Stereotype Threat Reinterpreted as a Regulatory Mismatch

Lisa R. Grimm

Arthur B. Markman

W. Todd Maddox

Grant C. Baldwin

University of Texas at Austin

[JPSP: in press]

Corresponding author:

Lisa R. Grimm
Department of Psychology
University of Texas at Austin
1 University Station A8000
Austin, TX 78712
grimmlr@mail.utexas.edu

Abstract

Research documents performance decrements resulting from the activation of a negative task-relevant stereotype. We combine a number of strands of work to identify causes of stereotype threat in a way that allows us to reverse the effects and improve the performance of individuals with negative task-relevant stereotypes. We draw on prior work suggesting that negative stereotypes induce a prevention focus, and other research suggesting that people exhibit greater flexibility when their regulatory focus matches the reward structure of the task. This work suggests that stereotype threat effects emerge from a prevention focus combined with tasks that have an explicit or implicit gains reward structure. We find flexible performance can be induced in individuals who have a negative task-relevant stereotype by using a losses reward structure. We demonstrate the interaction of stereotypes and the reward structure of the task using chronic stereotypes and GRE math problems (Experiment 1), and primed stereotypes and a category learning task (Experiments 2a and 2b). We discuss implications of this research for other work on stereotype threat.

Keywords: Regulatory Fit, Stereotype Threat, Motivation, Category Learning, Math

Stereotype Threat Reinterpreted as a Regulatory Mismatch

Stereotypes are a pervasive part of human psychological experience. 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, Lustina, Good, Keough, & Steele, 1999; Stone, Lynch, Sjomeling, & Darley, 1999). Not confined to laboratory settings, stereotype threat effects can be found in real-world contexts. 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 that women are avoiding the possibility of confirming a negative stereotype about their group by switching into fields like the social sciences that do not have negative stereotypes for women.

Because stereotypes are ubiquitous, it is imperative that researchers determine how to mitigate their negative effects on performance. We present data in support of one such method. Using Regulatory Focus Theory (Higgins, 1987, 1997), we suggest that stereotype threat effects are the result of a regulatory mismatch between the motivational state of the individual and the reward structure of the task. This explanation allows us to suggest a straightforward method to reverse stereotype threat effects. Simply, we demonstrate that negative stereotypes can produce better performance than positive ones given a "matching" task reward structure. We call the beneficial pairing of stereotype and task reward structure a *stereotype fit.* This result has profound practical implications, because in real-world contexts it may be possible to change the reward structure of a task without changing any other fundamental task characteristics or underlying group stereotypes thereby improving performance by negatively-stereotyped groups.

In this paper, we begin with an overview of stereotype threat effects (Steele & Aronson, 1995) and a brief review of Regulatory Focus Theory (Higgins, 1987, 1997). We review and elaborate on the concept of regulatory fit (Higgins, 1997, 2000; Higgins, Idson, Freitas, Spiegel, & Molden, 2003; Maddox, Baldwin, & Markman, 2006). Finally, we present our experiments that test the interaction of stereotypes and task reward structure using GRE math problems and a classification task that requires flexible processing and discuss the implications of our results. We find that the impact of negative or positive stereotypes depends directly on the nature of the task environment. For high performance domains, like academic testing situations, the task environment is very important and can be manipulated easily. This provides one method for eradicating the performance decrements documented when individuals encounter a negative stereotype.

Stereotype Threat

Starting with Steele and Aronson (1995), laboratory research documents that the activation of a negative task-relevant stereotype has an adverse effect on participants' performance on tasks. In Steele and Aronson's studies, Black participants underperformed White participants on tests of intellectual ability when the test was framed as diagnostic of their ability. This framing activates the cultural stereotype that Black participants should underperform relative to White participants on tests of intelligence.

This paradigm can be applied generally when groups have task-relevant negative stereotypes even when the groups are not typically disadvantaged. Aronson, Lustina, Good, and Keough (1999) found that White men, who were told that the purpose of the experiment was to study the superiority of Asians on mathematics tests, scored worse on a math test as compared to men in the control group. In a different domain, Stone, Lynch, Sjomeling, and Darley (1999) demonstrated that Black participants performed worse than the control condition when a golf task was framed as diagnostic of "sports intelligence," but better than the control when the task was framed as diagnostic of "natural athletic ability." In contrast, White participants performed worse than the control condition when the task was framed as diagnostic of "natural athletic ability."

Researchers have manipulated stereotype threat in a number of ways. The most subtle manipulation merely asks participants to note 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 be 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.

Research on this phenomenon has led to a number of theories that aim to explain stereotype threat. For stereotype-threat to occur, researchers argue that 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, Spencer, Quinn, & Gerhardstein, 2002), and the environment needs to allow for stereotype confirmation in that the stereotype should be applicable (Ben-Zeev, Fein, & Inzlicht, 2005; Spencer, Steele, & Quinn, 1999). For example, Inzlicht and Ben-Zeev (2003) argue that women in mixed-gender environments are more likely to exhibit behaviors consistent with stereotype-threat than are women in same-gender settings.

A number of social-cognitive explanations for stereotype threat have been put forward, such as participants putting forward too much effort or too little effort, self-handicapping, anxiety, and low performance confidence (Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; Smith, 2004). Studies also suggest a connection between the represented stereotype and the corresponding stereotypic behavior (Bargh, Chen, & Burrows, 1996; Wheeler & Petty, 2001). For example, Cadinu et al. (2003) argue that stereotype threat effects occur because of lower performance expectancies, and Schmader, Johns, and Barquissau (2004) provide behavioral data differentiating individuals based on stereotype endorsement. 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.

In addition, some work has related stereotype threat to working memory (Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Schmader & Johns, 2003; Schmader, Johns, & Forbes, 2008). Schmader and Johns (2003) argued that stereotype threat effects are mediated by working memory capacity. Beilock, Jellison, Rydell, McConnell, and Carr (2006) extend this idea and demonstrate that the working memory impairment is caused by explicit monitoring of performance for tasks that have been proceduralized (also see Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003 for an earlier discussion of the role of divided attention). This claim is supported by demonstrations of the role of negative thinking under stereotype threat (Cadinu, Maass, Rosabianca, & Kiesner, 2005).

An important part of our research is that stereotype threat influences a person's motivational state. At present, there are a few motivational accounts of stereotype threat. Stereotype threat has been conceptualized as activation and inhibition of specific stereotypes based on active goals (Fein, von Hippel, & Spencer, 1999; Sinclair & Kunda, 1999). It has also been suggested that stereotype threat produces an increase in system arousal (see Brehm & Self, 1989 for a general discussion on the role of arousal) that affects performance on difficult tasks but not on easy ones (Ben-Zeev, Fein, & Inzlicht, 2005; O'Brien & Crandall, 2003).

Most relevant to our Experiments, 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. To evaluate this claim, we provide an overview of Regulatory Focus Theory.

Regulatory Focus Theory

Regulatory focus is a motivational mechanism that influences people's sensitivity to potential gains and losses in their environment (Higgins, 1987, 1997). 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; and Miller, 1959 for further discussion). Orthogonal to this distinction, Higgins (1987, 1997) 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*. People differ in the chronic accessibility of these foci, but often situations that have salient potential gains or losses may induce a regulatory focus that overcomes a person's chronic focus (Shah, Higgins, & Friedman, 1998).

Using this framework, Seibt and Förster (2004) advanced an insightful proposal that differences in regulatory focus cause stereotype threat effects. In a series of experiments, they demonstrated that priming individuals with a negative stereotype induces a prevention focus while priming individuals with a positive stereotype induces a promotion focus. On this view, decrements in performance on difficult cognitive tasks arise because the cognitive processes associated with a promotion focus are better-suited to performance on these tasks than are the cognitive processes associated with a prevention focus.

There are several reasons to believe that stereotype threat effects and regulatory focus are related phenomena. Other work has explicitly linked stereotype threat effects with regulatory focus (Forster, Higgins, & Strack, 2000; Keller & Dauenheimer, 2003) by examining the role of regulatory focus in the processing of stereotypic information (Forster, Higgins, & Strack, 2000) and by studying the mediation of stereotype threat by emotions induced by regulatory focus states (Keller & Dauenheimer, 2003). Further, a study of regulatory focus (Keller & Bless, 2006) and a study examining stereotype threat (Brown & Josephs, 1999) used the same manipulation. Brown and Josephs manipulated stereotype threat by framing a test as diagnostic of weak or strong ability. They argued 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. Keller and Bless manipulated situational focus using the same test framing. However, they argued that the weak ability condition primed a situational-prevention focus and the strong ability condition primed a situational-promotion focus.

More recent work on regulatory focus demonstrates that 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. For example, Higgins and colleagues found that the value people give to items in the environment depends on the fit between a person's regulatory focus and aspects of the items being evaluated (Forster, Higgins, & Idson, 1998; Higgins, 2000; Shah, Higgins, & Friedman, 1998). Higgins argues that a regulatory fit enhances task engagement, which increases the perceived value of the task (Higgins, 2000). On this view, match states feel better than mismatch states (Aaker & Lee, 2006; Cesario, Grant, & Higgins, 2004; Kruglanski, 2006; Sassenberg, Jonas, Shah, & Brazy, 2007).

Another form of fit between regulatory focus and tasks can occur 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 fit 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 fit between individuals with a prevention focus and tasks for which they must avoid losses, but a regulatory mismatch for those participants for tasks for which they must achieve gains. Some of these studies use chronic regulatory focus, while others induce a situational focus by having people try to achieve or try to avoid losing a raffle ticket to win money. The reward structure of the task is manipulated to either match or mismatch the regulatory focus by presenting participants with opportunities to gain or lose points for each response.

Stereotype Fit

Table 1 summarizes the interaction between regulatory focus and task reward structure. Our 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. 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.

These predictions are also supported by some prior work on positive stereotypes (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. In our Table 1, this effect is located in the leftmost column of Table 1. That is, individuals with positive stereotypes are expected to do well in gains tasks.

Much of the work on stereotype threat has been completed using verbal and math tests and has used a gains context. 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. 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.

Similarly, Spencer, Steele, and Quinn (1999) asked participants to take the GRE (see also Quinn & Spencer, 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 a gains environment and the incorrect scoring is a small loss that may not be well understood by participants. As such, this test context is more of a gains environment than a losses environment. Thus, we suggest that the female participants in this study had a situational prevention focus because of the negative stereotype. In contrast, men have a positive self-relevant stereotype (or perhaps no active self-stereotype), and so they are likely to have a promotion focus. Because this was gains environment, females were likely to be in a regulatory mismatch, but males were likely to be in a regulatory fit, and so women should (and did) perform worse than men on this task.

Our analysis suggests that if we assessed the performance of participants in a loss condition (the rightmost column of Table 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 because individuals with a negative stereotype are experiencing stereotype fit.

We test our predictions in two experiments. Experiment 1 uses problems from the quantitative GRE. We replicate the method used by Spencer et al. (1999) to create a situation where stereotypes would be active, thereby inducing regulatory foci. Students were told that they were going to take a math test given to a large group of students. Relying on the stereotype threat literature, we assume that women have a negative math stereotype, while men do not. We manipulated the task reward structure in a manner consistent with prior work on regulatory fit (Maddox et al., 2006). Half of the students gained more points for correct responses than incorrect responses (i.e., the gains version) while half lost fewer points for correct responses than incorrect responses (i.e., the losses version). We predict that women will perform better in the losses version of the GRE test than in the gains version, while men will show the opposite pattern of data. Importantly, this result would show that it is possible to improve the performance of women on a standardized test by altering the reward framing of the test, while leaving the actual test unchanged.

Experiments 2a and 2b further investigate our predictions and a possible mechanism for our effects. We transfer our results to a new domain, classification learning. We picked this domain because work in classification learning suggests that flexibility (defined as the degree to which people test many rules to correctly solve the task) may be a possible mechanism to explain the interaction of regulatory focus and reward structure (Grimm, Markman, Maddox, & Baldwin, 2008; Maddox et al., 2006). We discuss this mechanism in more detail in the introduction to Experiment 2. Furthermore, we have models from the classification literature that we can use to analyze participant responses (Ashby & Maddox, 1993; Maddox & Ashby, 1993). These models allow a more detailed understanding of how participants completed the classification task.

EXPERIMENT 1

This experiment examines performance on quantitative GRE problems. Previous research suggests that women have a negative math stereotype, while men do not. We manipulate the reward structure of the task, so that half of the participants gain points for each response, but get more points for correct than incorrect responses, and half of the participants lose points, but lose fewer points for correct responses than for incorrect responses. We predict that women will experience stereotype fit and perform better in the losses version of the GRE test than in the gains, while men will perform better in the gains version relative to the losses version. Furthermore, we predict that we will replicate the stereotype threat literature, as we interpret it, and find that men perform better than women in the gains version of the GRE test.

Method

Participants and Design

Seventy-nine undergraduate students (37 men and 42 women) at the University of Texas at Austin participated for course credit. Half of the women and 20 men were randomly assigned to the gains reward structure. The remaining participants were assigned to the losses reward structure. This Experiment had a 2 (Gender: Male, Female) × 2 (Reward Structure: Gains, Losses) design. Reward Structure was manipulated between subjects.

Materials and Procedure

Participants were tested in individual cubicles in a room containing approximately equal numbers of men and women. Participants first completed the Regulatory Focus Questionnaire (RFQ: Higgins et al., 2001), and questionnaires for two constructs, worry and anxiety, that have been linked to a prevention focus (Higgins, 1997), the Beck Anxiety Inventory (BAI: Beck, Epstein, Brown, & Steer, 1988), and the Penn State Worry Questionnaire (PSWQ: Meyer, Miller, & Metzger, 1990). We used the RFQ as a measure of chronic promotion and chronic prevention focus. This questionnaire asks participants to rate the frequency of specific events in their lives. The PSWQ requires the participants to rate how true displayed items are of them and 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”. We used all of these questionnaires to determine if there were any group differences prior to telling participants about the purpose of study.

Next, using a slightly altered stereotype manipulation from Spencer et al. (1999), participants were told, “We are developing some new tests and we are evaluating across a large group of University of Texas students. Today you will be taking a math test. This test is designed to be diagnostic of your math ability.” Participants in the gains condition were told that they would earn two points for each correct answer and zero points for each incorrect answer and that their goal was to get 36 points (e.g., 90% correct) and participants in the losses condition were told they would lose 1 point for each correct response and 3 points for each incorrect response and that their goal was to lose no more than 24 points (e.g., 90% correct).[1] To continue to the next screen, participants were told to press “F’ if they were female or “M” if they were male to continue.

Directly after reading about the math test, we asked participants to rate: “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, the participants took 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. The PANAS gives us a measure of the positive and negative affect prior to completing the problems.

Participants completed 20 questions from the quantitative section of the general section of the Graduate Record Examination (GRE). These problems assume knowledge of arithmetic, algebra, trigonometry, and geometry (Educational Testing Service, 2004). Problems were presented in a box on the left side of the screen one at a time. 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 answered a question, they heard a “ching” sound 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. Also, the 90% criterion line was labeled “36 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 “-24 points”. Samples of the gains and losses task screens are in Figure 1.

After the GRE test, we asked participants to report on a 9-point scale, anchored by strongly disagree and strongly agree, the extent they agreed with the following statements: (1) I am good at math and (2) It is important to me that I am good at math (see Spencer et al., 1999). Also, we asked subjects to report their typical grade in a math course. We collected this information after the GRE test, unlike Spencer et al., because we did not want these ratings interfering with our results.

Results

To test our hypotheses, we first report our analyses for the interaction of Gender and Reward using accuracy as a dependent measure. Next, we examine alternative explanations for our findings by looking at the individual difference measures collected before participants began the study. We consider whether chronic regulatory focus can account for our effects and examine whether there were prior group differences between men and women that might explain the results using Analysis of Covariance with questionnaire scores as potential covariates. We also consider the influence of math importance ratings using regression and break down our data set to include only those participants who endorsed the statements, “I am good at math” and “It is important to me that I am good at math.” We include these analyses to parallel those done by Spencer et al. (1999).

Stereotype Fit

The data were analyzed using an analysis of variance (ANOVA) with Gender (Male, Female) and Reward Structure (Gains, Losses) as between-participants’ factors. The dependent measure was the percent of problems correctly solved out of the number attempted. All participants attempted all 20 problems. This analysis revealed a significant two-way interaction between Gender and Reward Structure, $F(1,75) = 6.46$, $MSE = 249.26$, $p < .05$ (see Figure 2). To examine this interaction, we compared the average accuracy scores within each gender for gains and losses. As predicted, women (i.e., negative math stereotype) who performed the losses GRE test performed significantly better ($M = 50.0$) than women who performed the gains GRE test ($M = 37.62$) $F(1,40) = 6.45$, $p < .05$. There was not a statistically reliable difference for the men in the gains ($M = 50.75$) and losses ($M = 45.0$) tests, despite being in the predicted direction. Critically, we also tested within Reward Structure for Gender to replicate the classic stereotype threat effect. In the gains GRE test, men ($M = 50.75$) performed significantly better than women ($M = 37.62$) $F(1,39) = 7.09$, $p < .05$.

Chronic Regulatory Focus and other possible Mediators

We predict that the stereotypes activated in the testing situation, which induce the situational regulatory focus states, override the influence of chronic regulatory focus, which is assessed by the RFQ. The RFQ does not assess situationally-induced focus and we collected the RFQ prior to the experimental manipulation. To ensure that our observed differences did not reflect chronic regulatory focus, we used the RFQ to categorize participants as chronic promotion or chronic prevention. Those who scored higher on promotion relative to prevention were categorized as chronic promotion and vice versa. We analyzed our data using an ANOVA with Chronic Focus (Promotion, Prevention) and Reward Structure (Gains, Losses) between participants and percent correct as the dependent measure. The interaction between Chronic Focus and Reward Structure was not statistically significant, $F = .05$.

We analyzed the other questionnaire data collected during the experimental session. We found several pre-existing differences (i.e., prior to the stereotype-relevant task instructions) between the men and women in our sample. Women scored higher on the PSWQ ($M = 54.9$) than men ($M = 48.1$), $t\ (77) = 2.56$, $p < .05$; and higher on the BAI ($M = 33.7$) than men ($M = 29.3$), $t\ (77) = 2.83$, $p < .05$. After the description of the math test, women reported that they expected to like the task less ($M = 6.3$) than men ($M = 7.0$), $t\ (77) = 1.97$, $p = .053$.

While we find these differences interesting (and potentially important) we do not try to explain them here. Instead, we use Analysis of Covariance (ANCOVA) to determine whether the significant gender effects found in the questionnaire data could account for our interaction effect of interest. To this end, we completed ANCOVAs with Gender and Reward and each of the questionnaire scores above as continuous predictors (i.e., covariates) of task performance. We included the interaction between the covariate and Reward in our model to ensure that our interaction between Gender and Reward was estimated without bias (see Yzerbyt, Muller, & Judd, 2004 for a detailed discussion). When the PSWQ scores were used in an ANCOVA, there was an interaction of Gender and Reward ($F(1,73) = 9.48$, $MSE = 231.66$, $p < .05$), an interaction of Reward and PSWQ ($F(1,73) = 4.86$, $MSE = 231.66$, $p < .05$), and a main effect of Reward ($F(1,73) = 5.41$, $MSE = 231.66$, $p < .05$). The inclusion of BAI scores in an ANCOVA resulted in only an interaction of Gender and Reward ($F(1,73) = 7.27$, $MSE = 252.79$, $p < .05$). Lastly, when the liking scores were used in an ANCOVA, there was an interaction of Gender and Reward ($F(1,73) = 6.67$ $MSE = 235.84$, $p < .05$), and a main effect of Liking ($F(1,73) = 4.36$, $MSE = 235.84$, $p < .05$). These analyses demonstrate that our Gender x Reward Structure interaction is robust even after controlling for differences between men and women. In addition, performance expectations did not drive our effects. Women expected to perform worse, but in fact, performed just as well as men in the losses version of the task. Likewise, positive or negative affect did not influence our effects.

To align our theoretical perspective with previous work on stereotype threat, we examined the relationship between math importance and our effects. Math importance is positively correlated ($r = .4$) with accuracy. To examine this relationship, we ran a multiple regression using math importance (centered) as a continuous predictor, and Gender, Reward and the interaction of Gender and Reward as categorical predictors of percent correct. The regression was significant, $F(4,74) = 5.71$, $MSE = 216.53$, $p < .05$ and R-square = .24. Both math importance (B = 2.59) and the interaction component (B = -3.79) were significant predictors, $p < .05$, $t = 3.5$ and -2.26, respectively.

Second, we performed a median split and selected participants who more strongly endorsed the claims “I am good at math” and “It is important to me that I am good at math.” We had 18 women (8 in gains and 10 in losses) and 20 men (10 in both gains and losses) in this sample. The data were analyzed using an ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses) between participants and percent correct as the dependent measure. This analysis revealed a significant main effect of gender, $\underline{F}(1,34) = 4.71$, $\underline{MSE} = 189.92$, $\underline{p} < .05$, qualified by a two-way interaction between Gender and Reward Structure $\underline{F}(1,34) = 6.84$, $\underline{MSE} = 189.92$, $\underline{p} < .05$. Men ($\underline{M} = 56.5$) performed significantly better than women ($\underline{M} = 46.75$). To examine the interaction, we compared the average percent correct within each gender for gains and losses. As predicted, women in the losses GRE test performed significantly better ($\underline{M} = 53.5$) than women who performed the gains GRE test ($\underline{M} = 40.0$) $\underline{F}(1,34) = 4.26$, $\underline{p} < .05$. There was not a statistically reliable difference for the men in the gains and losses tests, $\underline{p} = .11$. Men in the gains task ($\underline{M} = 61.5$) performed better than men in the losses GRE test ($\underline{M} = 51.5$). Furthermore, analyzing the data in a manner consistent with stereotype threat, men performed significantly better than women in the gains GRE test $\underline{F}(1,16) = 10.82$, $\underline{p} < .05$.

Discussion

Men and women completed problems from the quantitative section of the general GRE. Half of the men and half of the women completed a gains version of the GRE test, while the remainder completed a losses version. We theorized that work done on stereotype threat has typically used a gains-type environment and that individuals with negative stereotypes underperform because they are experiencing a regulatory mismatch. Therefore, we predicted that we would replicate stereotype threat effects, with men performing better than women, in the gains version of our GRE test because men would be experiencing stereotype fit. However, we also predicted that women would experience stereotype fit in the losses GRE test and perform better than in the gains GRE test.

We found support for our interpretation of the stereotype threat literature and stereotype fit. In the gains version of the GRE test, men performed better than women as predicted. More importantly, women in the losses version performed 12.38% better than women in the gains version. This is a meaningful performance improvement. Moreover, women in the losses GRE test ($\underline{M}$ = 50.0) performed just as well as men in the gains GRE test ($\underline{M}$ = 50.75). This result suggests that our method can eliminate the classic stereotype threat effect by changing the task environment to produce a stereotype fit for those with negative task-relevant stereotypes.

We did not find a full cross-over interaction. Men in the gains GRE test did better than those in the losses GRE test, but not statistically. We believe that the negative math stereotype for women is stronger than the positive math stereotype for men. As such, the positive stereotype may not have influenced the behavior of men to the same degree. We do not believe stereotype fit effects to be unique to women and will explore this issue in Experiment 2.

To align our study with previous work on stereotype threat and math importance, we also focused our analyses on participants who endorsed statements about math importance and math ability. First, we found that both math importance and the interaction of Gender and Reward were significant predictors of our effects. Second, using a subset of our data in secondary analyses, which had less power because they included only approximately 10 participants per group, we still find an interaction of Gender and Reward for these participants, and a significant effect for women who do better in losses than gains. We do find a 10% advantage for men in the gains GRE test over men in the losses GRE test, but because of the small number of participants included in this analysis, the difference is not statistically reliable.

This study demonstrates a stereotype threat effect using a pre-existing stereotype and a task that people often perform outside of the lab. This study also connects directly to previous research that has used a similar paradigm. However, it is difficult to use this task to provide support for the claim that the root of this effect lies in the degree of flexibility engendered by the interaction of a motivational state created by a negative self-relevant stereotype and the reward structure of the task. To explore this issue more directly, we turn to an experimental setting that permits us to describe changes in people's behavior in a more fine-grained way.

For this purpose, we use a classification task in which participants learn to classify lines that vary in length, orientation, and position on the screen. We chose this domain because it is well-understood and there are data analytic models that provide a means to analyze the strategies participants use to solve the task. These qualities allow us a greater chance to uncover possible mechanisms behind our effects than was possible with the GRE problems used in Experiment 1.

Because there are no pre-existing stereotypes related to perceptual classification, we were able to create arbitrary stereotypes for participants. In Experiments 2a and 2b, we use gender stereotypes, but across studies, we vary the stereotype given. In Experiment 2a, participants are told that women are better at the classification task than men, while in Experiment 2b, participants are told that men are better at the classification task than women. Like Experiment 1, half of the participants gain points and half lose points. We predict that our effects will not only be true for chronic stereotypes but for primed stereotypes as well because both activate regulatory focus states. Therefore, we predict that the two-way interaction between Gender and Reward observed in Experiment 1 will go in different directions in Experiments 2a and 2b leading to a 3-way interaction.

We use a classification task from Maddox, Baldwin, and Markman (2006). In Maddox et al., 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 of the lines. A simple rule involving only the highly salient position dimension would yield good performance, but not sufficiently good performance to achieve the performance criterion. Thus, this task requires flexibility to stop using an obvious but suboptimal rule and to try less obvious but more effective strategies for classifying the items. Simply, participants need to continue to search the rule space until they find the correct rule to use to classify the items. Flexibly trying rules leads to better performance, because the participants must try and abandon a number of incorrect rules prior to discovering the correct one.

Maddox et al. (2006) gave participants either 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 or 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. 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. Participants with a regulatory fit (i.e., a promotion focus with gains, or a prevention focus with losses) performed better and were more likely to achieve the performance criterion than were participants with a regulatory mismatch (i.e., a prevention focus with gains or a promotion focus with losses).

An important reason for using this classification task is that it allows researchers to fit mathematical models to the data in order to describe the strategies used by individual participants on a block-by-block basis. Maddox, Baldwin, and Markman (2006) found that early in learning, participants' performance was best characterized as using a simple rule along one dimension. Later, participants learned to classify on the basis of the correct two-dimensional rule. Participants with a regulatory fit found the correct two-dimensional rule earlier in the task than did those with a regulatory mismatch. That is, they engaged in more flexible processing.

The study just described is one in a series of experiments demonstrating that regulatory fit leads to flexibility and exploration in a variety of settings including classification, decision making, and foraging (Grimm, Markman, Maddox, & Baldwin, 2008; Maddox, Baldwin, & Markman, 2006; Maddox, Markman, & Baldwin, 2007; Markman, Baldwin, & Maddox, 2005; Markman, Maddox, & Baldwin, 2005; Markman, Maddox, & Worthy, 2006; Markman, Maddox, Worthy, & Baldwin, 2007; Worthy, Maddox, & Markman, 2007). Across these studies, the effects of regulatory fit are nearly identical for participants with a promotion focus and a gains reward structure and participants with a prevention focus and a losses reward structure.

These regulatory fit findings are consistent with those from the literature on chronic regulatory focus and on stereotype threat if we apply our interpretation of the literatures. For example, if we assume that studies of chronic focus (i.e., promotion versus prevention) typically used gains tasks, then they were comparing promotion participants experiencing fit to prevention participants experiencing mismatch. 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, 1999), 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 the prediction that a promotion focus leads to greater creativity 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.

Applying our regulatory fit framework, if most tasks are implicit or explicit gains environments, then the evidence found in favor of a promotion focus supporting more elaborative/flexible/creative processing is in fact evidence for flexible processing in regulatory fit. Critically, flexible abstract processing is a hallmark of a regulatory fit, not of a promotion focus, just as detailed local processing is a hallmark of a regulatory mismatch, not of a prevention focus.

In the stereotype threat literature, many of the tasks used require flexible and elaborative processing, such as the verbal GRE (Steele and Aronson, 1995) and the quantitative GRE (Spencer et al., 1999). Most closely related to the present study, Quinn and Spencer (2001) found 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. Quinn and Spencer coded the number of problem solving strategies used by participants. They found that 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 on to our claim that participants in a regulatory mismatch (i.e., negative stereotype in a gains task) will display less flexible processing or rule testing as compared to participants in a regulatory match or stereotype fit (i.e., positive/neutral stereotype in a gains task).

To analyze our data from the classification task and test for evidence that flexibility is the mechanism responsible for our effects, we turn to decision-bound modeling (Ashby & Maddox, 1993) to uncover the strategies used by participants to classify lines. We use models to determine if more participants in a stereotype fit than in a stereotype mismatch find and use the correct rule to classify the stimuli. Finding this correct rule requires participants to test and discard simpler rules and then expand their problem space to test rules that use two dimensions. There is an established literature suggesting that people start with simple unidimensional rules and change to more complex rules in most classification tasks (Bruner, Goodnow, & Austin, 1956). 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 stereotype fit will be more likely to abandon the simple rules in favor of the more complex conjunctive rule.

EXPERIMENTS 2A AND 2B

To summarize our design and predictions, in this study we told participants about gender stereotypes that relate to their performance in a perceptual classification task. In Experiment 2a, we told male and female participants that this classification task is one for which women have previously been demonstrated to do better than men. In Experiment 2b, we presented participants with the opposite story, so participants were told that men perform better than women at this classification 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, we predict that participants with a negative task-relevant stereotype will have a stereotype fit when the task has a losses reward structure, and so they should perform better and be more likely to find and use the correct classification rule than when the task has a gains reward structure and they have a mismatch. In contrast, we predict that participants with a positive task-relevant stereotype will have a stereotype fit for the gains reward structure, and thus should perform better and be more likely to find and use the correct classification rule than when they perform the task with a losses reward structure and have a regulatory mismatch.

Method

Participants and Design

Eighty undergraduate students (40 men and 40 women) at the University of Texas at Austin were given \$8 for their participation in Experiment 2a and another group of 80 students (40 men and 40 women) at the University of Texas at Austin were given \$8 for participating in Experiment 2b. Half of the men and half of the women were randomly assigned to the gains and losses reward structures. Each 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 Figures 3 and 4. 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.[2] 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 results in 83% accuracy for a block of trials. For example, Figure 3 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 and all stimuli falling below the bound into Category B. These 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.

There is an optimal decision bound for this task that, if used, yields 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 4 for a graphical representation of this rule). In order for participants to perform well in the task, they need to abandon the use of easier unidimensional rules in favor of the more complex conjunctive one. This switch requires cognitive flexibility.[3]

Materials and Procedure

As for Experiment 1, participants were tested in individual cubicles in a room with approximately the same number of men and women. 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. Following the questionnaires, to induce a stereotype our participants in Experiment 2a read: "This is an experiment testing sex differences in spatial 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 Experiment 2b, all participants read: "This is an experiment testing sex differences in spatial abilities. Previous research has shown that men perform better than women on tests of spatial ability." This primes men with a positive task-relevant stereotype and women with a negative task-relevant stereotype. Participants in both Experiments read text on the screen requiring them to note their gender by pressing "F" for female and "M" for male to advance in the computer task.

In the gains version of each experiment, participants were told that the group assigned the positive stereotype tended to earn more than 86 points per block, which is equivalent to the 90% correct criterion (correct on 43 of 48 trials), and the other group tended to earn fewer. In the losses version, participants were told that the group assigned the positive stereotype tended to lose less than 58 points per block, which is again equivalent to the 90% criterion (correct on 43 of 48 trials), and the other group tended to lose more. Next, we asked participants to rate: "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)" and then participants took the PANAS to get a measure of their positive and negative affect prior to completing the classification task.

We used the same progress meter and stimulus presentation box from Experiment 1. Because a different number of points were available, in the gains condition the 90% criterion line was labeled "86 points". For participants in the losses task, the bonus criterion was labeled "-58 points".

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 Gender influenced a basic performance metric. We computed the first block that each participant met or exceeded the criterion (90% correct) and the average accuracy for each participant in each block of trials. Second, we used quantitative models to examine the strategies used by participants to learn the task. By identifying the strategies likely to be implemented by participants, we are able to make claims about the processes used during the perceptual classification learning task and the possible mechanisms of stereotype fit. Third, we consider the influence of chronic regulatory focus and other possible mediators.

Behavioral data and Stereotype Fit

To test the interaction of Gender and Reward Structure across Experiments, we analyzed the first block participants reached or exceeded the criterion using an ANOVA with Experiment (2a, 2b), Gender (Male, Female), and Reward Structure (Gains, Losses) between participants. Any participant who failed to meet the criterion during the experiment was coded as a 13. This was done because this was the minimum value possible for a participant who had not met the criterion during the course of the 12 block experiment. This analysis revealed a significant three-way interaction between Experiment, Gender, and Reward Structure, $\underline{F}(1,152) = 7.39$, $\underline{MSE} = 12.3$, $\underline{p} < .05$. To examine this three-way interaction, we looked for our predicted two-way interaction between Gender and Reward Structure within each Experiment. For Experiment 2a, an ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses) revealed the predicted interaction, $\underline{F}(1,152) = 4.56$, $\underline{MSE} = 12.3$, $\underline{p} < .05$. For Experiment 2b, an ANOVA with Gender (Male, Female) and Reward Structure (Gains, Losses) revealed the predicted interaction, $\underline{F}(1,152) = 2.98$, $\underline{MSE} = 12.3$, $\underline{p} = .08$ (marginally-significant).

Within each of these interactions, we examined group differences using independent samples t-tests. For Experiment 2a, 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 is 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) = 1.92, p < .05 (one-tailed)]. For Experiment 2b, 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 (one-tailed). 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 is not statistically reliable.

Second, while the preceding analyses focus on a global performance metric, this metric does not allow us to take advantage of the correlations that exist in our accuracy data over time. Each participant has a score for each of the 12 blocks of trials. To take advantage of these correlations across time, we performed a discriminant function analysis. This analysis creates a linear discriminant function that distinguishes the groups based on their data over time. That is, a function was generated using the accuracy data from the 12 blocks as continuous predictor variables; one variable representing each block of trials. Next, we used Bayes' rule and the discriminant function to predict to which experimental group each participant belonged. We then tested to see if the predictions were significantly better than chance assignment of participants to groups. If the predictions are above chance, then our groups differed significantly when the pattern of their accuracy data over the course of the experiment is taken into account.

First, we modeled the performance of participants in a stereotype fit and those in a mismatch. The model correctly classified 70% of the participants into these two groups in Experiment 2a and correctly classified 67.5% of the participants in Experiment 2b, both classifications are significantly greater than chance, $p < .05$ (chance classification is .5 because there are two groups).[4] Looking within Gender for each Experiment, we tested for whether the model could correctly classify gains and losses participants better than chance. In Experiment 2a, the model correctly classified men and women into gains and losses tasks 75% and 70% of the time, respectively, both significantly greater than chance, $p < .05$. In Experiment 2b, the model correctly classified men and women into gains and losses tasks 85% and 67.5% of the time, respectively, both significantly greater than chance, $p < .05$.

The reason for the good performance of the models is obvious when the overall patterns in the data are considered (see Figure 5). As predicted, for Experiment 2a, men in the losses task performed better than men in the gains task and in fact were more accurate in all 12 experimental blocks and women in the gains task outperformed women in the losses task and were more accurate on 10 of the 12 blocks (both significant using binomial sign tests, $p < .05$). Similarly, as predicted, for the gains task, women outperformed men on 9 of the 12 blocks of trials and performed equally well on one block, and for the losses task, men outperformed women on every block (both significant using binomial sign tests, $p < .05$).

For Experiment 2b, the pattern reverses. As predicted, men in the gains task performed better than men in the losses task in 11 of the 12 experimental blocks, and women in the losses task outperformed women in the gains task and obtained higher accuracy on 8 of the 12 blocks (male data significant using a binomial sign test, $p < .05$; female data pattern critical to modeling is obvious in the first four blocks). Again as predicted, for the gains task, men performed better than women on all 12 blocks of trials, and for the losses task women performed better than men on all 12 blocks of trials (both significant using binomial sign tests, $p < .05$).

Strategy use and Stereotype Fit

To test for specific strategy use by participants, we fit a series of decision-bound models to the data from each participant for each block (Ashby & Maddox, 1993; Maddox & Ashby, 1993).[5] The models used provided a good account of our data.[6] 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).

For Experiment 2a, Figure 6 (Panel A) displays the proportion of data sets best fit by a conjunctive rule model for men in the gains and losses classification tasks separately by block. Because men in the losses task are in a stereotype fit relative to men in the gains task, 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 $p < .05$. The opposite pattern was predicted for women. Specifically, women in the gains task are in a stereotype fit and should be more likely to use a conjunctive rule then women in the losses task 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 6 Panel B).

For Experiment 2b, as shown in Figure 6 (Panel C) and as predicted, for men, the binomial tests for blocks 2, 3, and 9 revealed the stereotype fit 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 task was better fit by the conjunctive rule more frequently than the data in the men losses task, $p < .05$, with a higher proportion of the participants likely using the conjunctive rule in 11 of the 12 blocks. 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 6 Panel D). A binomial sign test across blocks revealed that the women/losses task was not better fit by the conjunctive rule than the women/gains task.

Chronic Regulatory Focus and other possible Mediators

As for Experiment 1, we collected the RFQ as a measure of chronic regulatory focus before the experimental manipulation and created regulatory focus groups (i.e., promotion and prevention groups) using the RFQ. We expect that our manipulation of stereotypes removed any influence of chronic focus. To test this possibility, we examined the influence of chronic Regulatory Focus by testing the interaction of Regulatory Focus and Reward Structure across Experiments. We analyzed the first block participants reached or exceeded the criterion using an ANOVA with Experiment (2a, 2b), Chronic Regulatory Focus (Promotion, Prevention), and Reward Structure (Gains, Losses) between participants. This analysis revealed a non-significant three-way interaction between Experiment, Chronic Regulatory Focus, and Reward Structure, $\underline{F} = 1.17$. As such, we believe chronic Regulatory Focus cannot account for our effects.

We analyzed the questionnaire data collected during the experimental session. In Experiment 2a, we found several pre-existing differences (i.e., prior to the stereotype-relevant task instructions) between the men and women in our sample. Women scored higher on the Prevention subscale of the RFQ ($\underline{M} = 17.7$) than men ($\underline{M} = 15.9$), $\underline{t}\ (78) = 2.35$, $\underline{p} < .05$ and higher on the PSWQ ($\underline{M} = 51.9$) than men ($\underline{M} = 45.6$), $\underline{t}\ (78) = 2.26$, $\underline{p} < .05$. In Experiment 2b, women scored higher on the Prevention subscale of the RFQ ($\underline{M} = 17.5$) than men ($\underline{M} = 15.2$), $\underline{t}\ (54) = 2.04$, $\underline{p} < .05$.

In Experiment 2a, we found a significant interaction for the Negative 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 marginally significant two-way interaction between Gender and Reward Structure, $\underline{F}(1,76) = 3.56$, $\underline{MSE} = 26.0$, $\underline{p} = .06$. 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, $\underline{t}\ (38) = 1.88$, $\underline{p} = .06$. Lastly, in Experiment 2a, relative to men, women believed they performed worse ($\underline{M} = 6.3$) than men did ($\underline{M} = 7.1$), $\underline{t}\ (78) = 2.32$, $\underline{p} < .05$.

As for Experiment 1, we completed ANCOVAs to demonstrate that our Gender x Reward Structure interaction in the first block participants reached or exceeded the criterion is robust even after controlling for differences between men and women. Prevention scores were correlated with gender in both Experiments. The inclusion of prevention as a covariate resulted in an interaction of Experiment, Gender, and Reward, $F(1,126) = 7.46$, $MSE = 14.08$, $p < .05$, but there was neither a main effect of Prevention nor interactions of Prevention and Reward or of Prevention and Experiment. For covariates unique to Experiment 2a, the inclusion of PSWQ scores in an ANCOVA resulted in both an interaction of Gender and Reward ($F(1,74) = 6.38$, $MSE = 10.66$, $p < .05$) and a main effect of Gender ($F(1,74) = 4.23$, $MSE = 10.66$, $p < .05$). Similarly, including the Negative Affect scale of the PANAS resulted in both an interaction of Gender and Reward ($F(1,74) = 6.34$, $MSE = 10.64$, $p < .05$) and a main effect of Gender ($F(1,74) = 3.91$, $MSE = 10.64$, $p < .05$). Lastly, when performance expectation scores were used in an ANCOVA, there was interaction of Gender and Reward, $F(1,74) = 4.27$, $MSE = 10.12$, $p < .05$. There were no covariates unique to Experiment 2b.

These analyses demonstrate that our Gender x Reward Structure interaction is robust even after controlling for pre-existing differences between men and women. Likewise, positive affect did not influence our effects. Furthermore, as for Experiment 1, performance expectations did not drive our effects. Women expected to perform equally well in both Experiment 2a ($M = 6.2$) and 2b ($M = 6.3$), as did men ($M = 6.7$ and $M = 6.5$, respectively), despite performance differences. As such, our stereotype manipulation was not just influencing performance expectations, which then produced our effects.

Discussion

Using a primed stereotype, we found that women and men responded differently to the gains and losses reward structures in a classification task using task accuracy and proportion of participants reaching the task criterion. In this set of Experiments, we expected to replicate our results from Experiment 1 in a different domain using primed stereotypes instead of chronic stereotypes. We also predicted stereotype threat-consistent effects for the gains structure. We found results consistent with our interpretation of the stereotype threat literature and stereotype fit.

We found the predicted three-way interaction between Experiment, Gender, and Reward Structure for the first block participants reached the accuracy criterion. Furthermore, using discriminant function analysis we showed that participants' accuracy profile over blocks predicted their group membership. This analysis revealed that our groups differed significantly across time. In Experiment 2a, women (given a positive stereotype) outperformed men (given a negative stereotype) in the gains version of the task. In Experiment 2b, we reverse this pattern of data in all 12 experimental blocks by switching the gender stereotype. The opposite is true for the losses reward structure. In Experiment 2a, as predicted, men outperformed women on all 12 blocks of trials. Men have a stereotype fit in the losses task. In Experiment 2b, again when we switch the stereotype, we completely reverse the effect. Women performed better than men on all 12 blocks of trials. Lastly, our data analytic models demonstrate that the better task performance corresponded to more flexible strategy use.

General Discussion

In two experiments, we found results consistent with our interpretation of the stereotype threat literature and our concept of stereotype fit. Based on the prior work by Maddox, Baldwin, and Markman (2006), we predicted that individuals experiencing a regulatory fit would perform better in the tasks than participants in a regulatory mismatch. Like Seibt and Förster (2004), we argue that priming a negative stereotype induces a prevention focus while priming a positive stereotype induces a promotion focus. Participants completed GRE math problems in Experiment 1 and a rule-based perceptual classification task in Experiments 2a and 2b. Further, our participants completed a gains version of each task where they gained points for correct responses or a losses version of each task where they lost points for correct responses. For the gains version of the task, we predicted that participants with a positive stereotype would be experiencing a stereotype fit while participants with a negative stereotype would be experiencing a regulatory mismatch. We predicted the opposite would be true for the losses version of the task.

We suggest that most experimental tasks are gains environments (either implicitly or explicitly). As such, we expected to replicate stereotype threat effects in the gains versions of our tasks. Using GRE math problems (Experiment 1) and a classification task (Experiment 2a and 2b), we find the classic stereotype threat effect in the gains task. Women performed worse than men on GRE math problems in Experiment 1. In Experiment 2, when 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 (Experiment 2a). However, when we switched the valence of the stereotypes applied to gender, we got the predicted performance reversal: men outperformed women in all 12 blocks of trials (Experiment 2b).

We have further evidence for stereotype fit using the data from the losses versions of our tasks. Now, unlike the gains versions, participants with negative task-relevant stereotypes are experiencing a stereotype fit. In Experiment 1, women in the losses GRE test performed better than women in the gains GRE test, which coincidentally completely removed the performance difference between women and men in the gains GRE test. In Experiment 2, men in Experiment 2a and women in Experiment 2b were experiencing a stereotype fit in the losses task. In Experiment 2a, men outperformed women in all 12 blocks in the losses task and in Experiment 2b, women outperformed men in all 12 blocks of trials in the losses task.

Across these two Experiments, we have replicated our findings using different domains, math and classification learning, and obtained the same results using chronic and primed stereotypes. Our other goal was to uncover a possible mechanism behind our stereotype fit effects. One possibility is the ability to think more flexibly when in a stereotype fit.

In Experiment 1, our participants completed math problems from the GRE. One could argue that participants need to be able to think flexibly in order to solve these difficult problems. While this intuitively sounds correct, like many tasks used in psychology, this is a very complicated task that is not very well understood. It is not clear exactly what processes individuals use to solve problems and if there is a consistent way these problems are approached across people.

To test our flexibility hypothesis, in Experiment 2, 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 the participant to flexibly work in the rule space.

The modeling results support the flexibility hypothesis. In Experiment 2a, the female data in the gains task is more consistent with the use of conjunctive rules 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 2b, 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. While not ideal, we believe this result should be considered in the context of the rest of our strong results in support of stereotype fit.

We are excited about this line of work and hope that other researchers will join us to start investigating different possible mechanisms for stereotype fit. We present the flexibility hypothesis as one possibility. This hypothesis is consistent with other work on regulatory focus (Friedman and Forster, 2001; Forster and Higgins, 2005) and regulatory fit (Grimm et al., 2008; Maddox et al., 2006). Furthermore, in the regulatory fit literature, we have evidence that effects reverse when we use an information-integration category structure. Learning this type of category structure is hindered by flexible processing (Grimm et al., 2008).

Based on our questionnaire data, we believe that positive or negative affect, worry, anxiety, chronic regulatory focus, or performance ratings cannot account for our effects. While chronic focus may be important, in our studies, the situational context (e.g., the math test in Experiment 1 and the overt primes in Experiment 2) overrides the chronic focus state. Similarly, one might argue that our manipulation of stereotype threat in Experiment 2 more created performance expectancies than stereotype threat and these expectancies are creating our effects. However, in Experiments 1 and 2, participants expect to perform equally well, but perform better or worse than expected based on their experimental condition.

We realize our findings go against most of the literature on positive affect. Positive affect has been linked with both a promotion focus (Higgins, 1997) and with creativity (Isen, Johnson, Mertz, & Robinson, 1985). 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. As above, we argue positive affect may be a hallmark of a regulatory fit and not a promotion focus. This claim is supported by work demonstrating that people feel better when they are in a regulatory fit (Camacho, Higgins, & Luger, 2003; Higgins, 2000) and by research demonstrating similar neural mechanisms for positive affect and creativity (Ashby, Isen, & Turken, 1999) and flexibility in our category learning task (Maddox & Ashby, 2004). Furthermore, directly examining the connection between regulatory fit and creativity, Markman, Maddox, Worthy, and Baldwin (2007) manipulated regulatory focus and tested participants using the Remote Associates Test, which is the measure of creativity used in the positive affect study by Ashby et al. (1999). They found fit participants solved more problems than mismatch participants. We believe more work needs to be done to assess the exact relationship between regulatory fit and positive affect.

Theoretical Implications

Our results have theoretical, methodological and practical implications for the study of stereotype threat and cognition more generally. Theoretically, we hope cross-disciplinary work like this will lead to further research on the influence of motivation on cognitive processing. For example, one such variable, self-construal, has been linked to regulatory focus (Lee, Aaker, & Gardner, 2000) and to stereotype threat (Marx, Stapel, & Muller, 2005). By studying the relationships between these and other individual differences, researchers may be able to posit similar mechanisms. Of particular interest to us at present is the relationship that testosterone may play given work by Josephs and colleagues demonstrating the moderating influence of testosterone in stereotype threat (Josephs, Newman, Brown, & Beer, 2003) and the intersection between our work and the work by Beilock and her colleagues (Beilock & Carr, 2005; Beilock, Jellison, Rydell, McConnell, & Carr, 2006; Beilock, Kulp, Holt, & Carr, 2004). Beilock's research demonstrates the role of working memory in producing stereotype threat effects for proceduralized tasks. 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. We are investigating whether a working memory account can explain our regulatory fit effects. That is, whether working memory constraints are different for regulatory matches and mismatches.

Furthermore, the relationship between the positive affect and regulatory fit suggests that it will be important to assess the influence of stereotype fit on creativity. Our research suggests that placing individuals with negative stereotypes in a losses task will produce more creative performance than placing those individuals in a gains task. As in the regulatory fit literature, these individuals should "feel better" in a fit and feel like they can process information more fluently (Lee & Aaker, 2004).

Methodological Implications

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, Grimm, & Markman, 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 task, 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. We suggest that there are two ways to influence stereotype fit to improve performance: changing the regulatory focus induced and changing the reward structure of the task. Some manipulations of stereotype threat, like same-gender versus mixed testing situations (Inzlicht & Ben-Zeev, 2003), may be inducing regulatory focus states by highlighting stereotype relevance in the mixed environment. We join a group of researchers already working on ways to eradicate stereotype threat by providing obvious situational attributions for performance decrements (Brown & Josephs, 1999; Johns, Schmader, & Martens, 2005) or by reducing anxiety (Cohen, Garcia, Apfel, & Master, 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 et al. (2006) improved performance by African American students by reducing stress using self-affirmation techniques.

We do want to be very clear about the implications for changing the reward structure of testing situations for stigmatized individuals. We are not advocating punishment of stigmatized populations to improve performance. Clearly, like the classic Yerkes-Dodson arousal function, there are extreme implementations of our approach that would surely hurt performance. A serious losses situation, like a threat of punishment, would be expected to impact cognitive processing and swamp any effects of regulatory fit. We believe that our effects appear because the reward is directly linked to task performance and is relatively mild. These mild losses (and mild gains) allow for the expression of regulatory fit, in a way that either extreme may not.

Conclusions

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 for one’s group may be an excellent first step in curbing performance decrements caused by negative stereotype encounters.

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.

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*, 29-46.

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

Ashby, F. G., Isen, A. M., & Turken, U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*, 529-550.

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

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*, 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*, 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*, 246-257.

Bruner, J. S., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York: Wiley.

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

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*, 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*, 498-510.

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

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

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

Educational Testing Service. (2004). *GRE General Test Practice Book*. Princeton, NJ: Educational Testing Service.

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*, 49-54.

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*, 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*, 178-197.

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

Grimm, L. R., Markman, A. B., Maddox, W. T., & Baldwin, G. C. (2008). Differential Effects of Regulatory Fit on Category Learning. *Journal of Experimental Social Psychology, 44*, 920-927.

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

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

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

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*, 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.

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

Isen, A. M., Johnson, M. M. S., Mertz, E., & Robinson, G. F. (1985). The Influence of Positive Affect on the Unusualness of Word-Associations. *Journal of Personality and Social Psychology, 48*, 1413-1426.

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

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

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*, 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*, 371-381.

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.

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

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*, 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*, 1122-1134.

Liberman, N., Idson, L. C., Camacho, C. J., & Higgins, E. T. (1999). 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*, 5-18.

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., 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. In D.L. Medin (Ed.) *The Psychology of Learning and Motivation, Vol. 39.* (pp. 97-129) San Diego, CA: Academic Press.

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.

Markman, A. B., Maddox, W. T., Worthy, D. A., & Baldwin, G. C. (2007). Using regulatory focus to explore implicit and explicit processing in concept learning. *Journal of Consciousness Studies, 14*, 132-155.

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*, 432-446.

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.* Poster 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*, 782-789.

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*, 55-71.

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*, 249-267.

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

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

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336-356.

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*, 38-56.

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*, 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*, 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*, 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*, 4-28.

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

Steele, 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*, 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*, 1213-1227.

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

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

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.

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

Worthy, D. A., Maddox, W. T., & Markman, A. B. (2007). Regulatory fit effects in a choice task. *Psychonomic Bulletin & Review, 14*, 1125-1132.

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

Yzerbyt, V.Y., Muller, D., & Judd, C.M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of Experimental Social Psychology, 40*, 424-431.

Author Note

Lisa R. Grimm, Arthur B. Markman, W. Todd Maddox, and Grant C. Baldwin, Department of Psychology, University of Texas at Austin. Research supported by Air Force Office of Scientific Research Award FA9550-06-1- 0204 to Arthur B. Markman and W. Todd Maddox, and National Institutes of Health Grant R01 MH077708 to W. Todd Maddox and Arthur B. Markman. We thank Leland Lockhart for help with data collection and Robert Josephs, Jennifer Beer, Brian Glass, Darrel Worthy, Jonathan Rein, Jeff Laux, Tram Dinh, Francesca Fraga, Charles Judd and anonymous reviewers for useful comments.

Address correspondence to Lisa R. Grimm, Department of Psychology, University of Texas, 1 University Station, A8000, Austin, TX 78712. Send electronic mail to grimmlr@mail.utexas.edu.

Footnotes

[1]Maddox, Baldwin, and Markman (2006) demonstrated that a gains structure with 2 points for a correct response and zero points for an incorrect response produces the same pattern of results as a gains structure with 3 points for a correct response and 1 point for an incorrect response.

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

[3]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.

[4] If we use all four groups of participants in each Experiment in the same discriminant function analysis, the model correctly classified 56.3 % and 45% of the participants in Experiments 2a and 2b, respectively, both significantly greater than chance, $\underline{p} < .05$.

[5]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 contains 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 that 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 that 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.

[6]The suboptimal conjunctive model accounted for 91% and 89% of the total category responses in Experiments 2a and 2b, respectively. For both experiments, the unidimensional rules on length and orientation were rarely used by participants. Based on AIC, the unidimensional length and orientation models best fit the data 5% and 17% of the time, respectively. In contrast, the unidimensional position rule best fit 30% of the data overall 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 2a and over 45% of the data in Experiment 2b. The remaining model discussions will focus on the conjunctive model fits.

Figure Captions

Figure 1. Sample gains and losses screens with point meter and box where problems (Experiment 1) and lines (Experiment 2) appeared. The point meter reflects the scoring used in Experiment 1.

Figure 2. Percent correct for men and women in the gains and losses tasks in Experiment 1

Figure 3. Stimulus space used in Experiments 2a and 2b with a unidimensional rule on position represented. If using this rule, a subject would classify all stimuli above the plane into Category A and all below in Category B. However, this would give only 83% accuracy. Circles represent Category A items and plus-signs represent Category B items.

Figure 4. Stimulus space with correct conjunctive rule on length and orientation dimensions represented.

Figure 5. Proportion correct across blocks for men and women in the gains and losses tasks in Experiments 2a (Panels A and B, respectively) and 2b (Panels C and D, respectively)

Figure 6. Proportion best fit by the correct conjunctive rule for men and women in the gains and losses tasks in Experiments 2a (Panels A and B, respectively) and 2b (Panels C and D, respectively)

Table 1. Schematic representation of regulatory fit and mismatches

	Gains	**Losses**
Positive stereotype (“Promotion”)	Fit	Mismatch
Negative Stereotype (“Prevention”)	Mismatch	Fit

Figure 1

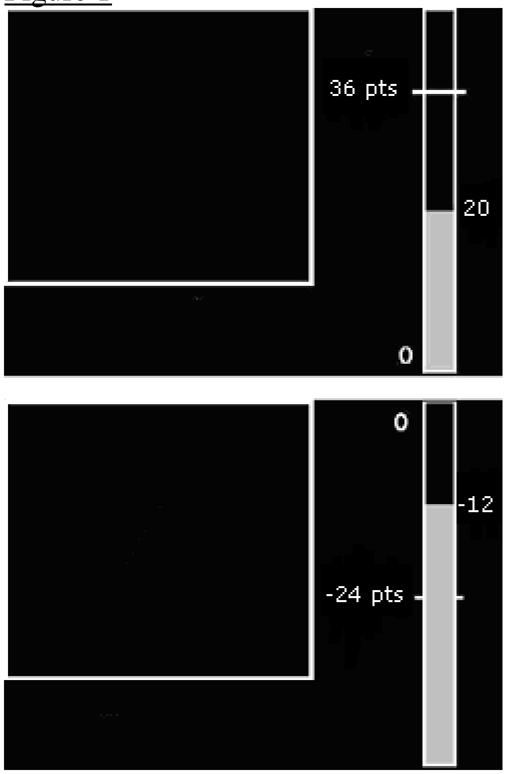


Figure 2

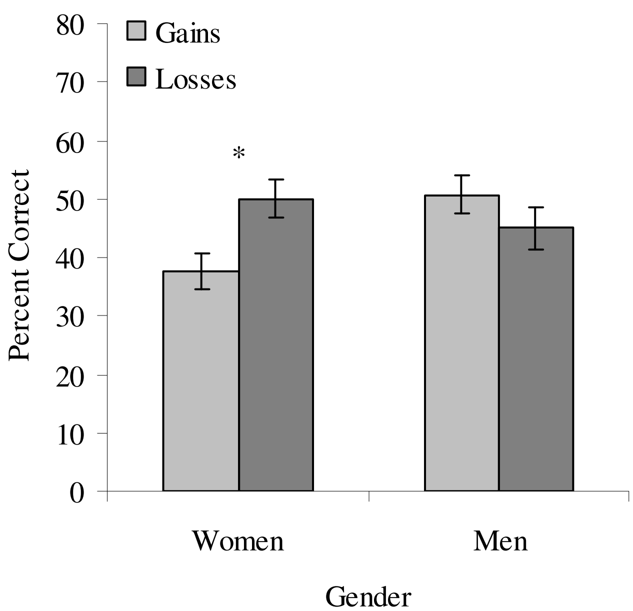


Figure 3

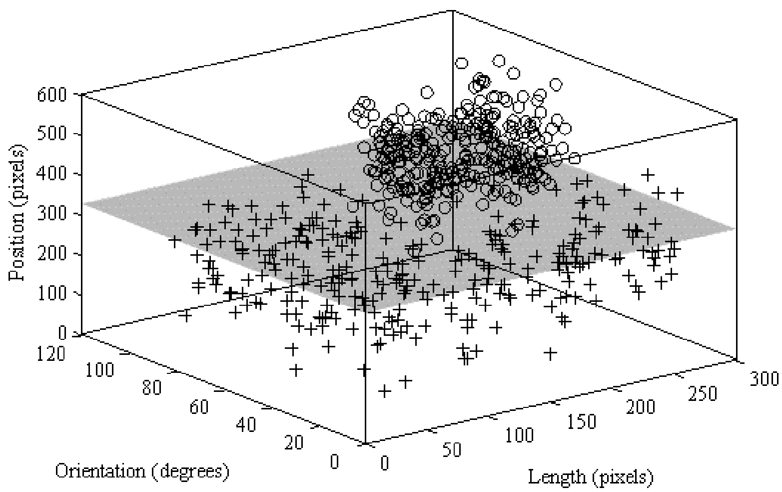


Figure 4

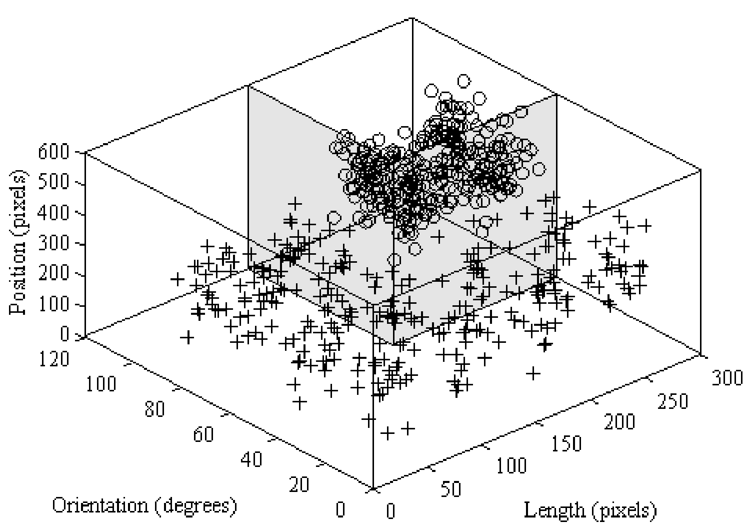


Figure 5

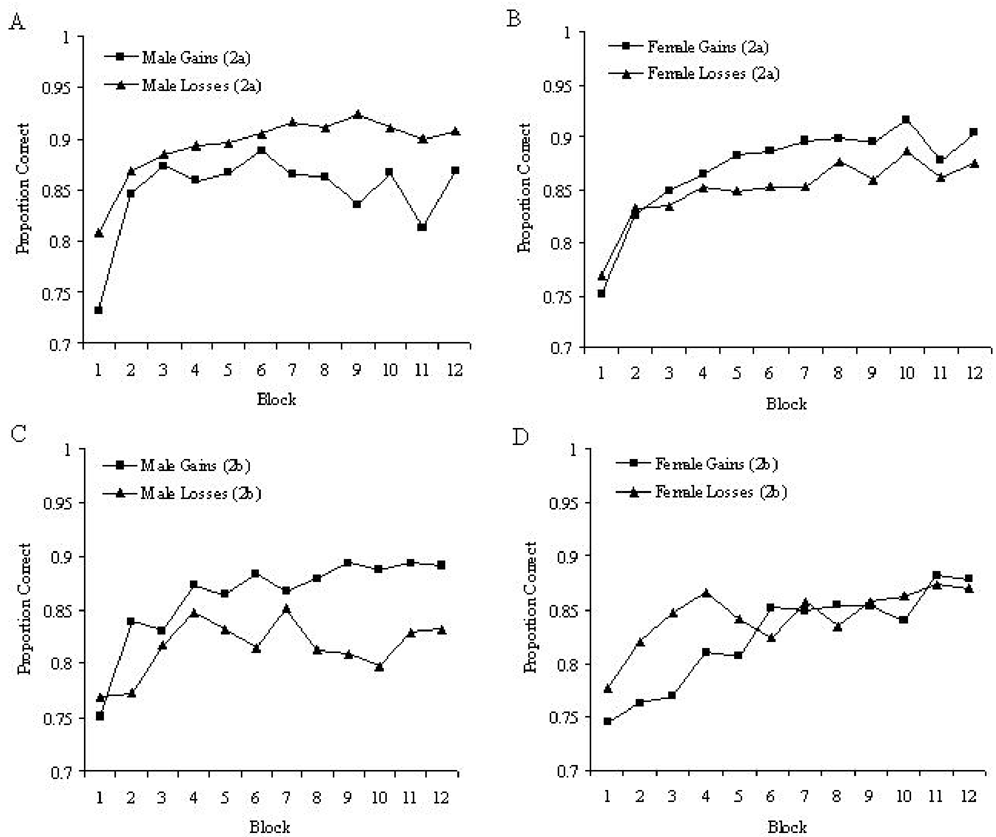
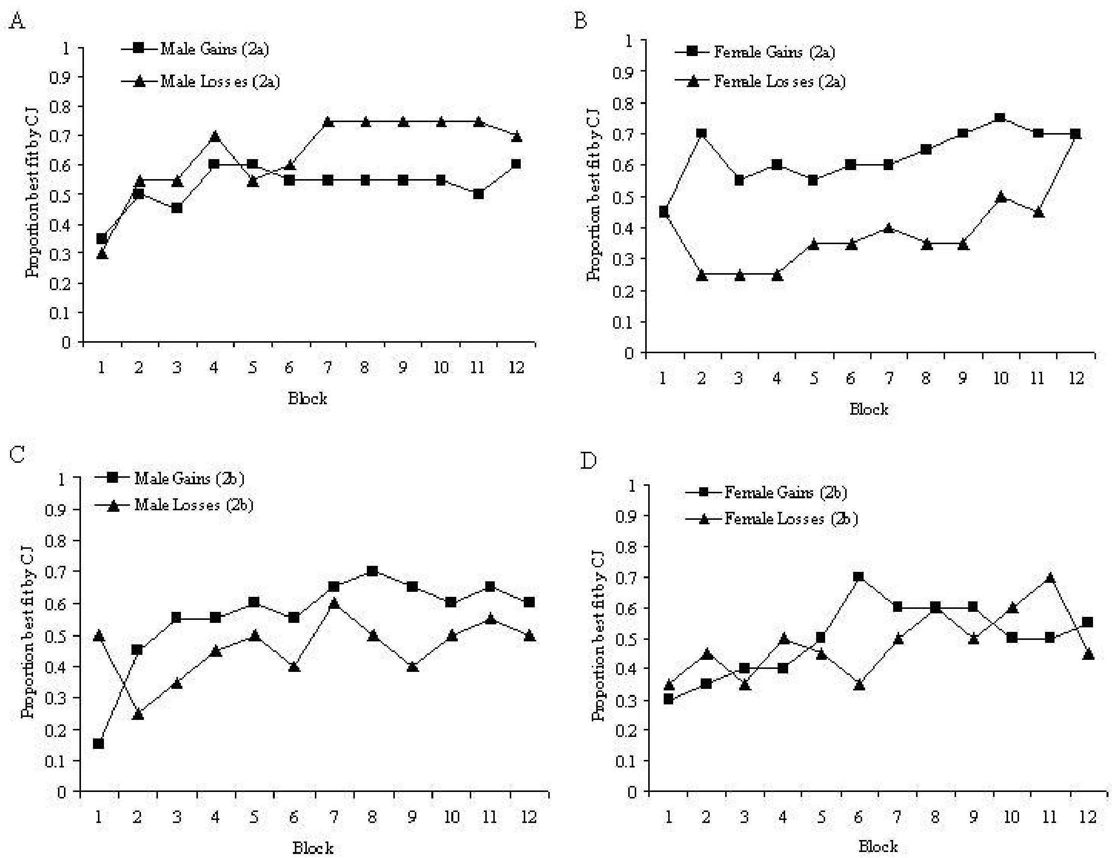

A
Male Gains (2a)
Male Losses (2a)
Proportion Correct
Block
B
Female Gains (2a)
Female Losses (2a)
Proportion Correct
Block
C
Male Gains (2b)
Male Losses (2b)
Proportion Correct
Block
D
Female Gains (2b)
Female Losses (2b)
Proportion Correct
Block

Figure 6


A
Male Gains (2a)
Male Losses (2a)
Proportion best fit by CJ
Block
B
Female Gains (2a)
Female Losses (2a)
Proportion best fit by CJ
Block
C
Male Gains (2b)
Male Losses (2b)
Proportion best fit by CJ
Block
D
Female Gains (2b)
Female Losses (2b)
Proportion best fit by CJ
Block