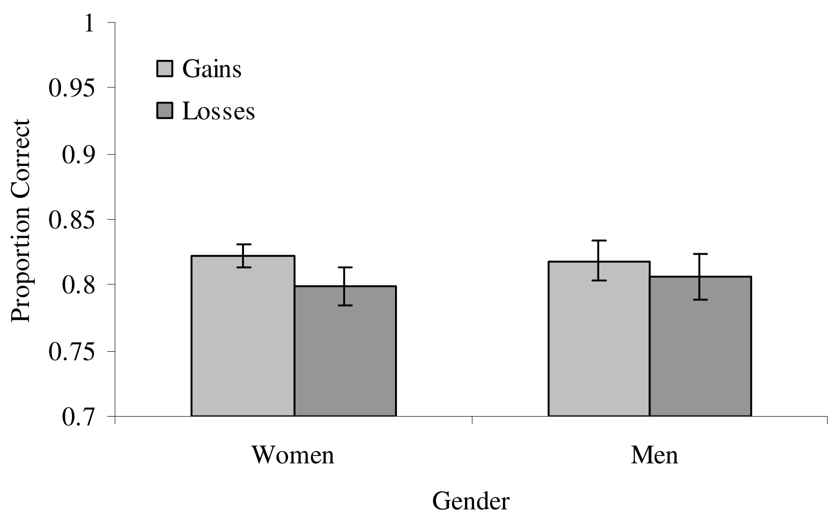


Figure 4.17. Proportion correct for women and men in the gains and losses tasks in Experiment 3

As shown in Figures 4.18 and 4.19, binomial sign tests find men (negative stereotype) in the gains task did not perform better than men in the losses task but women in the gains tasks did perform better than women in the losses task, $p < .05$. Women (positive stereotype) in the gains task outperformed women in the losses task and obtained higher accuracy on 10 of the 12 blocks.

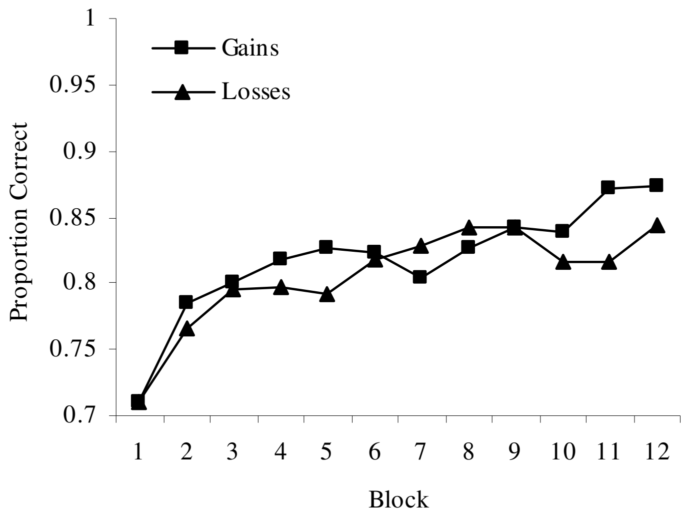


Figure 4.18. Proportion correct across blocks for men in the gains and losses tasks in Experiment 3

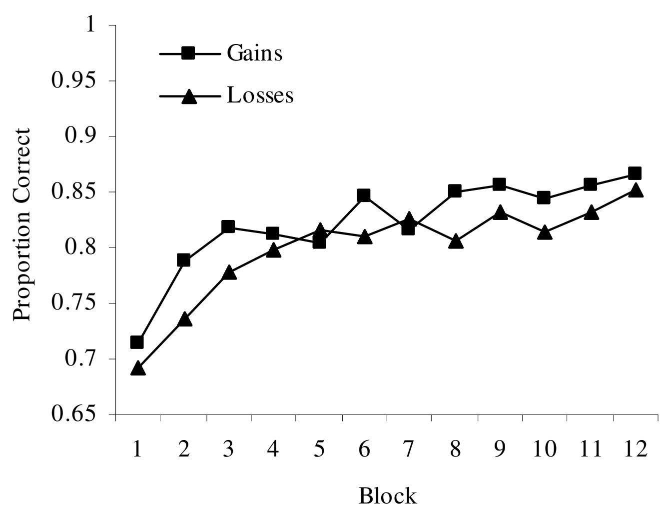


Figure 4.19. Proportion correct across blocks for women in the gains and losses tasks in Experiment 3

Additionally, I examined the interaction of Gender and Reward Structure by testing men versus women for the gains and losses tasks. This is the method that is consistent with the work on stereotype threat. For gains and losses, there was not a significant main effect of gender. Binomial tests also reveal no difference between men and women in gains or losses.

Proportion of participants reaching criterion for each block. I used binomial tests to assess whether the proportion of men who met or exceeded the criterion was larger in the gains task as compared to the losses task. As shown in Figure 4.20, the binomial tests for blocks 4, 11, and 12 revealed this advantage of the gains over the losses task, $p < .05$, and the tests for blocks 2 and 9 were marginally significant in the same direction, $p = .08$ and .07, respectively. A binomial sign test across blocks revealed no difference between the gains and losses tasks.

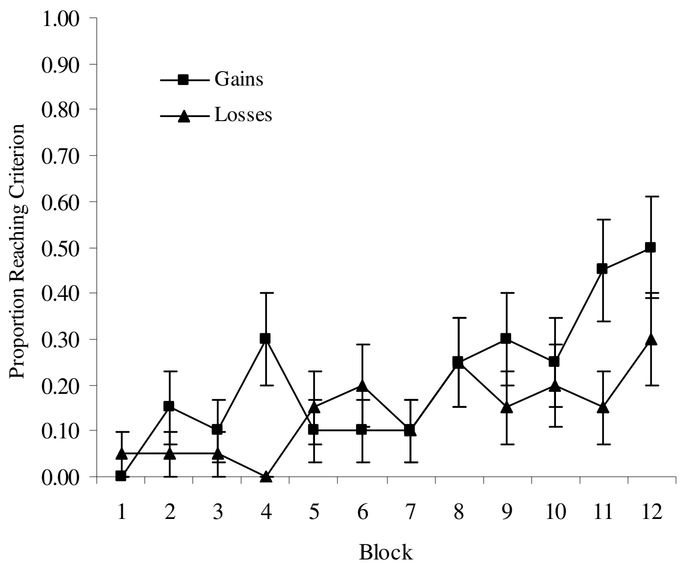


Figure 4.20. Proportion reaching criterion across blocks for men in the gains and losses tasks in Experiment 3

As shown in Figure 4.21, for women in the losses task as compared to women in the gains task, binomial tests revealed no differences. A binomial sign test across blocks also revealed no differences.

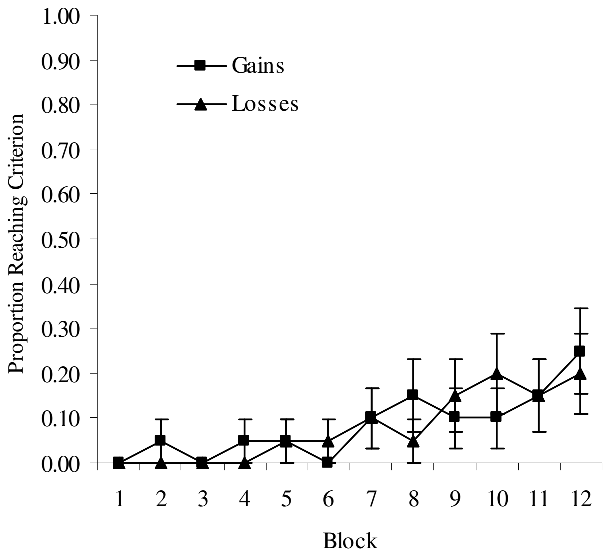


Figure 4.21. Proportion reaching criterion across blocks for women in the gains and losses tasks in Experiment 3

Analyzing the data in a manner consistent with stereotype threat, binomial tests reveal that men in the gains task (mismatch group) outperformed women in the gains task (match) in blocks 4, 9, 10, 11, and 12, $p < .05$, with a marginal effect in the same direction in block 2, $p = .08$. A binomial sign test also shows that men outperformed women in the gains task, $p < .05$, in 10 of the 12 experimental blocks.

In the losses task, women (mismatch) did not outperform men (match) as expected. In fact, the reverse is true. Binomial tests show that men outperformed women in blocks 6 and 8, $p < .05$, and performed better across blocks according to a sign test, $p < .05$. Men outperformed women in 7 of 12 experimental blocks and equaled women in the remaining 5 blocks.

First block exceeding criterion. Any participant who failed to meet the criterion during the experiment was coded as a 13. Men in the gains task exceeded the criterion sooner (after 8 blocks on average) as compared to men in the losses task (after 9.85 blocks on average). This difference is marginally significant, $t(38) = 1.4$, $p = .085$. Women in the gains task exceeded the criterion sooner (after 10.15 blocks on average) as compared to women in the losses task (after 11 blocks on average), but this difference was not statistically reliable.

Discussion

The goal of Experiment 3 was to reverse the effects for each gender found in Experiment 1. In Experiments 1 and 3, all participants were told women are better at the task. Women were experiencing a match in the gains task; men were experiencing a match in the losses task. This match was beneficial in Experiment 1 but was predicted to be disadvantageous in Experiment 3. I found results consistent with my interpretation of the stereotype threat literature. As measured by the proportion of participants reaching the criterion, men outperformed women in the gains version of the task. However, men outperformed women in the losses version of the task as well. Also consistent with regulatory fit theory, the men reached the criterion sooner on the gains version of the task as compared to the losses version. Inconsistent with my predictions, across blocks, women in the gains task performed better than those in the losses task.

Experiment 4

In Experiment 4, I give men a positive stereotype and women a negative stereotype. I predict that men will perform better in the losses task as compared to the gains task and that women will perform better in the gains task as compared to the losses task. Furthermore, I predict men will perform better than women in the losses task and women will perform better than men in the gains task. Like Experiment 3, participants in a regulatory mismatch should perform better on the task as compared to participants in a regulatory match. Based on the primed stereotype, I expect to reverse the effects found in Experiment 3. Further, based on the task performed, I expect to reverse the effects found in Experiment 2.

Method

Participants

One hundred undergraduate students (50 men and 50 women) at the University of Texas at Austin were given $8 for their participation. Half of the men and half of the women were randomly assigned to the gains and losses reward structures.

Design

This experiment used a 2 (Gender: Male, Female) × 2 (Reward Structure: Gains, Losses) design. Reward Structure was manipulated between subjects.

Stimuli and Stimulus Presentation

I used the information integration stimuli shown in Figure 3.2. The stimuli selection and presentation were identical to Experiment 3. There were 48 trials in each block and 12 blocks.

Materials

I used the Regulatory Focus Questionnaire (RFQ; Higgins et al., 2001), the Beck Anxiety Inventory (BAI; Beck et al. 1988), the Penn State Worry Questionnaire (PSWQ; Meyer et al., 1990), and the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988).

Procedure

Participants were tested in individual cubicles and all instructions were presented on the computer screen and auditorally using headphones. At the beginning of the classification task, participants were told that their job was to learn to classify items into two categories. To induce a stereotype our participants read that men perform better than women on tests of spatial ability. For example, participants in the gains task were presented with:

> This is an experiment testing sex differences in spatial abilities. Recent research (Simon & Small, 2005) found that college students differ in their ability to think spatially. Some students are better able to mentally separate objects into pieces. These students easily separate objects correctly and finish quickly. In a large review of 286 studies, Voyer, Voyer, and Bryden (1995) determined that men perform better than women across a wide-range of spatial tasks.

Researchers have speculated on the aspects of our evolutionary environment that would have led to these differences. For example, in early hunter-gatherer societies men were responsible for gathering wild game and fish for consumption, and so they had to develop accurate spatial maps of their environment. This association between gender-roles and the early evolutionary environment may have led to differences in male and female brain development as well as differences in the brain's response to estrogen and testosterone.

We are interested in further examining this sex difference.

In this experiment, you will earn some points for correct responses and no points for incorrect responses. In this task, men tend to earn more than 86 points per block of trials and women tend to earn fewer than 86 points per block of trials.

Please try your best in this task. Before continuing, please indicate whether you are male or female. If you are male, press the "M" key. If you are female, press the "F" key.

In the gains task, participants were told that the men tended to earn more than 86 points per block, which is equivalent to a 90% correct criterion (correct on 43 of 48 trials), and women tended to earn fewer. In the losses task, participants were told that men tended to lose less than 58 points per block, which is again equivalent to a 90% criterion (correct on 43 of 48 trials), and women tended to lose more.

The remainder of the task procedure was identical to Experiment 3.

Results

To test the hypotheses, I analyzed accuracy data and used quantitative models to give me insight into the strategies used by participants during the experimental task. I also present the results from the pre- and post-test questionnaires. The modeling results appear in a supplementary section following the discussion of this experiment.

Questionnaire Results

All of the significant comparisons in the questionnaire results appear in Table 4.10. Correlations appear in Table 4.11. Prior to the experimental manipulation, participants completed the RFQ, the PSWQ, and the BAI. There were neither significant interactions of reward structure and gender nor main effects of reward structure. However, there were main effects of gender. Prior to the experimental manipulation, women scored higher than men on the Prevention subscale of the RFQ and on the PSWQ. Post the manipulation, I collected the Positive and Negative subscales of PANAS and ratings of liking, motivation, and expected performance. Men expected to do better and predicted they would like the task more than women. I also found a significant interaction for the Positive Affect subscale of the PANAS. The data were analyzed using an ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses). This analysis revealed a significant two-way interaction between Gender and Reward Structure, $F(1,96) = 4.28$, $MSE = 73.1$, $p < .05$. Women in the losses and gains tasks averaged 33.8 and 27.7, respectively, and this difference was significant, $t(48) = 2.7$, $p < .05$. Men in the losses and gains tasks averaged 31.1 and 30.2, respectively.

	Women	Men	Test
RFQ: Prevention subscale	19.3 (3.6)	16.7 (3.9)	$t(98) = 3.38$, $p < .05$
PSWQ	50.4 (13.4)	45.3 (14.5)	$t(98) = 1.82$, $p = .07$
Expected Performance	6.0 (1.9)	7.1 (1.2)	$t(98) = 3.4$, $p < .05$
Predicting Liking	5.1 (1.7)	6.0 (1.9)	$t(98) = 2.3$, $p < .05$

Table 4.10. Significant comparisons in the Questionnaire data in Experiment 4 with means (standard deviations in parentheses).

After the manipulation and classification task, I collected the PANAS and ratings of perceived performance. I found a significant interaction for the performance ratings when the comparison group was women. The data were analyzed using an ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses). This analysis revealed a significant two-way interaction between Gender and Reward Structure, $F(1,96) = 8.1$, $MSE = 2.7$, $p < .05$. Women in the losses and gains tasks averaged 6.3 and 7.0, respectively. Men in the losses and gains tasks averaged 6.6 and 5.4, respectively and this difference was significant, $t(48) = 2.3$, $p < .05$. ANCOVAs revealed that none of the significant effects found in the questionnaire data produced significant main effects or interactions in the accuracy data after being accounted for.

	Total Accuracy	Promotion	Prevention	PSWQ	BAI	Expectation	Liking	Motivation	Positive PANAS (pre)	Negative PANAS (pre)	Positive PANAS (post)	Negative PANAS (post)	Performance	Performance versus men	Performance versus women
Total Accuracy	1	0.007	0.052	-0.023	-0.094	-0.065	-0.141	-0.033	-0.094	0.016	0.093	-.226(*)	.235(*)	0.14	0.167
Promotion		1	-0.074	-.286(**)	-0.138	.334(**)	-0.007	.220(*)	.256(*)	0.028	.273(**)	0.039	0.148	.254(*)	.202(*)
Prevention			1	0.088	-0.099	0.01	0.171	.236(*)	.213(*)	-0.046	0.002	-0.075	0.064	0.188	0.149
PSWQ				1	.362(**)	-.205(*)	0.085	0.001	-0.107	.316(**)	-0.175	.204(*)	-0.09	-0.189	-0.082
BAI					1	-0.059	.219(*)	0.036	0.014	.362(**)	-0.04	.368(**)	-0.043	-0.131	-0.041
Expectation						1	.353(**)	.309(**)	.247(*)	-0.16	0.002	0.027	.228(*)	0.126	0.156
Liking							1	.432(**)	.467(**)	0	0.195	0.106	0.058	-0.009	0.038
Motivation								1	.631(**)	-0.111	.213(*)	0.005	0.154	0.053	0.19
Positive PANAS (pre)										0.105	.570(**)	0.121	0.174	0.113	0.113
Negative PANAS (pre)										1	0.172	.478(**)	-0.183	-0.093	-0.139
Positive PANAS (post)											1	-0.062	.357(**)	.210(*)	0.188
Negative PANAS (post)												1	-.303(**)	-0.104	-0.168
Performance													1	.561(**)	.493(**)
Performance versus men														1	.557(**)
Performance versus women															1

Note: * $p < .05$, ** $p < .01$.

Table 4.11. Correlations between scales and total accuracy in Experiment 4. Significant correlations appear in grey.

Accuracy Analysis

As for Experiments 1, 2, and 3, I computed the average accuracy for each participant in each block of trials, the proportion of participants reaching criterion (90% accuracy) for each block, and the first block that each participant met or exceeded this criterion. Again, the criterion was established as part of the stereotype threat manipulation. Participants were told that men tended to get the amount of points corresponding to 90% correct on the task, or more, whereas women did not.

Average accuracy for each participant for each block. The data were analyzed using a repeated measures ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses) between participants and Blocks of trials within participants. As shown in Figure 4.22, there was neither a significant interaction nor significant main effects of Gender or Reward Structure. Within men and women, there was not a statistically reliable difference for performance on the gains (M = .82 and M = .84, respectively) and losses (M = .80 and M = .81, respectively) tasks.

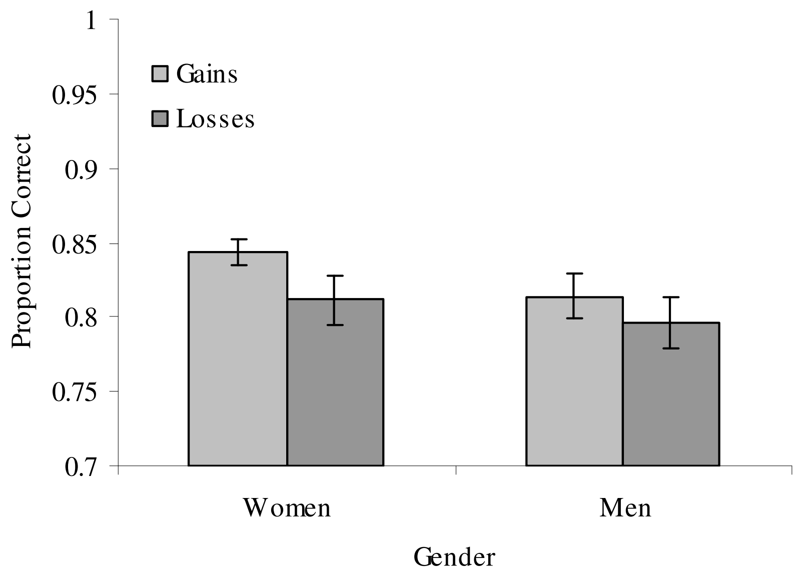


Figure 4.22. Proportion correct for women and men in the gains and losses tasks in Experiment 4

Using a binomial sign test, as shown in Figure 4.23, men in the gains task performed better than men in the losses task, $p < .05$, and in fact obtained higher performance accuracy in 10 of the 12 experimental blocks. Similarly, a binomial sign test reveals that women in the gains task performed better than women in the losses task, $p < .05$ (see Figure 4.24), in all 12 of the experimental blocks.

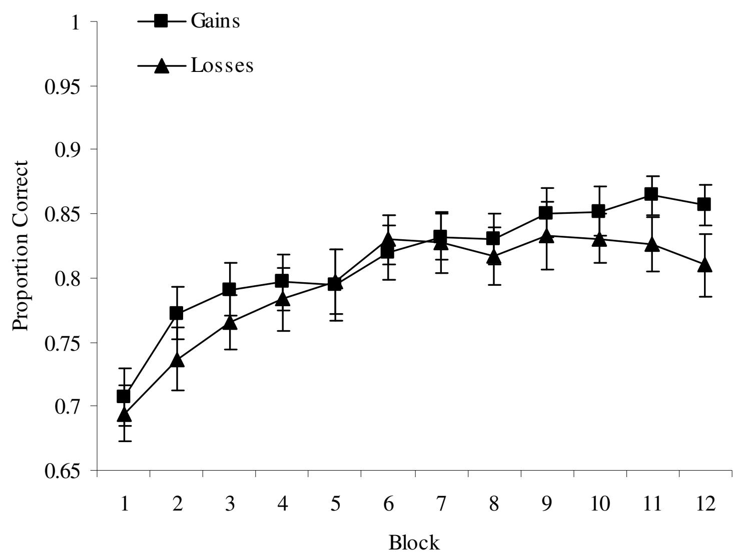


Figure 4.23. Proportion correct across blocks for men in the gains and losses tasks in Experiment 4

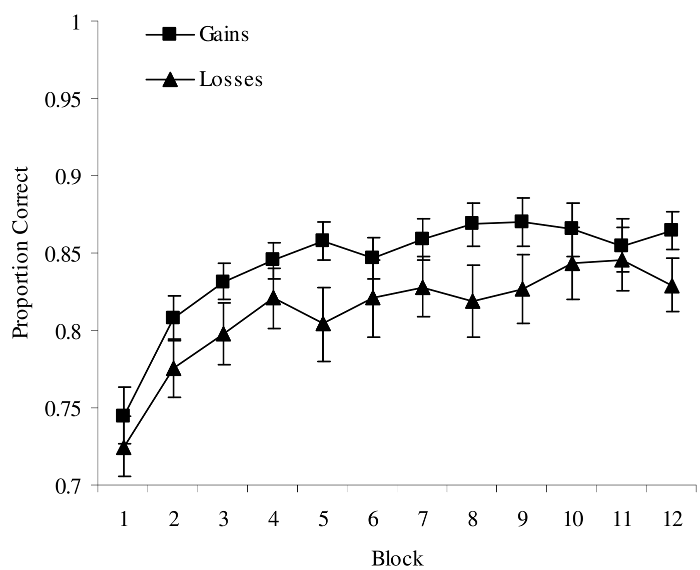


Figure 4.24. Proportion correct across blocks for women in the gains and losses tasks in Experiment 4

Additionally, I used binomial sign tests to explore the accuracy data in a manner consistent with work on stereotype threat. For gains, a binomial sign test reveals that women outperformed men, $p < .05$. They performed better on 11 of 12 blocks of trials. For losses, women outperformed men on 10 of the 12 blocks, $p < .05$.

Proportion of participants reaching criterion for each block. I used binomial tests to assess whether the proportion of men who met or exceeded the criterion was larger in the losses as compared to the gains task. As shown in Figure 4.25, the binomial test for block 8 revealed an advantage of the losses over the gains task, $p < .05$, and the test for block 12 revealed an advantage of the gains over the losses task, $p < .05$. A binomial sign test across blocks did not reveal any differences. Comparing male losses to female losses, a binomial test revealed better male performance in block 8, $p < .05$.

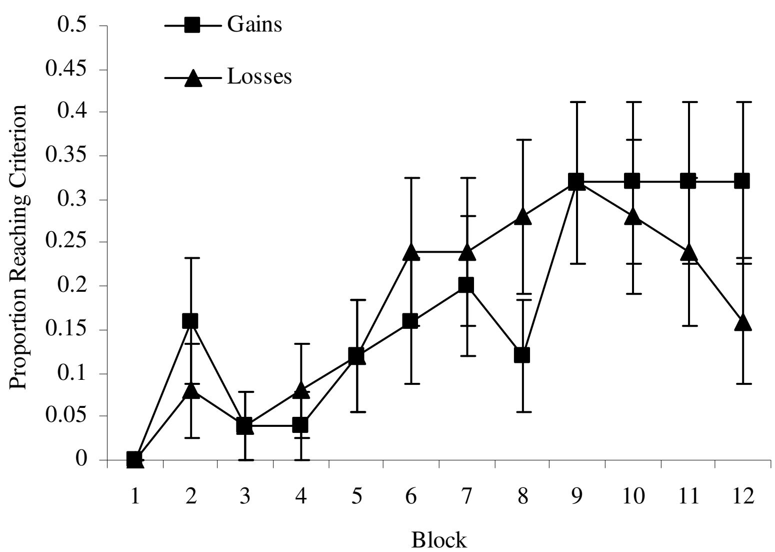


Figure 4.25. Proportion reaching criterion across blocks for men in the gains and losses tasks in Experiment 4

As shown in Figure 4.26, for women in the gains task as compared to women in the losses task, the binomial test for block 8 revealed the advantage for the gains task, $p < .05$. A binomial sign test across blocks revealed that women in the gains condition outperformed women in the losses condition in 8 of the 12 blocks. This difference was statistically reliable, $p < .05$ (3 of the other blocks are ties). Comparing female gains to male gains, binomial tests revealed better female performance in blocks 4 and 8, $p < .05$.

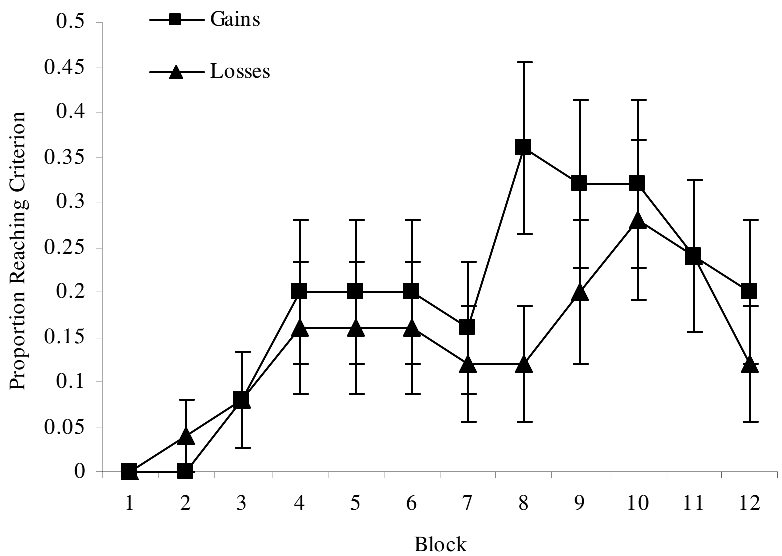


Figure 4.26. Proportion reaching criterion across blocks for women in the gains and losses tasks in Experiment 4

First block exceeding criterion. Again, any participant who failed to meet the criterion during the experiment was coded as a 13. Men in the gains condition exceeded the criterion sooner (after 8.7 blocks on average) as compared to men in the losses condition (after 9.4 blocks on average). Women in the gains condition exceeded the criterion sooner (after 8.4 blocks on average) as compared to women in the losses condition (after 8.9 blocks on average). Neither of these differences was statistically reliable.

Discussion

As for Experiment 3, I predicted that the regulatory mismatch groups (e.g., men in losses and women in gains) would outperform the match groups (e.g., men in gains and women in losses). This is the opposite pattern of data obtained in Experiment 2. Consistent with the predictions from regulatory fit theory, women in the gains task were more accurate across blocks and more reached the criterion as compared to women in the losses task. Further, as predicted, women in the gains task were more accurate than men in the gains task. However, inconsistent with predictions, men in the gains task and women in the losses task were more accurate than men in the losses task.

Model-based Analyses

As in Experiments 1 and 2, an advantage of using the information-integration classification task is that I have a set of computational models that allow me to characterize participants' responses on a block-by-block basis. I fit the series of decision-bound unidimensional and conjunctive models used in Experiments 1 and 2 and incorporated general linear contrast models and conjunctive models that used the position dimension (Ashby & Maddox, 1993; Maddox & Ashby, 1993). I fit two different general linear models. First, I fit an optimal model which assumes the participant used the optimal criterion which has an intercept of zero and a slope of 1. This model only has one free noise parameter. Second, I fit a suboptimal model which assumes suboptimal intercept and slope criteria. Therefore, this model has three free parameters: one for the intercept, one for the slope, and one noise parameter. I also fit conjunctive models that used position and length or position and orientation. Each of these models has three free parameters: one for position, one for length or orientation, and one noise parameter. As for Experiments 1 and 2, model parameters were estimated using maximum likelihood (Ashby, 1992).

Before summarizing the model results, it is important to determine that the models provided a good account of the data. The general linear model accounted for an average of 81% and 82% of the total category responses in Experiments 3 and 4, respectively.

	Experiment	
Model	**3**	**4**
General Linear Classifier	81.2	81.6
Conjunctive	72.6	72.8
Length	59.5	59.4
Orient	40.6	40.3
Position	82.0	82.1

Table 4.12. Percent of total category responses accounted for by each model type

For both experiments, the unidimensional rules on length and orientation were rarely used by participants. The unidimensional length and orientation models best fit the data <1% and <2% of the time, respectively. In contrast, the unidimensional position model best fit 17.6% and 18.6% of the time and the conjunctive models best fit 49% of the time. On average, the general linear classifier fit 27.7% and 26.1%, respectively.

To determine the success of participants using each of the model types, the best fitting model for each block for each participant was matched with block accuracy. On average, participants best fit by the general linear classifier were the most accurate for both experiments (*M* = 84% and 87%, respectively). When participants were best fit with the position model or conjunctive models, they scored an average of 82% in both experiments.

The remaining model discussions will focus on the general linear classifier fits and specific predictions of regulatory fit theory.

Experiment 3: Women primed with positive stereotype. Figure 4.27 displays the overall proportion of data sets best fit by a general linear classifier model for men and women in the gains and losses conditions. Because men in the gains condition and women in the losses condition are in a regulatory mismatch, I predicted that a larger proportion of data sets in these conditions would be best fit by a general linear classifier model. For men, this pattern did not hold; a sign test across blocks did not reveal a significant difference between the gains and losses tasks. Further, relative to men in the losses task, men in the gains task were more often fit across blocks by the position model (significant based on a sign test, $p < .05$) and by conjunctive models involving position (marginally significant based on a sign test, $p = .07$). While not predicted, relative to women in the losses task, women in the gains task were more often fit by the general linear classifier in 7 of the 12 blocks of trials, marginally significant based on a sign test, $p = .07$. This trend was supported by a significant binomial test in block 1, $p < .05$.

Analyzing the data in a manner consistent with stereotype threat, men in the gains task were more often fit by the general linear classifier than women, significant based on a sign test, $p < .05$, and women were more often fit by conjunctive models involving position, marginally significant based on a sign test, $p = .07$. Using binomial tests, a higher proportion of data from men was fit by the general linear classifier in blocks 8, 9, and 10, $p < .05$. The same pattern held in the losses task; men were more often fit by the general linear classifier, $p < .05$, and women were more often fit by conjunctive position models, $p < .05$. Binomial tests reveal more male data sets fit than female sets in blocks 1, 4, 6, 7, 9, 10, 11, $p < .05$, with a marginally significant test in block 8.

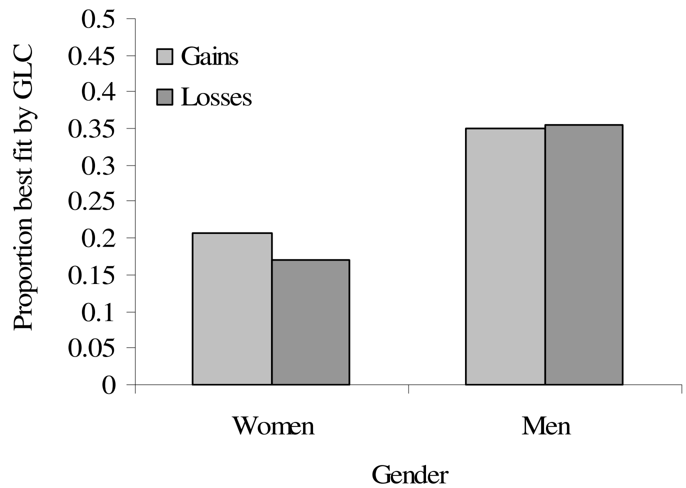


Figure 4.27. Proportion best fit by the correct general linear classifier rule for women and men in the gains and losses tasks in Experiment 3

In addition to examining proportions accounted for by different models, I considered when models best accounted for data sets. I predicted that the mismatch groups' (e.g., men with gains and women with losses) data sets should be fit by the general linear classifier earlier than the match groups. As shown in Figure 4.28, on average, men were fit earlier in the experiment than women, $F(1,76) = 5.68$, $MSE = 25.65$, $p < .05$, but there was not an interaction between gender and reward structure.

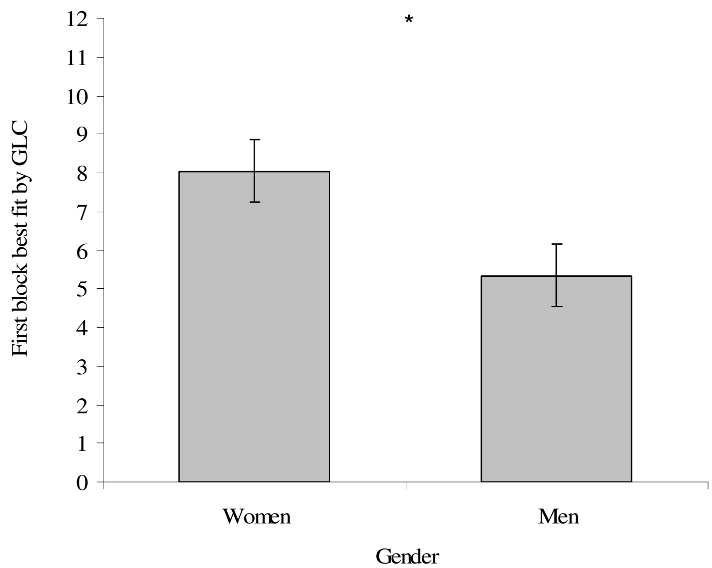


Figure 4.28. Average first block best fit by the general linear classifier model for women and men in Experiment 3

I also predicted that the match groups will exhibit behavior consistent with flexibility whereas the mismatch groups will not. I examined whether the match groups (e.g., men in losses and women in gains) switched strategies more often or tried more strategies as compared to the mismatch groups (e.g., men in gains and women in losses), and whether the mismatch groups had longer runs of using the same strategy as compared to the match groups. Using 2 Gender (Women, Men) x 2 Reward Structure (Gains, Losses) ANOVAs, there were not significant interactions of Gender and Reward Structure for the number of times a participant switched strategies, the number of strategies used by participants, or the longest run of consecutive strategy use. However, as shown in Figures 4.29 and 4.30, my data is consistent with the claims that the match groups switched strategies more often and the mismatch groups had longer runs of consecutive strategy use, respectively.

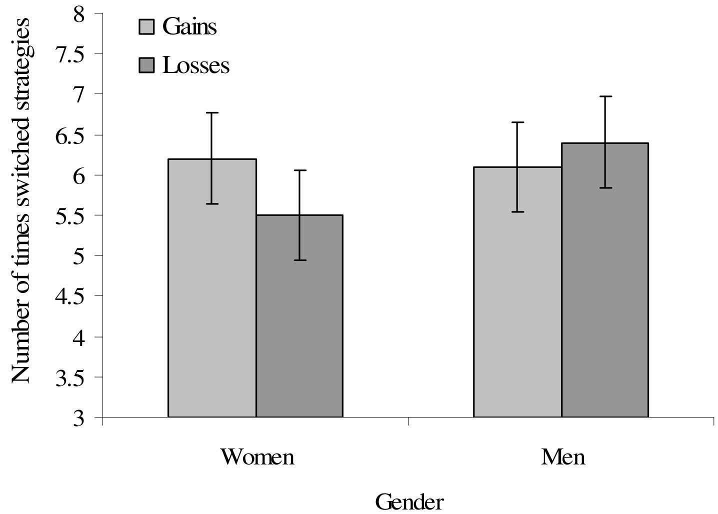


Figure 4.29. Number of times modeling suggested a strategy switch for women and men in gains and losses in Experiment 3

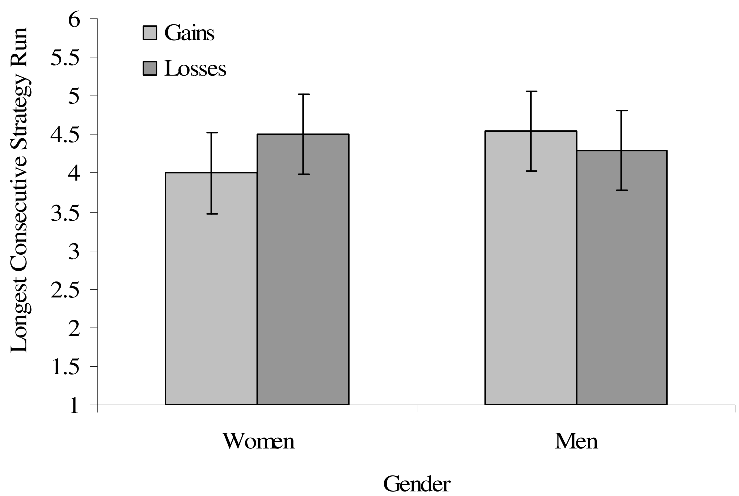


Figure 4.30. Longest run of the same strategy, measured in blocks, for women and men in gains and losses in Experiment 3

I examined the goodness of fit of the general linear model when it was selected as the best fitting model for a particular block. For each subject, I extracted the percent of the data accounted for by the model. I assume that the benefit of being in a regulatory mismatch will occur earlier in learning and predict that the mismatch participants will have more of their data fit in the first five blocks in terms of average data fits, a greater maximum value fit, and more improvement between the first and second consecutively fit blocks. If participants were not fit best by the general linear model in the first five blocks of trials, they were dropped from the analyses. Using 2 Gender (Women, Men) x 2 Reward Structure (Gains, Losses) ANOVAs, there were not significant interactions of Gender and Reward Structure or main effects for the percent of the data accounted for by the general linear model in the first five blocks or the maximum value fit in the first five blocks.

Because very few women in the losses task were best fit by the general linear model in consecutive blocks, I was only able to examine the improvement between the first and second consecutively fit blocks for the men. The data for both men and women appear in Figure 4.31. For men, participants in the gains task showed more improvement in fit than participants in the losses task, $t(11) = 2.4$, $p < .05$.

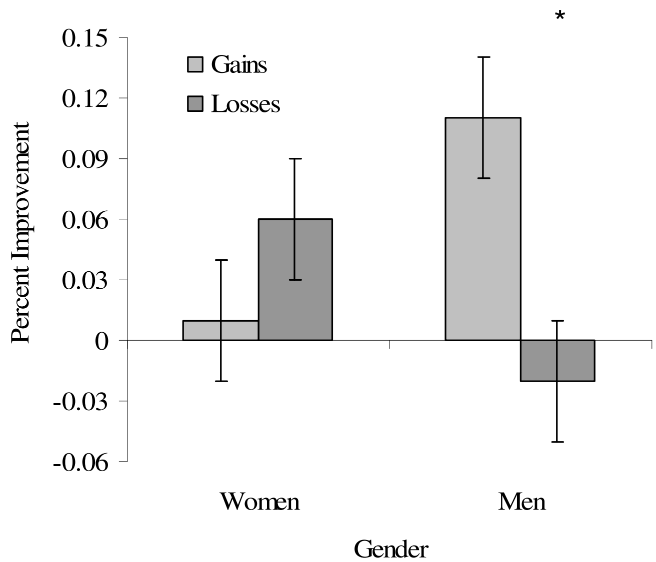


Figure 4.31. Percent improvement in general linear model fits between the first and second consecutively fit blocks for women and men in gains and losses in Experiment 3

Experiment 4: Men primed with a positive stereotype. As predicted, as shown in Figure 4.32, men in the losses task were better fit overall by the general linear model as compared to men in the gains task, binomial test $p < .05$. This overall difference is supported by significant binomial tests for blocks 1, 3, 7, and 9 demonstrating better fits for losses task participants, $p < .05$. However, a binomial sign test across blocks did not reveal a significant difference between the gains and losses tasks. For women, binomial tests demonstrate that those in the gains task were better fit by the general linear model overall as compared to those in the losses task, $p < .05$ (see Figure 4.32). This overall test is supported by binomial tests showing significant advantages for women in the gains task in blocks 3, 5, 9, 12, $p < .05$, and marginally significant differences in blocks 1, 4, and 8, $p < .08$. Further, women in the gains task were better fit in 10 of 12 blocks as compared to women in the losses task, significant based on a sign test, $p < .05$.

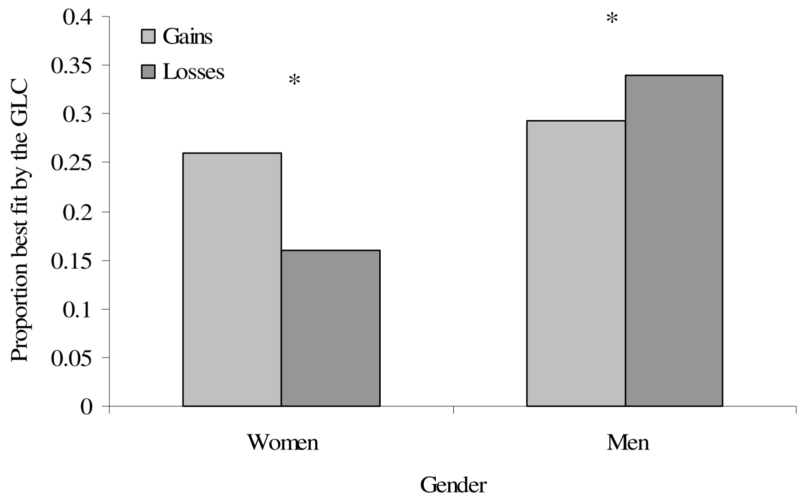


Figure 4.32. Proportion best fit by the correct general linear classifier rule for women and men in the gains and losses tasks in Experiment 4

Analyzing the data using binomial tests, in a manner consistent with prior stereotype threat work, men in gains were better fit by the general linear model in block 3, marginally significant $p < .08$, while women in gains were better fit in block 7, marginally significant $p < .08$. Further, women in the gains task were more often fit across blocks by conjunctive models using position as compared to men in the gains task, significant based on a sign test, $p < .05$. Lastly, men in the losses task were more often fit across blocks by the general linear model as compared to women in the losses, significant based on a sign test, $p < .05$. This difference is supported by significant binomial tests where men were better fit than women in blocks 1, 2, 3, 5, 6, 7, $p < .05$, and marginally significant tests in blocks 4, 8, and 12, $p < .08$. In contrast, as compared to men, women in the losses task were more often fit across blocks with conjunctive models that included position, significant based on a sign test, $p < .05$, and were better fit in 10 of the 12 blocks. Similarly, a marginally significant sign test, $p < .08$, reveals that women were better fit in 7 of the 12 blocks (with 4 ties) by the position model as compared to men.

As for Experiment 3, I analyzed how quickly participants were best fit by the general linear classifier model using a 2 Gender (Women, Men) x 2 Reward Structure (Gains, Losses) ANOVA. As shown in Figure 4.33, men were fit earlier than women, $F(1,99) = 3.79$, $MSE = 3.79$, $p < .05$, but this main effect was qualified by a marginally significant interaction between gender and reward structure, $F(1,99) = 3.05$, $MSE = 3.79$, $p = .08$. Women in the gains task were fit earlier ($M = 7.04$) than women in the losses task ($M = 9.04$), $t(48) = 1.38$, $p < .05$.

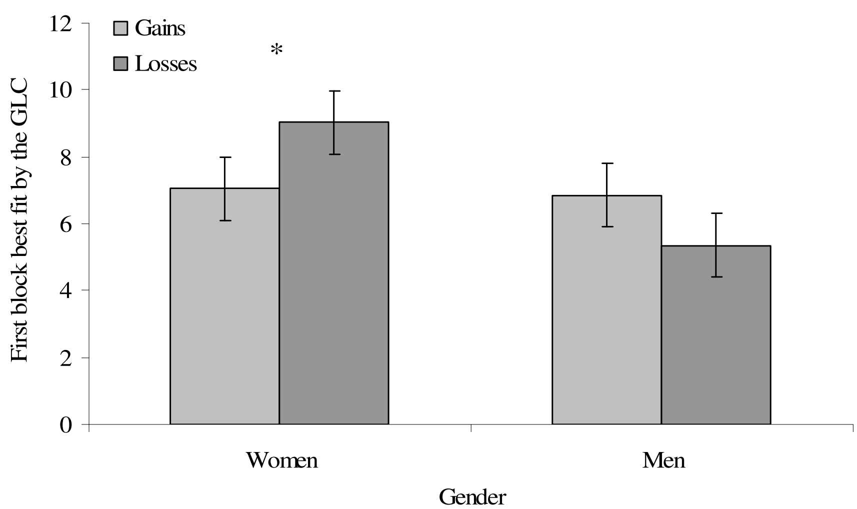


Figure 4.33. Average first block best fit by the general linear classifier model for women and men in the gains and losses tasks in Experiment 4

Similarly, I analyzed the number of strategies used, the number of times the strategy changed from block to block, and the longest run of the same strategy using 2 Gender (Male, Female) x 2 Reward (Gains, Losses) ANOVAs. There were no significant interactions between Gender and Reward Structure for any of these dependent measures. However, as shown in Figure 4.34, there is a pattern consistent with the mismatch groups (e.g., women in gains and men in losses) having longer runs of the same consecutive strategy. There was also a marginally significant main effect of Reward Structure on the number of different strategies used, $F(1, 96) = 3.49$, $MSE = .73$, $p = .06$. Participants in the losses task used more strategies ($M = 3.2$) on average than participants in the gains task ($M = 2.9$).

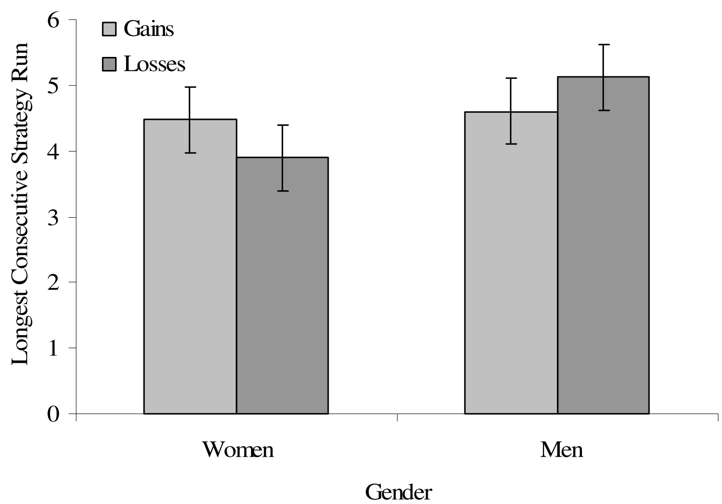


Figure 4.34. Longest run of the same strategy, measure in blocks, for women and men in the gains and losses tasks in Experiment 4

Again, as for Experiment 3, I examined the goodness of fit when the general linear contrast model was selected as the best fitting model for a given block. Using 2 Gender (Women, Men) x 2 Reward Structure (Gains, Losses) ANOVAs, there were not significant interactions for the average fit in the first five blocks of trials or for the maximum best fit in the first five blocks. However, there was a significant main effect of Gender for the average fit, $F(1,43) = 9.38$, $MSE = .005$, $p < .05$; women had a higher percentage of their data fit ($M = 85\%$) than men ($M = 79\%$). There was also a marginally significant main effect of Gender for maximum fit in the first five blocks $F(1, 43) = 3.45$, $MSE = .007$, $p = .07$. Again, women had a higher percentage of their data fit (M = 87%) relative to men (M = 82%). Lastly, I analyzed the fit improvement between the first and second fit block in the first 5 blocks. As for Experiment 3, too few women had two blocks fit and therefore as a group their data cannot be analyzed. There was no difference in the fit improvement for men in the gains and losses tasks.

General Discussion

In Experiments 3 and 4, I found some data consistent with regulatory fit predictions and our interpretation of the stereotype threat literature. Consistent with regulatory fit predictions, in Experiment 3, men in the gains task reached the criterion sooner than men in the losses task, and, in Experiment 4, women in the gains task performed better than women in the losses task. Likewise, consistent with my interpretation of the stereotype threat literature, in the gains task, more men reached the criterion than women in Experiment 3, and women performed better than men in Experiment 4.

Participants learned to classify lines into two categories. These lines varied in length, orientation, and position on the screen. Participants could perfectly classify the lines if they used an information-integration strategy and placed lines that were longer than they were steep into one category and all of the other lines into another category. I predicted that this information-integration category structure would be better learned by individuals in a regulatory mismatch because those in a regulatory match would persist in explicit hypothesis testing. Explicit hypothesis testing is not useful in learning information-integration category structures because the rule that separates the categories is not easily verbalizable.

In Experiment 3, I primed women with a positive stereotype and men with a negative stereotype. I reversed the stereotype assignment in Experiment 4. In both Experiments, half of the participants completed a gains version of the task where they gained points for correct responses and half completed a losses version of the task where they lost fewer points for correct responses than incorrect responses. I predicted that a regulatory mismatch would occur when those primed with a negative stereotype completed the gains task and those primed with a positive stereotype completed the losses task. Further, I predicted that participants in a regulatory mismatch would perform better in the classification task relative to participants in a regulatory match.

In Experiment 3, I found limited support for my predictions. As predicted, in the gains task, more men reached the criterion than women in 10 of the 12 blocks of trials, and exceeded the criterion sooner as compared to men in the losses task. However, contrary to my expectations, more men in the losses task also reached the criterion than women in the losses task in 7 of the 12 blocks. And, women in the gains task performed better than women in the losses task in 10 of the 12 blocks.

Examining the questionnaire data, participants were well calibrated to their level of performance. Women in the gains task believed they performed better, relative to other women in the experiment, than women in the losses task did. Further, women rated their performance lower than men. Interestingly, at the start of the experiment, women in the gains task expected to perform better than women in the losses task did and men in the losses task expected to perform better than men in the gains task did, although this comparison was not statistically reliable. This first reaction to the task is what would have been predicted from regulatory fit theory. Individuals experiencing a fit benefited from the presence of a match between their situationally-primed focus and the task context.

In Experiment 4, I find some support for my predictions in the accuracy data and questionnaire data. Women in the gains task performed better than women in the losses task in all 12 blocks of trials. Further, in the gains task, women outperformed men in 11 of the 12 blocks. However, men in the gains task performed better than men in the losses task in 10 of the 12 experimental blocks. In the losses task, women performed better than men in 10 of the 12 blocks.

Examining the questionnaire data, again I find some support for the benefits of experiencing fit. Women in the losses task rated their mood as higher at the beginning of the task relative to women in the gains task. At the end of the task, relative to women in the experiment, women in the gains task rated their performance as higher than women in the losses task and men in the losses task rated their performance as higher than men in the gains task. Interestingly, this performance rating corresponds to the presence of a regulatory match or mismatch and does not reflect actual task performance relative to the other groups.

The modeling results are generally consistent with my predictions, particularly in Experiment 4. In Experiment 3, women were primed with a positive stereotype and men were primed with a negative stereotype. I predicted that women in the gains task and men in the losses task would exhibit more flexible performance as compared to the mismatch groups and be less likely fit by the optimal general linear classifier model. Consistent with my predictions, men in the gains task and women in the losses task were more often fit by models that involve position as compared to men in the losses task suggesting limited rule testing. These groups continued to use the likely default strategy and closely related strategies. In addition, men in the gains task were more often fit by the general linear model as compared to women in the gains task. I also found data patterns consistent with the match groups switching strategies more often and the mismatch groups using the same strategy for more consecutive blocks, but these differences were not statistically reliable.

Contrary to my predictions, women in the losses task were less often fit by the general linear classifier as compared to women in the gains task and men in the losses task. I predicted that women in the losses task were experiencing a regulatory mismatch and therefore were predicted to use a strategy consistent with the general linear classifier model. Further, in the gains task, women were more often fit by conjunctive models involving position as compared to men. Overall, men were best fit by the general linear classifier in earlier blocks relative to women.

In Experiment 4, men were primed with a positive stereotype and women with a negative stereotype. As predicted, men in the losses task and women in the gains task were more often fit by the general linear classifier model as compared to men in the gains task and women in the losses task, respectively. These groups were also fit earlier in the experiment by the general linear model. In the losses task, men were also more often fit than women. Further, in the gains task, women were more often fit by conjunctive position models relative to men. Contrary to my predictions, in the losses task women were more often fit by position models than men. Overall, women had a greater percentage of their data fit by the general linear model when it was selected as the best fitting model than men.

Across these two experiments, the accuracy and modeling data is not completely consistent with my predictions from regulatory fit and stereotype threat. I believe in part this inconsistency stems from the chosen stimulus space. The stimulus structure may have allowed participants to get decent performance simply by using strategies consistent with conjunctive rules that use position instead of relying on the more optimal information-integration strategy. Future research will use a stimulus structure that allows for more differentiation between the optimal strategy and strategies that are more easily verbalizable. My results and recent results by other researchers have prompted a revision of how information-integration category structures are constructed.

In addition, comparing across the two experiments, there is a Gender x Experiment interaction. Men performed better than women in Experiment 3 but women performed better than men in Experiment 4. In Experiment 3, this main effect was coupled with earlier general linear classifier fits for the men. This generally supports a claim that negative stereotypes do produce better performance in information-integration learning.

Chapter 5: Conclusions

This dissertation examines the interaction between motivational factors that drive behavior, the stereotypes that induce them, the reward structure of tasks, and the type of task performed. I argue that regulatory focus states, motivational variables, are induced when positive and negative stereotypes are primed. Specifically, priming a negative stereotype induces a prevention focus and priming a positive stereotype induces a promotion focus. Prior work on regulatory focus demonstrates that a prevention focus leads to sensitivity to losses and non-losses while a promotion focus leads to sensitivity to gains and non-gains (Higgins, 1987, 1997a). Further, I demonstrate that task performance depends on the induced regulatory focus and the reward structure of the task environment. When the focus matches the reward structure, participants are more cognitively flexible and engage in more explicit strategy testing than when participants are in a mismatch. Lastly, I show that cognitive flexibility is advantageous in some tasks but not in others.

In four experiments, I manipulated the primed stereotype assigned to men and women, the type of task performed, and the reward structure of the task. I predicted a four-way interaction between Stereotype, Gender, Reward structure, and Task. Participants learned to classify lines that varied in length, orientation, and position on the screen into two categories. The rule that could yield perfect classification varied across experiments. In Experiments 1 and 2, the two categories to be learned could be perfectly distinguished using a verbalizable conjunctive rule. In Experiments 3 and 4, the rule that distinguished the categories was an information-integration rule which is not likely to be generated by testing easily verbalizable rules. As such, individuals who flexibly tested rules were predicted to perform better in Experiments 1 and 2 but worse in Experiments 3 and 4. In the first two Experiments, testing lots of rules should provide good coverage of the relevant rule space allowing an individual to find the correct rule. In contrast, the correct rule in Experiments 3 and 4 lies outside of the rule space likely searched by participants. Therefore, persistence in searching the rule space will hinder performance because the correct rule will not be found and used.

In all Experiments, participants were assigned to a gains reward structure where they gained more points for correct responses than for incorrect ones or were assigned to a losses reward structure in which they lost fewer points for correct responses relative to incorrect ones. Some participants were primed with a positive stereotype and some with a negative stereotype. Using regulatory fit theory, I predicted priming a positive stereotype induces a promotion focus leading to sensitivity to gains. Therefore, individuals primed with a positive stereotype would experience a regulatory match in the gains version of the task and a mismatch in the losses version of the task. In contrast, priming a negative stereotype induces a prevention focus leading to sensitivity to losses. Negative-stereotype priming places individuals in a regulatory match in the losses environment and a regulatory mismatch in the gains environment.

Further, prior research on regulatory fit demonstrates that regulatory matches promote more flexible rule testing than regulatory mismatches (Maddox et al., 2006; Maddox et al., 2007; Markman, Baldwin et al., 2005; Markman, Maddox et al., 2005; Markman et al., 2006). Therefore, I predicted individuals in a regulatory match should perform better in Experiments 1 and 2, where flexible rule testing is advantageous, but worse in Experiments 3 and 4, where flexible testing is disadvantageous. The opposite pattern was predicted for individuals in a regulatory mismatch.

In Experiments 1 and 2, men and women performed a classification task for which flexible rule testing is advantageous. In Experiment 1, participants were told women tended to perform better than men while in Experiment 2, participants were told men tended to perform better than women. Behavioral accuracy data and model-based analyses supported our predictions. Individuals experiencing a regulatory match performed better on the task and their data was more consistent with an application of the correct classification rule relative to individuals in a regulatory mismatch.

In Experiments 3 and 4, participants completed a classification task for which flexible rule testing is disadvantageous. Participants were told women tended to perform better in Experiment 3 and men tended to perform better in Experiment 4. In both Experiments, the accuracy data was not entirely consistent with my predictions. Overall, individuals in the gains task performed better than those in the losses task. However, in the gains task, the individuals primed with a negative stereotype performed better. That is, the mismatch groups performed better; men performed better in Experiment 3 and women performed better in Experiment 4. The model-based analyses revealed a pattern closer to our predictions. In Experiment 3, male data was more often fit by the correct rule than female data. In Experiment 4, the mismatch groups' data was more consistent with using the correct rule than the match groups' data.

In this chapter, I will revisit stereotype threat effects in the literature. I will discuss how my theoretical approach can illuminate prior results and what factors need to be included to explain stereotype threat. Second, I will briefly review neural mechanisms that may underlie flexible rule testing behaviors given a regulatory match. Third, I will highlight a set of related individual differences that should be examined in future work, and lastly discuss some practical implications of my approach and present data from an experiment that applies our theoretical approach to math performance.

STEREOTYPE THREAT IMPLICATIONS

This dissertation claims to be able to account for stereotype threat effects using regulatory fit. As demonstrated in rule-based classification, negative stereotype priming leads to better performance in losses environments than gains and positive stereotype priming lead to better performance in gains than losses. However, it is still an open question as to whether this effect is merely another way to demonstrate and ameliorate stereotype threat or if this perspective can account for a range of findings in the stereotype threat literature.

To address this issue, I revisit some of the most influential and most often cited studies discussed in Chapter 2. I consider specific task contexts to classify prior work as instantiations of gains environments and describe possible ways to create losses environments. My goals are to postdict study results and predict results given losses versions of the tasks. Further, I will discuss what factors are necessary to include in explaining stereotype threat.

Much of the work on stereotype threat has been completed using verbal and math tests. Unintentionally creating a gains context, Steele and Aronson (1995) told subjects that they should not expect to get many questions correct in all experimental conditions.

Merely mentioning correct responding may be enough to frame a test as a gains environment. Therefore, Steele and Aronson created a regulatory mismatch when Black participants were told the test was diagnostic of their ability or had their race highlighted. These Black participants were prevention-focused in a gains environment. To reverse this effect, the test could be described as one where subjects should expect to get many questions incorrect. This creates a losses context and should improve performance of Black participants in the diagnostic and race prime conditions. Likewise, Keller and Dauenheimer (2003) created a gains environment by emphasizing to students that they needed to solve as many problems as possible and demonstrated the classic stereotype threat effect with women and math. A simple change of instructions telling students to avoid getting problems incorrect would reverse their effects.

Similarly, Spencer et al. (1999) asked participants to take the GRE (see also Spencer & Quinn, 2001). As part of the test instructions, participants read the standard GRE scoring from 1999: correct items get 1 point, blank items get no deductions, and incorrect items get a deduction to correct for guessing. Technically-speaking, this point structure is a mixed structure composed of both gains and losses. However, the correct and blank items' scoring matches the gains environment used in our studies and the incorrect scoring is a small loss that may not be well understood by participants. As such, it appears this test context is more of a gains environment than a losses environment. Therefore, their female participants underperformed relative to men on the difficult math test because they were experiencing a regulatory mismatch. To reverse this effect, correct items would have fewer points lost than blank items or guesses.

Using a different domain, Stone et al. (1999) had participants complete a golf course and measured performance using the number of strokes to complete the course. They found that groups with negative stereotypes performed worse than those without. There was an elaborate cover story given to participants. Each participant completed 10 rounds. In each round they were told to try to get the ball in a hole in the fewest strokes possible. There were multiple possible holes: a small hole earning 1 point, a medium hole earning 2 points, and a large hole earning 3 points. Subjects were told to try to maximize their points. This is clearly a gains environment. It is unclear why this point system was used and the researchers do not report actual points earned, which is very problematic because subjects were trying to maximize their earned points. One could imagine that subjects passed by closer lower-point holes and took more shots to get more points. At any rate, to reverse their effects, one would simply need to frame the test in terms of lost points and tell subjects to minimize points lost.

There are also a series of studies that fail to describe procedures well enough to determine if an explicit gains environment was created (Aronson et al., 1999; Shih et al., 1999). For example, Aronson et al. (1999) does not describe the instructions students read that relate to the math test. They only describe the content of the stereotype threat manipulations. However, like all laboratory studies, students are completing experiments to gain course credit or pay. Without disconfirming evidence, I believe that these studies could be consistent with my perspective.

As reviewed in Chapter 2, there are many highlighted mechanisms for stereotype threat some of which appear to match what would be expected if negative stereotypes induce a prevention focus. For example, stereotype avoidance (Steele & Aronson, 1995), self-doubt activation (Steele & Aronson, 1995), anxiety (Spencer et al., 1999), and negative thinking (Cadinu et al., 2006) are all factors that seem consistent with a prevention focus. The literature on regulatory focus suggests that a prevention focus is linked to anxiety and avoidance.

Further, there are other factors in the literature which may reduce the likelihood that a prevention focus is induced and therefore influence performance. For example, Spencer et al. (1999) frame a test as not showing gender differences in the past and demonstrate that stereotype threat effects went away (see also Keller & Dauenheimer, 2003). Similarly, providing an alternative explanation for performance or a self-handicap (Stone et al., 1999; Ben-Zeev et al., 2003) or giving participants a working memory dual-task (Beilock et al., 2006) reduces stereotype threat effects. I would argue these manipulations worked because a prevention focus was not induced or was masked thereby not interacting negatively with the gains task structure.

However, it may not be so simple. Brown and Josephs (1999) manipulated test framing. Some participants were told the test was diagnostic of weak ability while others were told the test was diagnostic of exceptional ability. Let us assume, as did Keller and Bless (2006), that the weak ability condition primes prevention and the exceptional ability condition primes promotion. Further, let us assume that the environment was an implicit gains environment; students were trying to get math questions correct and were earning credit for a course research requirement. If these assumptions are correct, regulatory fit cannot account for the pattern of data. For men, a "prevention" (mismatch) state produces better performance than a "promotion" (match) state while a "prevention" state produces worse performance for women than a "promotion" state.

If, however, we attend to Brown and Josephs' argument that the weak ability condition corresponds to the performance concerns of women and the exceptional ability condition corresponds to the performance concerns of men, our predictions change. Given that men want to avoid not appearing exceptional and women want to avoid appearing weak, these conditions may be inducing a prevention focus, causing a regulatory mismatch, which reduces performance levels. Interpreting the test framing in this way allows the data to be interpreted as consistent with regulatory fit predictions. This study highlights the importance of determining what conditions induce a prevention focus.

Lastly, there are other factors that probably are working in concert with regulatory fit to create experimental effects, such as domain identification (Aronson et al., 1999; Stone et al., 1999) with testosterone as a moderator (Josephs et al., 2003), task difficulty (Spencer et al., 1999), and working memory capacity (Schmader & Johns, 2003; Beilock & Carr, 2005; Beilock et al., 2006; Beilock et al., 2004). For example, Aronson et al. (1999, Experiment 2) demonstrate that high math identifiers perform worse in a stereotype threat condition but better in a control condition relative to moderate math identifiers. Using prior work, the stereotype threat condition is a regulatory mismatch and the control is a regulatory match. If we assume that domain identification magnifies the effect of the induced focus state, high identifiers should do worse in a mismatch than low identifiers but better in a match. This is the pattern of data found by Aronson et al. Further, as demonstrated by Josephs and colleagues, testosterone may moderating influence domain investment/identification on stereotype threat (Josephs, Newman, Brown, & Beer, 2003), potentially magnifying further the effect of the induced state. Future research should investigate this possibility.

This interpretation is supported by a meta-analytic review of stereotype lift (Walton & Cohen, 2003). Walton and Cohen found that individuals with positive stereotypes (e.g., men and Whites) perform better when tests are framed a diagnostic of ability. This effect is completely consistent with the claim in this dissertation assuming most tasks are gains environments. Further, Walton and Cohen found that individuals from these groups performed better when they were high identifiers relative to low identifiers. I would argue these groups are in a strong regulatory match.

Likewise, task difficulty surely interacts with regulatory fit effects. In our prior work we demonstrate the interaction of fit and task difficulty. In Grimm et al. (in press), the effect of regulatory fit on a rule-based task was smaller than the effect of regulatory fit on a rule-based task by Maddox et al. (2006). In that study, however, the rule to be learned was more complex than the simple unidimensional rules in Grimm et al. We expect larger motivational effects when tasks are difficult than when they are easy.

Beilock's research demonstrates the role of working memory in producing stereotype threat effects for proceduralized tasks (Beilock & Carr, 2005; Beilock et al., 2006; Beilock et al., 2004). In tasks that are well-learned, Beilock et al. (2004) argue that stereotype threat and situational pressure situations use similar mechanisms to produce performance decrements. Like "choking under pressure," stereotype threat induces explicit monitoring of performance which hurts tasks that are automatized. In contrast, Schmader and Johns (2003) demonstrates that stereotype threat reduces working memory capacity and therefore hurts performance. As Beilock et al. (2006) note, the exact impact of stereotype threat on working memory appears to greatly depend on the type of task performed.

The Schmader and Johns (2003) task appears to be an explicit more rule-based task and the Beilock et al. (2006) task to be an implicit task. Based on the results from this dissertation, the negatively stereotyped group (given an implicit gains environment) should have done worse in Schmader and Johns because of less flexible processing, possibly the result of a lower working memory capacity. For a more procedural task, I predicted better performance for this group in the information-integration task and got mixed results, which may be consistent with in Beilock et al. Subjects did persist in using rules on the position dimension. This could suggest explicit inflexible monitoring of performance, instead of a reliance on the procedural system. Future research should continue to investigate how working memory relates to stereotype threat effects.

POSSIBLE NEURAL MECHANISMS

The long-term success of this research program does rely on creating a neurally plausible model to account for differences in regulatory fit. In this section, I speculate on regions and systems that might be involved. Most related to the work in this dissertation, Seger and colleagues study the role of the caudate in classification learning (Cincotta & Seger, 2007; Seger & Cincotta, 2005, 2006). The caudate is part of the basal ganglia. Within this structure, the caudate and the putamen make up a substructure, the striatum. Seger (2006) argues that the striatum is connected to the cortex in 4 different pathways that have different functions. The executive pathway connects the head of the caudate with the dorsolateral prefrontal and posterior parietal regions. The visual pathway connects the body and tail of the caudate with the temporal cortex and ventrolateral prefrontal cortex. The motor pathway connects the putamen to the premotor and somato-sensory areas, and the motivational pathway connects the ventral striatum to the orbitofrontal cortex and the anterior cingulate.

We are most interested in regions that process feedback because the task reward environment is a critical component of regulatory fit. Prior work by Seger and Cincotta suggests that our classification task relies on the head of the caudate which processes feedback (Cincotta & Seger, 2007). In a rule-based task, this region was more active at the start of learning, while areas in the cortex were more active later (Pasupathy & Miller, 2005; Seger & Cincotta, 2005), which may suggest a processing of the reward context. Furthermore, the head of the caudate shows stronger patterns of activation given positive feedback relative to negative feedback (Filoteo et al., 2005; Seger & Cincotta, 2005). Seymour, Dew, Dayan, Singer, and Dolan (2007) argue that more anterior regions of the striatum show relative selectivity for gains while more posterior regions of the striatum show relative selectivity for losses.

While there are probably many regions involved when subjects complete our classification tasks, some researchers have argued that different regions are recruited for the rule-based and information-integration versions of our classification tasks. The COVIS model of multiple memory systems (Ashby et al., 1998; Ashby & Waldron, 1999), suggests that the rule-based task mostly takes place in frontal brain regions which are used for flexible processing while the information-integration task is likely mostly learned by a procedural system instantiated in sub-cortical areas, such as the basal ganglia

(Maddox & Ashby, 2004). Further, information-integration learning is supported by dopamine release in these areas (Ashby et al., 1998).

A growing body of research suggests that the basal ganglia and the cortex interact during the course of learning. As noted above, the basal ganglia is activated earlier in learning relative to the cortex. Pasupathy and Miller (2005) trained monkeys to respond to two visual cues with distinct saccadic eye movements. They argue that the time-course in monkeys demonstrates sub-cortical areas are recognizing the reward context and then 'train' the prefrontal cortex. Interestingly, the behavioral data corresponds better to cortical activation patterns than to caudate activation patterns. So, while the caudate may assess the context more quickly, this system may be reliably slower to learn. As a caveat, work by Poldrack and colleagues (Poldrack & Rodriguez, 2004) demonstrates that the medial temporal lobe is active before the caudate but quickly deactivates. They suggest that the medial temporal lobe and the caudate may be learning independently and that the prefrontal cortex selects which region governs responding. The medial temporal lobe (e.g., the hippocampus) guides earlier responding but after extended training the caudate directs behavior (Poldrack & Packard, 2003).

It may be that regulatory fit determines whether cortical or sub-cortical areas govern responses or the speed at which a transition occurs from cortical to sub-cortical systems. In a regulatory match, behavioral and model-based analyses demonstrate participants are engaged in more explicit rule-based testing. This explicit testing seems consistent with use of cortical areas. In contrast, in a regulatory mismatch, less rule-based strategy testing is employed, which suggests that sub-cortical systems are guiding responses.

While all of this is speculative, these different strategies seem to correspond to the behavior of the locus-coeruleus. The locus-coeruleus is a group of neurons located in the pons, part of the brainstem. Aston-Jones and Cohen (2005) argue that the locus-coeruleus tracks task performance. There are two main modes of neural firing: tonic and phasic. The tonic mode is associated with disengagement from the task and a search for other options, while the phasic mode focuses processing on the current task. Aston-Jones and Cohen name these two modes exploration and exploitation, respectively. Intuitively, exploration and exploitation strategies map onto previously documented regulatory match and mismatch behavior. Those in a regulatory match seem to explore the rule space on tasks while those in a regulatory mismatch do not.

While these parallels are intriguing, it is unclear whether the firing modes in the locus-coeruleus really could scale-up to create strategy differences seen in our regulatory match and mismatch participants. The locus-coeruleus projects to cortical and sub-cortical areas and is projected to from the orbitofrontal cortex and the anterior-cingulate. Interestingly, it does not project to the caudate. Based on these connections, one plausible model of fit effects starts with a match or a mismatch state being recognized in the anterior-cingulate or orbitofrontal cortex, which tracks reward. These areas then project to the locus-coeruleus initiating a phasic or tonic mode of firing. The projections from the locus-coeruleus may then promote exploration in cortical areas or cause exploitation by inhibiting responding, which would allow the sub-cortical structures to take over and guide behavior.

THEORETICAL AND METHODOLOGICAL IMPLICATIONS

My results have theoretical and methodological implications for the study of stereotype threat and cognition more generally. Theoretically, I hope cross-disciplinary work like this will lead to further research on the influence of motivation on cognitive processing. There are a host of related individual difference variables, such as self-construal (Lee, Aaker, & Gardner, 2000; Marx, Stapel, & Muller, 2005), defensive pessimism (Elliot & Church, 2003; Yamawaki, Tschanz, & Feick, 2004), sensitivity to reward and punishment (Avila & Parcet, 2002; Poy, Eixarch, & Avila, 2004; Torrubia, Avila, Molto, & Caseras, 2001), fear of isolation (Kim & Markman, 2006; see Walton & Cohen, 2007 for a FOI type manipulation with racial stereotypes), mortality salience (Arndt, Lieberman, Cook, & Solomon, 2005; Greenberg, Pyszczynski, Solomon, Simon, & Breus, 1994; Landau et al., 2006; Wisman & Goldenberg, 2005), and achievement motivation (Elliot, 1999; Hyde & Kling, 2001).

For example, one such variable, self-construal, has been linked to regulatory focus (Lee et al., 2000) and to stereotype threat (Marx et al., 2005). Self-construal is an individual's view of self. The self is thought to be a collection of attributes, such as brown hair, and roles the self plays in relation to others, such as parent (Markus & Kitayama, 1991). Individuals with relatively more attributes are *independent* and those with relatively more roles are *interdependent*. There is evidence that self-construals vary across cultures (Markus & Kitayama, 1991) and within cultures (Cross & Madson, 1997). Furthermore, priming studies that manipulate self-construal demonstrate effects of self-construal on cognitive processing (Geodert, Grimm, Markman, & Spellman, 2007; Kim, Grimm, & Markman, 2007). Lee et al. (2000) provide evidence that an independent self-construal leads to greater preference for items framed in terms of promotion and an interdependent self-construal leads to greater preference for items framed in terms of prevention. Further, Marx et al. (2005) demonstrate that a group with a negative stereotype completed a pronoun task with more interdependent pronouns (e.g., we, our) than independent pronouns (e.g., me, mine).

My approach suggests that self-construal will also interact with the reward structure of the environment. Independent individuals may prefer promotion-framed items because that primed self-construal induces a promotion focus while an interdependent-primed self-construal induces a prevention focus. Likewise, Marx et al. find self-construal priming because of the induced regulatory focus. Future studies will investigate this possibility.

A related individual difference variable is fear of isolation (Kim & Markman, 2006). Fear of isolation (FOI) can occur in situations in which one feels lonely or anxious by virtue of being isolated from others. Like self-construal, FOI varies cross-culturally. Kim and Markman (2006) verified that East Asians have a greater FOI than American college students or European-born students. Furthermore, their experimental studies demonstrate that priming FOI leads to preference for dialectical proverbs and a greater sensitivity to contextual elements. Future work by Art Markman, Kyungil Kim, and I will consider the link between self-construal and fear of isolation. Art Markman and I already have some unpublished data suggesting these variables behave similarly.

A high fear of isolation seems to be related to a prevention focus and to stereotype threat. A high FOI is associated with anxiety, an emotion that is a hallmark of a prevention focus. In addition, Walton and Cohen (2007) manipulated social belonging by telling some students to list 8 friends who would fit in well in their department while other students listed 2 friends. Minority students who listed 8 friends reported a lower sense of fit in the major than those who listed 2 friends. Dissertation data collected by Kim (2005), under the supervision of Markman, suggests that asking students to list an unrealistically large number of friends induces a high FOI. Current work by Jason Lee, Art Markman, and I investigates whether the interaction between FOI (high, low) and the task reward structure (gains, losses) will mimic regulatory fit effects. We predict that priming a high FOI induces a prevention focus and priming a low FOI induces a promotion focus. This would suggest that Walton and Cohen's stereotype threat effect resulted from regulatory fit-related priming.

Another closely related variable is mortality salience. Mortality salience is derived from Terror Management Theory (Arndt et al., 2005; Greenberg et al., 1994; Landau et al., 2006; Rosenblatt, Greenberg, Solomon, Pyszczynski, & Lyon, 1989; Wisman & Goldenberg, 2005) , which maintains that individuals are afraid of death and want to be immortal. To support this claim, terror management researchers prime mortality salience by asking participants to write about what it will feel like to die and to discuss how thoughts of their death make them feel. Then, researchers measure expressions of a desire to be immortal. For example, Rosenblatt et al. (1989) examined bail amounts handed down by municipal court judges. Some judges were asked to write about their death (e.g., high mortality salience) while others were not (e.g., low mortality salience). Judges then read a case study about an arrested prostitute and recommended a bail amount. Judges in the high mortality salience condition generated higher bail amounts than judges in the low mortality salience condition. Rosenblatt et al. argue that higher bail amounts demonstrate a desire to uphold social norms. Upholding social norms is one means to attaining immortality. That is, people are more likely to remember you when you are gone.

Much like fear of isolation, high mortality salience seems like a prevention focus. Individuals are primed to be vigilant and anxious. Prior work on regulatory focus has argued that a prevention focus leads to more vigilant processing where individuals worry about misses while a promotion focus leads to more elaborative processing where individuals focus on hits in a recognition memory task (Friedman & Forster, 2001). Art Markman and I have some unpublished data linking mortality salience with fear of isolation and self-construal. Future work will investigate whether mortality salience primes induce a prevention focus and can produce regulatory fit effects on cognitive tasks. By studying the relationships between these and other individual differences, researchers may be able to posit similar mechanisms.

Furthermore, methodologically, there are a host of individual difference variables that could benefit from using well-understood tasks from cognitive psychology (Narvaez & Markman, 2006). For example, self-construal has been studied in the domain of casual induction (Kim et al., 2007). This is a domain with a rigorous mathematical definition for what it might mean to be sensitive to the presence or absence of causes. In our tasks, we used mathematical models to characterize performance. These models allow us to be more confident about what participants are actually doing in the service of completing our task.

PRACTICAL IMPLICATIONS

Practically, our work suggests a way around stereotype threat. By elaborating on the observation by Seibt and Förster (2004), we demonstrate that it is possible to change relatively minor aspects of the task environment to get large differences in performance. This suggests that it is possible to reverse the negative effects of negative stereotypes by changing small task characteristics. Performing well in a domain typically associated with a negative stereotype may be an excellent first step in curbing performance decrements caused by negative stereotype encounters.

To investigate this possibility, we extended my dissertation work to consider regulatory fit effects in mathematics. There have been numerous studies in stereotype threat comparing the performance of women and men on math tests. It is assumed that individuals have the stereotype that women are bad at math. Using our theoretical perspective, women who take a math test would have a prevention focus while men would have a promotion focus. Therefore, men should do well in a gains test environment relative to a losses environment while the opposite is predicted for women.

Without manipulating task-relevant stereotypes, we told students they were going to take a math test that was diagnostic of their ability in math. Students then completed 20 questions from the GRE. Students in the gains environment earned points for correct responses and students in the losses environment lost fewer points for correct responses. There were approximately 20 women and 20 men in each reward structure. We analyzed our data using a 2 Gender (Female, Male) x 2 Reward structure (Gains, Losses) ANCOVA. We took math importance into account as a covariate. As shown in Figure

5.1, there is an interaction of Gender and Test Type, $F(1,74) = 5.12$, $MSE = 216.53$, $p < .05$.

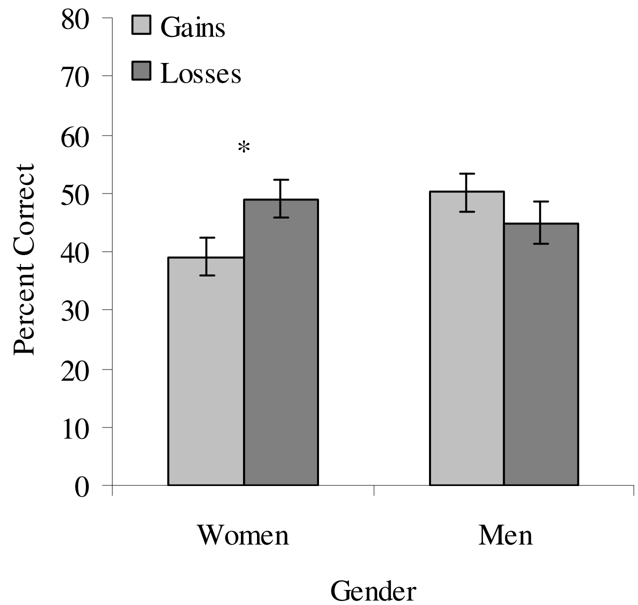


Figure 5.1. Percent correct on the GRE for men and women in gains and losses test environments

Women in the losses version of the test got more test problems correct as compared to women in the gains version of the test, $F(1,39) = 7.23$, $MSE = 146.09$, $p < .05$. In contrast, men in the gains test got more problems correct than men in the losses test. However, this difference is not statistically significant. This is exactly the pattern of data predicted. Women are assumed to have a negative stereotype and were predicted to be sensitive to a test with losses and non-losses emphasized and therefore perform better. The losses environment matches their underlying motivational focus.

I would also like to note the size of the difference between the gains and losses versions of the tests for female students. While the test was clearly difficult for this group of students, women in the losses test scored 10% better than women in the gains test. This clearly is a meaningful improvement.

Future work will continue to examine the role of regulatory fit effects on the performance of women in mathematics. The regulatory fit perspective suggests that there are two key methods for reducing or eliminating the performance decrements that result from regulatory mismatches (e.g., negative math stereotypes and gains environments). One is altering the elements that activate the prevention focus and the other is to alter the test environment. Our future work will investigate both of these methods. Importantly, we will work on developing strategies that female students can use when they encounter a situation that primes a prevention focus or a reward structure that mismatches their focus.

Appendix

REGULATORY FOCUS QUESTIONNAIRE (RFQ: HIGGINS ET AL., 2001)

This set of questions asks you HOW FREQUENTLY specific events actually occur or have occurred in your life. Please indicate your answer to each question by pressing the appropriate key.

	Response Options for Statements	Certainly False				Certainly True
	Response Options for Questions	Never or seldom		Sometimes		Very often
1	Compared to most people, are you typically unable to get what you want out of life?	1	2	3	4	5
2	Growing up would you ever “cross the line” doing things that your parents would not tolerate?	1	2	3	4	5
3	How often have you accomplished things that got you “psyched” to work even harder?	1	2	3	4	5
4	Did you get on your parents’ nerves often when you were growing up?	1	2	3	4	5
5	How often did you obey rules and regulations that were established by your parents?	1	2	3	4	5
6	Growing up, did you ever act in ways that your parents thought were objectionable?	1	2	3	4	5
7	Do you often do well at different things that you try?	1	2	3	4	5
8	Not being careful enough has gotten me into trouble at times.	1	2	3	4	5
9	When it comes to achieving things that are important to me, I find that I don’t perform as well as I ideally would like to.	1	2	3	4	5
10	I feel like I have made progress toward being successful in my life.	1	2	3	4	5
11	I have found very few hobbies or activities in my life that can capture my interest or motivate me to put effort into them.	1	2	3	4	5

BECK ANXIETY INVENTORY (BAI: BECK ET AL., 1988)

For the following items, give an answer from the 1 through 4 scale indicating how much you have been bothered by each symptom during the PAST WEEK, INCLUDING TODAY, by pressing the appropriate number key.

		Not at all	Mildly	Moderately	Severely
1	Numbness or tingling	1	2	3	4
2	Feeling hot	1	2	3	4
3	Wobbliness in legs	1	2	3	4
4	Unable to relax	1	2	3	4
5	Fear of the worst happening	1	2	3	4
6	Dizzy or lightheaded	1	2	3	4
7	Heart pounding or racing	1	2	3	4
8	Unsteady	1	2	3	4
9	Terrified	1	2	3	4
10	Nervous	1	2	3	4
11	Feelings of choking	1	2	3	4
12	Hands trembling	1	2	3	4
13	Shaky	1	2	3	4
14	Fear of losing control	1	2	3	4
15	Difficulty breathing	1	2	3	4
16	Fear of dying	1	2	3	4
17	Scared	1	2	3	4
18	Indigestion or discomfort in abdomen	1	2	3	4
19	Faint	1	2	3	4
20	Face flushed	1	2	3	4
21	Sweating (not due to heat)	1	2	3	4

PENN STATE WORRY QUESTIONNAIRE (PSWQ: MEYER ET AL., 1990)

For the following items, give an answer from the 1 through 5 scale that best describes how typical or characteristic each item is of you by pressing the appropriate number key.

		Not at all typical		Somewhat typical		Very typical
1	If I don't have enough time to do everything, I don't worry about it.	1	2	3	4	5
2	My worries overwhelm me.	1	2	3	4	5
3	I don't tend to worry about things.	1	2	3	4	5
4	Many situations make me worry.	1	2	3	4	5
5	I know I shouldn't worry about things, but I just can't help it.	1	2	3	4	5
6	When I am under pressure I worry a lot.	1	2	3	4	5
7	I am always worrying about something.	1	2	3	4	5
8	I find it easy to dismiss worrisome thoughts.	1	2	3	4	5
9	As soon as I finish one task, I start to worry about everything else I have to do.	1	2	3	4	5
10	I never worry about anything.	1	2	3	4	5
11	When there is nothing more I can do about a concern, I don't worry about it anymore.	1	2	3	4	5
12	I've been a worrier all my life.	1	2	3	4	5
13	I notice that I have been worrying about things.	1	2	3	4	5
14	Once I start worrying, I can't stop.	1	2	3	4	5
15	I worry all the time.	1	2	3	4	5
16	I worry about projects until they are all done.	1	2	3	4	5

POSITIVE AND NEGATIVE AFFECT SCHEDULE (PANAS: WATSON ET AL., 1988)

This scale consists of a number of words that describe different feelings and emotions. Read each item and press the appropriate number key. Indicate to what extent you feel this way CURRENTLY.

		Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
1	Interested	1	2	3	4	5
2	Distressed	1	2	3	4	5
3	Excited	1	2	3	4	5
4	Upset	1	2	3	4	5
5	Strong	1	2	3	4	5
6	Guilty	1	2	3	4	5
7	Scared	1	2	3	4	5
8	Hostile	1	2	3	4	5
9	Enthusiastic	1	2	3	4	5
10	Proud	1	2	3	4	5
11	Irritable	1	2	3	4	5
12	Alert	1	2	3	4	5
13	Ashamed	1	2	3	4	5
14	Inspired	1	2	3	4	5
15	Nervous	1	2	3	4	5
16	Determined	1	2	3	4	5
17	Attentive	1	2	3	4	5
18	Jittery	1	2	3	4	5
19	Active	1	2	3	4	5
20	Afraid	1	2	3	4	5

References

Aaker, J. L., & Lee, A. Y. (2006). Understanding regulatory fit. *Journal of Marketing Research, 43*, 15-19.

Akaike, H. (1974). A new look at the statistical model identification. *Transactions on Automatic Control, 19*, 716-723.

Arndt, J., Lieberman, J. D., Cook, A., & Solomon, S. (2005). Terror management in the courtroom - Exploring the effects of mortality salience on legal decision making. *Psychology Public Policy and Law, 11*(3), 407-438.

Aronson, J., Lustina, M. J., Good, C., Keough, K., & Steele, C. M. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*(1), 29-46.

Ashby, F. G. (Ed.). (1992). *Multidimensional models of categorization*. Hillsdale, NJ: Erlbaum.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442-481.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology, 37*, 372-400.

Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review, 6*(3), 363-378.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28*, 403-450.

Avila, C., & Parcet, M. A. (2002). Individual differences in reward sensitivity and attentional focus. *Personality and Individual Differences, 33*(6), 979-996.

Avnet, T., & Higgins, E. T. (2006). How regulatory fit affects value in consumer choices and opinions. *Journal of Marketing Research, 43*(1), 1-10.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*(2), 230-244.

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measureing clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*, 893-897.

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science, 16*, 101-105.

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin, 32*(8), 1059-1071.

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General, 133*, 584-600.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*(2), 174-181.

Brehm, J. W., & Self, E. A. (1989). The Intensity of Motivation. *Annual Review of Psychology, 40*, 109-131.

Brockner, J., Paruchuri, S., Idson, L. C., & Higgins, E. T. (2002). Regulatory focus and the probability estimates of conjunctive and disjunctive events. *Organizational Behavior and Human Decision Processes, 87*, 5-24.

Brown, R. P., & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76*(2), 246-257.

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary Determinants of Attitudes .2. Arm Flexion and Extension Have Differential-Effects on Attitudes. *Journal of Personality and Social Psychology, 65*(1), 5-17.

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*(2), 267-285.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 16*(7), 572-578.

Camacho, C. J., Higgins, E. T., & Luger, L. (2003). Moral value transfer from regulatory fit: What feels right is right and what feels wrong is wrong. *Journal of Personality and Social Psychology, 84*(3), 498-510.

Carver, C. S., & Scheier, M. F. (1990). Origins and Functions of Positive and Negative Affect - a Control-Process View. *Psychological Review, 97*(1), 19-35.

Cesario, J., Grant, H., & Higgins, E. T. (2004). Regulatory fit and persuasion: Transfer from "feeling right". *Journal of Personality and Social Psychology, 86*, 388-404.

Cincotta, C. M., & Seger, C. A. (2007). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of Cognitive Neuroscience, 19*(2), 249-265.

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*(5791), 1307-1310.

Croizet, J. C., Desert, M., Dutrevis, M., & Leyens, J. P. (2001). Stereotype threat, social class, gender, and academic under-achievement: When our reputation catches up to us and takes over. *Social Psychology of Education, 4*, 295-310.

Cross, S. E., & Madson, L. (1997). Models of the self: Self-construals and gender. *Psychological Bulletin, 122*(1), 5-37.

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin, 28*(12), 1615-1628.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34*(3), 169-189.

Elliot, A. J., & Church, M. A. (2003). A motivational analysis of defensive pessimism and self-handicapping. *Journal of Personality, 71*(3), 369-396.

Fein, S., von Hippel, W., & Spencer, S. J. (1999). To stereotype or not to stereotype: Motivation and stereotype activation, application, and inhibition. *Psychological Inquiry, 10*(1), 49-54.

Filoteo, J. V., Maddox, W. T., Simmons, A. N., Ing, A. D., Cagigas, X. E., Matthews, S., & Paulus, M. P. (2005). Cortical and subcortical brain regions involved in rule-based category learning. *Neuroreport, 16*(2), 111-115.

Forster, J., & Higgins, E. T. (2005). How global versus local perception fits regulatory focus. *Psychological Science, 16*(8), 631-636.

Förster, J., & Higgins, E. T. (2005). How global versus local perception fits regulatory focus. *Psychological Science, 16*, 631-636.

Forster, J., Higgins, E. T., & Idson, L. C. (1998). Approach and avoidance strength during goal attainment: Regulatory focus and the "goal looms larger" effect. *Journal of Personality and Social Psychology, 75*(5), 1115-1131.

Forster, J., Higgins, E. T., & Strack, F. (2000). When stereotype disconfirmation is a personal threat: How prejudice and prevention focus moderate incongruency effects. *Social Cognition, 18*(2), 178-197.

Freitas, A. L., Liberman, N., & Higgins, E. T. (2002). Regulatory fit and resisting temptation during goal pursuit. *Journal of Experimental Social Psychology, 38*(3), 291-298.

Friedman, R. S., & Forster, J. (2001). The effects of promotion and prevention cues on creativity. *Journal of Personality and Social Psychology, 81*, 1001-1013.

Geodert, K. M., Grimm, L. R., Markman, A. B., & Spellman, B. A. (2007). *Self-contrual and the processing of base rate information in a contingency learning task.* Paper presented at the Posert presented at the 29th Annual Meeting of the Cognitive Science Society, Nashville, TN.

Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L., & Breus, M. J. (1994). Role of Consciousness and Accessibility of Death-Related Thoughts in Mortality Salience Effects. *Journal of Personality and Social Psychology, 67*(4), 627-637.

Grimm, L. R., Markman, A. B., Maddox, W. T., & Baldwin, G. C. (in press). Differential effects of regulatory fit on category learning. *Journal of Experimental Social Psychology*.

Higgins, E. T. (1987). Self-Discrepancy - a Theory Relating Self and Affect. *Psychological Review, 94*(3), 319-340.

Higgins, E. T. (1997a). Beyond pleasure and pain. *American Psychologist, 52*(12), 1280-1300.

Higgins, E. T. (1997b). Beyond pleasure and pain. *American Psychologist, 52*, 1280-1300.

Higgins, E. T. (2000a). Making a good decision: Value from fit. *American Psychologist, 55*, 1217-1230.

Higgins, E. T. (2000b). Social cognition: Learning about what matters in the social world. *European Journal of Social Psychology, 30*(1), 3-39.

Higgins, E. T. (2005). Value from regulatory fit. *Current Directions in Psychological Science, 14*, 209-213.

Higgins, E. T., Friedman, R. S., Harlow, R. E., Idson, L. C., Ayduk, O. N., & Taylor, A. (2001). Achievement orientations from subjective histories of success: promotion pride versus prevention pride. *European Journal of Social Psychology, 31*(1), 3-23.

Higgins, E. T., Idson, L. C., Freitas, A. L., Spiegel, S., & Molden, D. C. (2003). Transfer of value from fit. *Journal of Personality and Social Psychology, 84*, 1140-1153.

Higgins, E. T., Klein, R., & Strauman, T. (1985). Self-Concept Discrepancy Theory - a Psychological Model for Distinguishing among Different Aspects of Depression and Anxiety. *Social Cognition, 3*(1), 51-76.

Higgins, E. T., Roney, C. J. R., Crowe, E., & Hymes, C. (1994). Ideal Versus Ought Predilections for Approach and Avoidance - Distinct Self-Regulatory Systems. *Journal of Personality and Social Psychology, 66*(2), 276-286.

Higgins, E. T., Shah, J., & Friedman, R. (1997). Emotional responses to goal attainment: Strength of regulatory focus as moderator. *Journal of Personality and Social Psychology, 72*(3), 515-525.

Hyde, J. S., & Kling, K. C. (2001). Women, motivation, and achievement. *Psychology of Women Quarterly, 25*(4), 364-378.

Idson, L. C., Liberman, N., & Higgins, E. T. (2004). Imagining how you'd feel: The role of motivational experiences from regulatory fit. *Personality and Social Psychology Bulletin, 30*(7), 926-937.

Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95*(4), 796-805.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle - Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16*(3), 175-179.

Josephs, R. A., Newman, M. L., Brown, R. P., & Beer, J. M. (2003). Status, testosterone, and human intellectual performance: Stereotype threat as status concern. *Psychological Science, 14*(2), 158-163.

Keller, J., & Bless, H. (2006). Regulatory fit and cognitive performance: The interactive effect of chronic and situationally induced self-regulatory mechanisms on test performance. *European Journal of Social Psychology, 36*(3), 393-405.

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29*(3), 371-381.

Kim, K. (2005). Fear of negative evaluation, subject size of social network, and risk taking. *DIssertation Abstracts International: Section B: The Sciences and Engineering, 66(5-B)*, 2848.

Kim, K., Grimm, L. R., & Markman, A. B. (2007). Self-construal and the processing of covariation information in causal reasoning. *Memory and Cognition, 35*, 1337-1343.

Kim, K., & Markman, A. B. (2006). Differences in Fear of Isolation as an explanation of cultural differences: Evidence from memory and reasoning. *Journal of Experimental Social Psychology, 42*(3), 350-364.

Kruglanski, A. W. (2006). The nature of fit and the origins of "feeling right": A goal-systemic perspective. *Journal of Marketing Research, 43*(1), 11-14.

Landau, M. J., Goldenberg, J. L., Greenberg, J., Gillath, O., Solomon, S., Cox, C., Martens, A., & Pyszczynski, T. (2006). The Siren's call: Terror management and the threat of men's sexual attraction to women. *Journal of Personality and Social Psychology, 90*(1), 129-146.

Lee, A. Y., & Aaker, J. L. (2004). Bringing the frame into focus: The influence of regulatory fit on processing fluency and persuasion. *Journal of Personality and Social Psychology, 86*(2), 205-218.

Lee, A. Y., Aaker, J. L., & Gardner, W. L. (2000). The pleasures and pains of distinct self-construals: The role of interdependence in regulatory focus. *Journal of Personality and Social Psychology, 78*(6), 1122-1134.

Leyens, J. P., Desert, M., Croizet, J. C., & Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Personality and Social Psychology Bulletin, 26*, 1189-1199.

Liberman, N., Idson, L. C., Camacho, C. J., & Higgins, E. T. (1999a). Promotion and prevention choices between stability and change. *Journal of Personality and Social Psychology, 77*(6), 1135-1145.

Liberman, N., Idson, L. C., Camacho, C. J., & Higgins, E. T. (1999b). Promotion and prevention choices between stability and change. *Journal of Personality and Social Psychology, 77*, 1135-1145.

Liberman, N., Molden, D. C., Idson, L. C., & Higgins, E. T. (2001). Promotion and prevention focus on alternative hypotheses: implications for attributional functions. *Journal of Personality and Social Psychology, 80*(1), 5-18.

Lockwood, P., Jordan, C. H., & Kunda, Z. (2002). Motivation by positive or negative role models: Regulatory focus determines who will best inspire us. *Journal of Personality and Social Psychology, 83*(4), 854-864.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics, 53*, 49-70.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes, 66*, 309-332.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes, 66*(3), 309-332.

Maddox, W. T., Baldwin, G. C., & Markman, A. B. (2006). A test of the regulatory fit hypothesis in perceptual classification learning. *Memory and Cognition, 34*, 1377-1397.

Maddox, W. T., Markman, A. B., & Baldwin, G. C. (2007). Using classification to understand the motivation-learning interface. *Psychology of Learning and Motivation, 47*, 213-249.

Markman, A. B., Baldwin, G. C., & Maddox, W. (2005). The interaction of payoff structure and regulatory focus in classification. *Psychological Science, 16*, 852-855.

Markman, A. B., & Brendl, C. M. (2000). The influence of goals on value and choice. *Psychology of Learning and Motivation: Advances in Research and Theory, Vol 39, 39*, 97-128.

Markman, A. B., Maddox, W. T., & Baldwin, G. C. (2005). The implications of advances in research on motivation for cognitive models. *Journal of Experimental and Theoretical Artificial Intelligence, 17*, 371-384.

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science, 17*, 944-948.

Markus, H. R., & Kitayama, S. (1991). Culture and the Self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224-253.

Marx, D. M., Stapel, D. A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparisons under threat. *Journal of Personality and Social Psychology, 88*(3), 432-446.

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.

Meyer, T. J., Miller, M. L., & Metzger, R. L. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28*, 487-495.

Miller, N. E. (1959). Liberalization of basic S-R concepts: Extensions to conflict behavior, motivation, and social learning. In S. Koch (Ed.), *Psychology: A study of a science. General and systematic formulations, learning, and special processes* (pp. 196-292). New York: McGraw Hill.

Narvaez, L. R., & Markman, A. B. (2006). *Individual differences lead to increased context sensitivity in causal induction.* Paper presented at the 47th Annual Meeting of the Psychonomic Society, Houston, TX.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29*(6), 782-789.

Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature, 433*(7028), 873-876.

Poldrack, R. A., & Packard, M. G. (2003). Competition among multiple memory systems: converging evidence from animal and human brain studies. *Neuropsychologia, 41*(3), 245-251.

Poldrack, R. A., & Rodriguez, P. (2004). How do memory systems interact? Evidence from human classification learning. *Neurobiology of Learning and Memory, 82*(3), 324-332.

Poy, R., Eixarch, M. D., & Avila, C. S. (2004). On the relationship between attention and personality: covert visual orienting of attention in anxiety and impulsivity. *Personality and Individual Differences, 36*(6), 1471-1481.

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.

Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T., & Lyon, D. (1989). Evidence for Terror Management Theory .1. The effects of mortality salience on reactions to those who violate or uphold cultural-values. *Journal of Personality and Social Psychology, 57*(4), 681-690.

Sassenberg, K., Jonas, K. J., Shah, J. Y., & Brazy, P. C. (2007). Why some groups just feel better: The regulatory fit of group power. *Journal of Personality and Social Psychology, 92*(2), 249-267.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*(3), 440-452.

Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles, 50*(11-12), 835-850.

Seger, C. A. (2006). The basal ganglia in human learning. *Neuroscientist, 12*(4), 285-290.

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience, 25*(11), 2941-2951.

Seger, C. A., & Cincotta, C. M. (2006). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cerebral Cortex, 16*(11), 1546-1555.

Seibt, B., & Forster, 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.

Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *Journal of Neuroscience, 27*(18), 4826-4831.

Shah, J., & Higgins, E. T. (1997). Expectancy x value effects: Regulatory focus as determinant of magnitude and direction. *Journal of Personality and Social Psychology, 73*(3), 447-458.

Shah, J., Higgins, E. T., & Friedman, R. S. (1998). Performance incentives and means: How regulatory focus influences goal attainment. *Journal of Personality and Social Psychology, 74*, 285-293.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*(1), 80-83.

Sinclair, L., & Kunda, Z. (1999). Reactions to a black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology, 77*(5), 885-904.

Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review, 16*(3), 177-206.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4-28.

Spiegel, S., Grant-Pillow, H., & Higgins, E. T. (2004). How regulatory fit enhances motivational strength during goal pursuit. *European Journal of Social Psychology, 34*(1), 39-54.

Steele, C. M. (1997). A threat in the air - How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613-629.

Steele, C. M., & Aronson, J. (1995). Stereotype Threat and the Intellectual Test-Performance of African-Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811.

Steele, J., James, J. B., & Barnett, R. C. (2002). Learning in a man's world: Examining the perceptions of undergraduate women in male-dominated academic areas. *Psychology of Women Quarterly, 26*(1), 46-50.

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.

Takane, Y., & Shibayama, T. (1992). *Structure in stimulus identification data*. Hillsdale: Erlbaum.

Torrubia, R., Avila, C., Molto, J., & Caseras, X. (2001). The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences, 31*(6), 837-862.

Walton, G. M., & Cohen, G. L. (2003). Stereotype Lift. *Journal of Experimental Social Psychology, 39*(5), 456-467.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology, 92*(1), 82-96.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of belief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063-1070.

Whaley, A. L. (1998). Issues of validity in empirical tests of stereotype threat theory. *American Psychologist, 53*(6), 679-680.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*(6), 797-826.

Wisman, A., & Goldenberg, J. L. (2005). From the grave to the cradle: Evidence that mortality salience engenders a desire for offspring. *Journal of Personality and Social Psychology, 89*(1), 46-61.

Wraga, M., Helt, M., Duncan, L. E., & Jacobs, E. C. (2006). *Altering test instructions to improve women's spatial cognition performance.* Paper presented at the 47th Annual Meeting of the Psychonomic Society, Houston, TX.

Yamawaki, N., Tschanz, B., & Feick, D. (2004). Defensive pessimism, self-esteem instability, and goal strivings. *Cognition & Emotion, 18*(2), 233-249.

Vita

Lisa Renee Grimm Narvaez was born in Washington, DC on February 17, 1980, the daughter of Lynn Ann Grimm and Stanley Brian Grimm. After completing her International Baccalaureate Diploma at George Mason High School, Falls Church, Virginia, in 1998, she entered Grinnell College in Grinnell, Iowa. During the fall of 2000, she attended the University of Stirling in Stirling, Scotland. She received the degree of Bachelor of Arts from Grinnell College in December, 2001. In December, 2004, she earned the degree of Master of Arts in Psychology from The Graduate School at The University of Texas.

Permanent Address: 805 Maryland Dr.

Austin, Texas 78758

This dissertation was typed by the author.