ORIGINAL ARTICLE

Variations of Gender–math Stereotype Content Affect Women's Vulnerability to Stereotype Threat

**Dustin B. Thoman · Paul H. White ·
Niwako Yamawaki · Hirofumi Koishi**

Abstract To determine whether variations in stereotype content salience moderates stereotype threat effects, 66 US female undergraduate students were given a standardized math exam, and the salience of specific gender–math stereotype content was manipulated before the exam. Women exerted more effort on each problem and performed better on a math exam when threatened with an effort-based stereotype compared to when threatened with the ability-based stereotype or control (where no stereotype was explicitly mentioned). Implications of these results are discussed in terms of stereotype and social identity threat theory, as well as how the socio-cultural salience of ability versus other components of the gender–math stereotype may impact women who pursue math and science-based domains.

Keywords Gender stereotypes · Stereotype threat · Sex · Mathematical ability · Academic motivation

Introduction

The purpose of the present research is to determine whether variations in stereotype content salience moderate stereotype

D. B. Thoman (✉) · P. H. White
University of Utah,
380 South 1530 East, Room 502,
Salt Lake City, UT 84112, USA
e-mail: dustin.thoman@psych.utah.edu

N. Yamawaki
Brigham Young University,
Provo, UT, USA

H. Koishi
Kobe University,
Kobe, Japan

threat effects on women's math performance. Specifically, we ask whether the salience of the gender–math stereotype affects women's math performance (i.e., stereotype threat) when any component of the stereotype is salient or whether these effects occur only when the ability component of the stereotype is salient. We tested this question experimentally by asking undergraduate women to complete a math exam in a typical stereotype threat paradigm, where we made salient the gender–math stereotype (except for those in the control condition) prior to taking exam. In one condition we made salient the ability component of the stereotype (i.e., men outperform women in math because they have more natural mathematic ability) and in another condition we made salient the effort component of the stereotype (i.e., men outperform women in math because they work harder in math). We tested for differences in exam performance, as well as self-reported attitudes and motivation toward the exam, across the two stereotype threat conditions and the control condition to determine whether stereotype threat effects on women's math performance differ as a function of the specific stereotype content that is salient during a math testing situation. The answer to our research question can inform our basic understanding of how stereotype threat functions and how socio-cultural beliefs about stereotype content (in particular, ability-based stereotypes) play a role in this process.

Background

Social psychological research has well documented that knowledge of socio-cultural stereotypes can influence the academic outcomes of stigmatized individuals through a number of indirect and direct paths (Crocker et al. 1998; Swim and Stangor 1998). For example, research based on Eccles's expectancy–value model of achievement (e.g., Eccles 1994; Eccles and Wigfield 2002) has demonstrated

that, indirectly, the cultural transmission of gender role stereotypes can influence individuals' goals and general self-schemata, which in turn influence specific thoughts, feelings, and behavior in direct encounters with stereotype-relevant activities. In terms of the direct influence of stereotypes on stigmatized individuals, research on stereotype and social identity threat (e.g., Steele and Aronson 1995; Steele et al. 2002) suggests that the situational salience of stereotypes can immediately impair stigmatized individuals' performance in and motivation toward diagnostic tests of stereotype-relevant ability, above and beyond the relatively more distal influences of resources and other sociological factors (Steele 1997).

As one might expect, indirect effects of stereotypes may be influenced by a number of variables in the individual's socio-cultural context over time, but research suggests that even when stereotype salience directly affects an individual in a given situation, whether and how the threat is experienced can be influenced by a number of variables in the immediate physical and social environment (see Steele et al. 2002). Examination of this literature, however, suggests that little research has explicitly connected these levels of analysis in terms of how broader socio-cultural variables may affect not only the indirect influences of stereotypes on individuals, but also the direct situational encounters with stereotype or identity threat.

One notable exception is the work of Aronson, Good, and their colleagues (e.g., Aronson et al. 2002; Good et al. 2003), whose intervention research with adolescents informs the link between implicit theories of intelligence (see Dweck and Leggett 1988) and stereotype threat vulnerability. Results from their interventions suggests that students' performance is more influenced by stereotypes about academic ability when they believe that intelligence is a fixed entity, compared to students who view intelligence as malleable. In their interventions, when they teach a group of students that intelligence is malleable (rather than fixed), this group performs better than a control group who did not get the intervention. This work draws attention to the fact that implicit socio-cultural beliefs about the nature of intelligence influence the degree to which students are vulnerable to stereotype threat. The belief that intelligence is based more on effort than ability undermines the effects of an ability stereotype because it disrupts the logical connection between the stereotype (i.e., that one's group has less ability than another group) and the outcome (i.e., that performance is based on one's ability), in this case by suggesting that performance is not based on one's ability but rather on how hard one studies. In the present analysis, we consider whether the link between the stereotype and performance outcome may be disrupted by changing the other side of the socio-cultural beliefs in the equation, i.e., that the *content* of the stereotype *emphasizes* that one's

group has less ability than another group. We recognize that other socio-cultural variables may also influence stereotype threat vulnerability, but our aim here is not to create an exhaustive list of these variables. Rather, the aim of the present research is to explicitly examine a socio-cultural variable that may influence direct effects of stereotype knowledge (or salience) on academic performance: stereotype content.

Specifically, we examine whether variations in the salient content of the gender–math stereotype can predict women's math performance. We focus here on differences in stereotype content because, as Steele et al. (2002) note about direct situational experiences of stereotype threat, "The nature of the threat—the kind of devaluation and mistreatment that is threatened—depends importantly on the specific content of the negative stereotype" (p. 390). The importance of stereotype content can be considered, on one hand, as reflecting the fact that within a culture specific stereotypes apply only to members of certain groups in certain situations (e.g., in the US, women may experience stereotype threat in math, whereas men may experience threat relevant to social sensitivity). On the other hand, we can consider that a given stereotype (e.g., gender–math stereotype) may be construed differently in terms of underlying content across situations and/or socio-cultural contexts (e.g., men outperform women because they have more math ability vs. more men than women have math careers because women are less interested in math). We hypothesize that by considering different socio-cultural beliefs about the underlying content (or construal) of a given stereotype we can predict whether or not individuals are vulnerable to stereotype threat effects in a given situation.

The question we raise here is whether women will be vulnerable to stereotype threat effects on a math exam if we make salient a component of the stereotype other than ability. We can phrase this question in terms of competing hypotheses: (1) does making any part of the gender–math stereotype salient lead to stereotype threat effects on math performance, or (2) do stereotype threat effects on performance only occur when the content of the stereotype made salient relates to ability? We can perform a crucial test of these competing hypotheses using the traditional stereotype threat paradigm and manipulating the specific content of the gender–math stereotype that is made salient.

The answer to our research question would inform our basic understanding of how stereotype threat functions and how socio-cultural beliefs about stereotype content may play a role in this process. There may be a number of socio-cultural beliefs that must be held psychologically for the salience of a stereotype to directly affect a student's academic performance, and understanding how these various beliefs form a foundation for linking stereotypes to performance may improve social psychological theory

on the relationship between socio-cultural beliefs and behavior. Research on stereotype threat to date has focused on ability-based stereotypes, and non-ability-based stereotypes have only been considered in the literature on indirect effects of stereotypes on performance (e.g., Eccles 1994). It is possible that stereotype threat may occur for women whenever the gender–math stereotype is salient (regardless of the specific content of the stereotype that is salient), or it may only occur when the ability content of the gender–math stereotype is salient but not when other content is salient. In addition, with a focus toward social justice and intervention, it may not be possible to convince students in actual classrooms now that the gender–math stereotype does not exist (i.e., eliminate the stereotype), but it may be possible to make salient some other content component of the stereotype in a given situation that could alleviate students' vulnerability to stereotype threat. An analogy here can be drawn to the social psychological research on latitudes of acceptance and rejection of attitudes (e.g., Fazio et al. 1977). That is, convincing students that the gender–math stereotype may be out of reach (in the latitude of rejection) because years of socialization have taught them about this stereotype. However, it may be within the latitude of acceptance to make salient some other component of the gender–math stereotype (even one that is perhaps already a smaller component of the stereotype) if doing so makes stereotype threat less likely to occur than when the dominant ability component of the stereotype is salient. Although we may not be able to eliminate stereotypes in one large step, moving students' beliefs about stereotypes could serve as a relatively smaller step toward changing the stereotype altogether toward a more egalitarian socio-cultural belief.

How Beliefs of Stereotype Content may Impact Individuals' Exam Performance and Motivation

In the United States, the gender–math stereotype at the socio-cultural level most prominently emphasizes that females possess less natural ability than males in mathematics (National Science Foundation 1996). This emphasis on ability is not surprising, considering that the US educational culture supports the normative belief that intelligence is fixed ability (Kurtz-Costez et al. 2005). Given this socio-cultural perspective, with such an emphasis on ability in terms of both stereotype content and educational culture, it makes sense that stereotype threat can directly impair the performance of stigmatized individuals in the US during situations that are most highly diagnostic of ability (Steele 1997); therefore it is understandable why previous stereotype threat research with women in math has rightly focused on the most prominent content of the stereotype.

We assume that when the general stereotype (that men outperform women at math) is salient students are most likely to think of the ability component because it is the most prominent content of the stereotype in this socio-cultural belief structure. This assumption is supported by research suggesting that women are likely to be affected by stereotype threat both when the stereotype is explicitly made salient *and* when nothing is said (because knowledge of the stereotype is implicit in a math testing situation for women), relative to when the stereotype is explicitly nullified (Smith and White 2002). However, because these students know all the various content of the gender–math stereotype it should be plausible to make any given part of the stereotype situationally salient. For our experimental design, therefore, we compare the performance of women who were explicitly threatened with the ability stereotype (i.e., men outperform women at math because they have more natural ability) to a no explicit threat control condition, and finally to a condition where women were explicitly threatened with the gender–math stereotype where the stereotype content other than ability was made explicitly salient.

In addition to a performance component, the gender–math stereotype has a strong social role component related to the idea that women may not be interested in math and women should not choose math careers (Seymour and Hewitt 1997). Stereotypes concerning social gender roles are prevalent in the US, and these stereotypes have been implicated in more long-term and indirect effects (Eccles and Wigfield 2002). Long-term interest and choices of major and careers are variables that necessarily play out over relatively longer periods of time, however, and to provide a strong test of our hypotheses we are looking for a variable related to stereotype content that could logically affect situational performance. We would not expect that presenting social gender role stereotypes to women would affect acute exam performance, but more likely they would affect only longer-term measures such as career interest, value, and choice.

Given the logical reasons not to use the social roles or interest stereotype content in this paradigm, we decided to explore effort as a threat on situational performance. One could think of effort in terms of relatively long-term preparation (e.g., working harder to prepare on exams) or in terms of how hard one works in an actual exam situation to earn a good score. In either case, effort theoretically could be considered as directly relevant to a performance situation. Therefore, we decided to also test whether or not females would demonstrate stereotype threat effects on a math exam when threatened with the stereotype that men outperform women in math because they work harder at math. This manipulation also parallels our lab study more closely with the intervention work by Aronson, Good, and

colleagues (e.g., Aronson et al. 2002; Good et al. 2003, who oppose ability and effort as the two bases for implicit theory of intelligence (see Dweck and Leggett 1988). Our study differs from theirs theoretically, again, in that we focus on shifting the stereotype content rather than shifting students' implicit theories of intelligence.

The Present Research

We reasoned that students in an educational culture that emphasizes the role of ability as well as a stereotype about ability (US students) would be susceptible to stereotype threat effects of impaired performance when presented with the stereotype that males are better than females in mathematics because of innate sex differences in natural mathematics ability, but not when presented the stereotype that males are better at math because they work harder at the subject. Thus, in our experimental design we present the gender–math stereotype to both groups of women, but vary the specific stereotype-related content that is made salient. We also included a control group, where no explicit stereotype was made salient. As mentioned above, we expect that this group may be implicitly threatened by the math situation (see Smith and White 2002) and therefore should perform similarly to those threatened with the ability stereotype. Compared to both this control group and the threat–ability group, we expect women in the threat–effort group to perform better on the math exam even though the gender–math stereotype is made salient (because content other than ability will be salient).

We employ two dependent measures of exam performance to gain a more complete description of how variations in stereotype content salience affect exam performance. First, we will measure the percent score on the exam, because a percent score more closely parallels the actual scoring of standardized exams than the number of items correct (cf. Spencer et al. 1999). Second, we will measure number of problems attempted as a proxy measure of effort on exam (cf. Steele and Aronson 1995). We use these two performance variables in conjunction to not only look at exam score, but to determine whether our stereotype content manipulation affects both the percent correct and how much effort participants exert on each question. Following the logic for hypotheses discussed above, we expect that those threatened with the effort component of the stereotype will get a higher percent of items correct and will exert more effort on each item than those in the threat ability and control conditions.

In addition to performance outcome measures, we included a number of process measures in an attempt to better understand how differences in performance may be accounted for by effects on attitudes, motivation, and/or feelings related to the exam. Though various process

measures have received mixed results in previous studies (see Smith 2004; Steele et al. 2002 for review) we include several process measures in this study because it is not clear which variables will explain how manipulating stereotype content may affect exam performance. We also include a measure of domain identification (Smith and White 2001) as a covariate, in order to isolate effects of our situational manipulation on performance from the portion of performance that is determined by pre-existing differences in math ability and identity.

To summarize the each of the specific hypotheses being tested, we list them below:

1. For all dependent variables, there should be no differences between those in the control group and those in the threat ability group. All predicted effects should emerge as a function of comparing the threat effort group to threat ability and control groups.
2. Those presented with effort component of the gender–math stereotype will perform better than those presented with ability stereotype and those in control group.
3. Those presented with effort component of the gender–math stereotype will attempt fewer problems than those presented with ability stereotype and those in control group.
4. Differences between the threat effort and other groups in performance should be explained by differences in reported thoughts, feelings, and motivation. We do not have specific a priori predictions for which specific variables will account for these effects because although there is much research on mediators of stereotype threat phenomenon (see Smith 2004; Steele et al. 2002) there is no work that has looked at process variables as a function of a stereotype component manipulation.

Method

Participants

Sixty-six female undergraduate students (mean age=21.23, SD=4.20) from a large private university in the Rocky Mountain region participated for class extra credit. They participated in sessions that ranged in size from 3 to 10 and were randomly assigned to one of three experimental conditions: (1) threat ability (N=19), (2) threat effort (N=22), and (3) control (N=25). In addition, participants' domain identification with mathematics was measured and included as a covariate, as described shortly.

Procedure

Participants arrived at the lab expecting to participate in the validation of a "newly developed mathematics exam." They

were told that only female participants were currently recruited because enough data had already been collected on male participants, and we needed to equal number of males and females for proper validation. Furthermore, participants were told that the exam was ostensibly being validated nationwide. The exam was described as the Polo Mathematics Aptitude Assessment (PMAA), with no other context provided.

Following this study introduction and informed consent, participants were told that we wanted to understand their views and evaluations of the exam, thus we would give them a packet of questionnaires to complete before and after taking the exam, but that people may receive different questions, so some people may finish faster than others. For those who do, they were to wait before starting the exam itself. This allowed us to use written instructions and manipulations, so individuals within a single session were not necessarily in the same condition.

Next, all participants were told we wanted to gather some background information regarding their experience and attitudes toward mathematics and other areas. In the background questionnaire, participants answered questions regarding how many math classes they had taken, how recently, etc. Embedded in the questionnaire was the Domain Identification Measure (DIM) by Smith and White (2001), which is described below.

After completion of the background questionnaire, participants randomly assigned to Threat Ability or Threat Effort conditions were given a recent journal article on mathematic assessment to read as a way to help them, "focus and get in the right mindset (thinking about math) before taking the exam." Depicting the Theory of Intelligence manipulation, participants read one of the two articles. Both articles were modeled articles used by Smith and White (2002). The articles, written with a scientific tone, described fictitious research and concluded that males are better than females at mathematics, but differed in the source of the discrepancy. The Ability article stated that males are better than females at mathematics because of innate, biological, and genetic differences in thinking and reasoning patterns associated with mathematics, and that any differences in effort had been ruled out as a possible cause of sex differences. The Effort article, on the other hand, suggested that research has ruled out innate differences between the sexes, and that considerable research shows that males are better than females at math simply because they try harder and exert more effort than females in math. After reading the article, these participants also read, "…the article suggests that men outperform women in math because of differences in (innate ability or effort). Indeed, we have found similar results in our study thus far, that men tend to outperform women on this exam." This was done to reinforce the potential threat of the stereotype

(e.g., Spencer et al. 1999). Participants assigned to the control conditions were not given an article to read. After completion of the background questionnaire, they completed the pre-exam questionnaire (described below).

Next, all participants read a short description of the exam and completed two sample exam items. Afterward, they completed the pre-exam questionnaire (described in more detail below) designed to measure motivational attitudes, expectations, and feelings related to of the upcoming exam. When all participants finished the questionnaire, the experimenter began the exam. The exam consisted of 20 moderately difficult to difficult items each with four possible answers. The exam was modeled after the Graduate Record Exam (GRE) and included items from the general and advanced GRE quantitative sections (Educational Testing Service 1995). The exam was divided into two 10-item sections and participants were allowed 10 min to complete each section. After the exam, participants completed a post-exam questionnaire (described in more detail below) designed to measure perceptions of the exam, feelings during the exam, and perceived performance. Finally, participants completed demographic items, including sex, age, and ethnicity. Participants were then fully debriefed and thanked for their participation.

Measure

Exam Performance

Two indices were selected a priori to assess performance on the exam. First, we looked at the ratio of correct problems to the number of problems attempted because not all participants were exposed to all exam items before time expired. As done in other research (e.g., Spencer et al. 1999; Steele and Aronson 1995), we divided the total number of correctly solved problems by the total number attempted to determine the percentage of correct items for each participant. The second a priori index to assess performance was the total number of problems completed by the participants, which is considered as a proxy of how much effort participants put forth on the exam (cf., Steele and Aronson 1995).

Domain Identification

Embedded within the filler items of the initial background questionnaire was Smith and White's (2001) Domain Identification Measure (DIM), which is designed to measure how self-identified a person is with particular domains. We used the nine- item scale (Cronbach's alpha= .87) to assess participants' identification with mathematics. Sample items include, "How much is math related to the sense

of who you are?" and "How important is it to you to be good at math?" where participants responded using a scale ranging from 1 (not at all) to 5 (very much).

Pre-exam Measures

The pre-exam questionnaire consisted of seven Likert-scale items asking participants to rate their commitment ("To what extent do you feel committed to doing your best on this exam?"), motivation ("To what extent do you feel motivated to perform well on this exam?"), effort ("How much effort will you put forth do perform well on this exam?"), anxiety ("To what extent do you feel anxious about taking this exam?"), and perceived competence ("To what extent are you certain that you have the academic knowledge to do well on this exam?"), as well as their anticipated difficulty of the exam ("How difficult do you anticipate this exam to be?") and expected performance ("How well do you expect to do on this exam?"). Participants rated each single-item measure using a scale that ranged from 1 to 11, where higher numbers indicated higher levels of the construct.

Post-exam Measures

This final questionnaire consisted of six Likert-scale items asking participants to rate the degree to which they agreed or disagreed with several items related to their perceptions of the exam, feelings while working on the exam, and perceived performance. Specifically, participants rated on 1 (strongly disagree) to 5 (strongly agree) scale their perceptions of the exam as difficult ("This test was very difficult") and fair ("Math tests like this are unfair") (reverse coded), their perceived performance (e.g., "What percent do you think you got correct on this exam?"), perceptions of the exam's validity as a test of math ability ("This was a good test of my math ability"), feelings of rushing through the exam ("I felt rushed to finish each section of the exam"), and general ability to perform on math exams ("I usually do well on math exams").

Results

Descriptive

Table 1 displays the means and standard deviations for each condition for all the dependent variables. Descriptive statistics are provided for the two measures of performance (percent correct and number of problems attempted), as well as for all pre-exam and post-exam measures of attitudes, motivation, and feelings toward the exam. Tables 2 and 3, respectively, display the correlations between the performance measures and pre- and post exam self-report measures for the entire sample.

We used ANOVA models to test for demographic differences in age and ethnicity across the conditions. Results suggest that there were no significant demographic differences by age or ethnicity across the conditions ($ps > .10$).

Table 1 Means and standard deviations for all dependent variables by condition.

	Control ($N=25$)	Threat–ability ($N=19$)	Threat–effort ($N=22$)
Performance			
Percent correct	58.99 (17.5)	62.10 (21.51)	68.19 (18.62)
Number attempted (of 20 items)	17.48 (1.82)	17.78 (1.31)	16.18 (2.36)
Pre-exam self-report (on 1–11 scale)			
Commitment	7.64 (1.84)	8.52 (1.98)	8.13 (1.95)
Motivation	7.28 (2.09)	8.42 (2.54)	7.63 (2.10)
Effort	8.20 (1.95)	8.79 (1.81)	8.41 (1.84)
Anxiety	4.60 (3.47)	4.68 (2.51)	5.36 (3.52)
Perceived competence	5.28 (3.02)	6.42 (1.86)	5.18 (2.65)
Anticipated difficulty	7.76 (1.76)	8.21 (1.84)	7.91 (1.54)
Expected performance	5.40 (2.39)	7.00 (1.82)	5.64 (2.13)
Post-exam self-report (on 1–5 scale)			
Difficulty	3.36 (.91)	2.79 (.98)	3.27 (.98)
Fairness	3.34 (1.28)	3.69 (1.10)	3.41 (.96)
Perceived performance	2.32 (1.07)	2.68 (1.00)	2.45 (.86)
Validity of exam	2.52 (1.08)	2.78 (.85)	2.63 (1.13)
Feelings of rushing through exam	3.56 (1.26)	3.52 (1.54)	3.91 (1.23)
General math exam ability	3.28 (1.06)	3.79 (1.03)	3.45 (.67)

For both pre-and post-exam scales, higher numbers indicate more of the construct.

Table 2 Correlations between exam performance indices and pre-exam process measures.

Variable	2	3	4	5	6	7	8	9
Percent correct	−.40*	.32*	.31*	.32*	−.17	.22	−.48*	.42*
Number attempted	–	−.12	−.13	.03	−.19	.24*	−.08	.18
Commitment		–	.71*	.77*	.04	.17	−.09	.37*
Motivation			–	.78*	.22	.27*	−.01	.43*
Effort				–	.20	.17	−.04	.34*
Anxiety					–	−.29*	.51*	−.30*
Perceived competence						–	−.41*	.80*
Anticipated difficulty							–	−.44*
Expected performance								–

*p<.05; N=66 for all correlations.

Testing Hypotheses

We planned to analyze the data using a linear regression model of Analysis of Covariance (ANCOVA) with participants' math identification scores as our covariate (e.g., Smith and White 2002) and a set of two planned orthogonal contrasts to test our specific hypotheses of mean differences between the three groups. First, a total domain identification score was calculated by summing together the items on the DIM (Smith and White 2001). Next, an analysis was conducted to verify that participants' DIM scores did not differ by condition, which was the case, $F(2, 63)=2.19$, $p>.10$. Thus, the data were analyzed using the planned regression model with DIM scores as the covariate.

To build the regression model, we next set up the two orthogonal contrast codes. The first contrast was designed to test Hypothesis 1: our assumption that the threat–ability and no threat control groups did not differ on the dependent variables, as found by Smith and White (2002). In this contrast, we coded the threat–ability group as −1 and the control group as +1 (and the threat–effort group was coded as 0). The second contrast was designed to test Hypothesis 2 and Hypothesis 3: that women threatened with the effort content of the gender–math stereotype would outperform

(Hypothesis 2) and exert more effort than (Hypothesis 3) those threatened with the ability content of the stereotype and the no threat control group. In this contrast, we coded the threat–effort group as +2 and coded both the threat ability and control group as −1. We entered this set of contrasts and the DIM covariate into the model simultaneously. The ratings and scores presented are estimated means, adjusted for DIM scores.

Performance on Exam

Ratio of Correct Problems

When the ratio of correct problems was regressed onto the set of orthogonal contrasts and the DIM, the overall model was significant, $F(3, 62)=5.46$, $p=.002$, $R^2=.21$. As predicted, the first contrast (between the control and threat–ability group) was not individually significant, $B= −.63$, $t(62)=−.23$, $p=.82$. Confirming Hypothesis 1 and replicating the results of Smith and White (2002), female participants explicitly threatened with the stereotype that men outperform women at math due to natural ability performed at the same level on the math exam as participants who were not explicitly threatened during the

Table 3 Correlations between exam performance indices and post-exam process measures.

| Variable	2	3	4	5	6	7	8
Percent correct	−.40*	−.25*	.09	.41*	.08	.07	.37*
Number attempted	–	−.33*	.13	.55*	.09	−.05	.48*
Difficulty		–	−.17	−.62*	.07	.16	−.09
Fairness			–	.33*	.29*	−.15	−.14
Perceived performance				–	.12	−.23	.38*
Validity of exam					–	.02	.03
Feelings of rushing						–	.09
General math exam ability							–

*p<.05, N=66 for all correlations

experiment. Next, as predicted by Hypothesis 2, the second contrast did significantly predict the ratio of correct problems, $B=3.39$, $t(62)=2.19$, $p=.03$. As can be seen in Fig. 1, participants threatened with the effort stereotype solved a higher ratio of problems than those in the ability stereotype and control conditions.
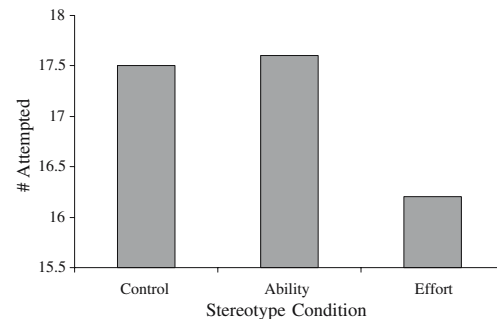
Problems Attempted

The second a priori index to assess performance was the total number of problems completed by the participants as a possible proxy of how much effort participants put forth on the exam (cf., Steele and Aronson 1995). When the number of problems attempted was regressed onto the set of orthogonal contrasts and the DIM, the overall model was significant, $F(3, 62)=3.61$, $p=.01$, $R^2=.15$. As predicted by Hypothesis 1, the first contrast (between the control and threat–ability group) was not significant, $B=-.006$, $t(62)=.20$, $p=.84$. Next, also as predicted, the second contrast did significantly predict the number of problems attempted, $B=-.49$, $t(62)=-2.68$, $p=.01$. As can be seen in Fig. 2, confirming Hypothesis 3, participants threatened with the effort stereotype attempted fewer problems than those in the ability stereotype and control conditions.

Number Correct

We did not plan a priori analyses on the number of items answered correctly because percent score more closely parallels the scoring of most standardized exams than the number of correct items (cf., Steele and Aronson 1995; Spencer et al. 1999). However, we conducted a post-hoc analysis with number of items correct as the dependent variable to clarify the pattern of results across our planned measures of performance. Specifically, one possible interpretation of results thus far would be that participants in the threat–effort condition adopt a different test taking strategy— that they go more slowly and so they attempted fewer problems, but they get more of the problems that they
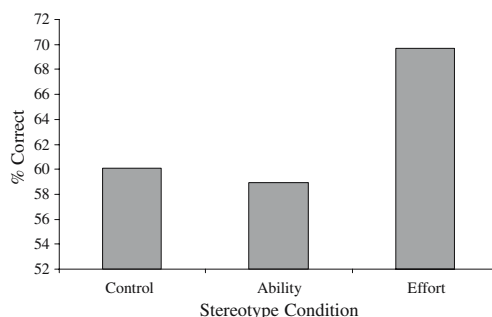
Fig. 1 Stereotype condition effects on the percent of problems correct. Scores represent estimated means by condition after adjusting for domain identification covariate.

Fig. 2 Stereotype condition effects on the number of problems attempted. Scores represent estimated means by condition after adjusting for domain identification covariate. The number of items attempted on the math exam ranges from 0 to 20.

attempt correct. From this interpretation, it is possible that although those in the threat–effort condition answer a higher percentage correct than all others they might actually answer fewer total items correct (thus perhaps offsetting the benefits of getting a higher percentage correct). To test this possibility we regressed the number of items correct on the same regression model used above. Although the overall model predicting number of correct was significant ($F(3,62)=8.29$, $p<.001$, $R^2=.29$), the only individually significant term in the model was the domain identification covariate ($p<.01$). Neither of the contrast terms significantly predicted the number of items correct ($ps>.10$). The pattern of estimated means (adjusted for the domain identification covariate) suggests that those in the threat–effort condition ($M=11.44$, $SE=.63$) did not answer fewer items correct than those in the threat–ability ($M=10.63$, $SE=.65$) or control ($M=1.92$, $SE=.59$) groups. Thus, though not statistically significant from other groups, those in the threat–effort condition actually answered the most items correct among the experimental groups.

Examining Process Measures

Hypothesis 4 predicts that performance differences across the experimental conditions would be explained by process measures related to perceptions, feelings, and/or motivation associated with the exam. However, we did not make a priori predictions for which specific variables would account for performance differences because many process variables have been found to be important in stereotype threat research (Smith 2004; Steele et al. 2002), but there is no clear basis in the literature for predicting which process variables would explain stereotype threat differences as a function of manipulating the stereotype content. Our analysis strategy to test this hypothesis reflects the lack of specificity in predicting process compared to our hypotheses related to exam performance. We used ANOVA analyses to test for differences in responses to all items on

the pre-exam and post-exam questionnaires, and planned to use follow-up pairwise comparisons with a Bonferroni adjustment to determine which conditions differed in cases where the omnibus F test is significant.

Pre-exam Responses

Before completing the exam, participants responded to a series of questions (1–11 scale) to assess their mindset just before taking the exam to see if these factors help explain any effects of the experimental conditions on math performance (e.g., Steele and Aronson 1995). All means by condition are displayed in Table 1. No statistically significant differences were found across conditions for participants' pre-exam commitment to do well on the exam, motivation, anxiety, perceived competence, anticipated difficulty of exam, effort, and expectations to perform well on the exam (all $ps > .25$). Such factors have shown mixed results in their influence of participants' performance in stereotype threat research (e.g., Aronson et al. 1999; Smith and White 2002; Spencer et al. 1999).

Post-exam Responses

After completion of the exam, participants responded to a series of questions (1–5 scale) to assess their mindset right after taking the exam. Means and standard deviations for all items are displayed by condition in Table 2. Surprisingly, no statistically significant differences were found for participants' evaluations of the exam's difficulty, fairness, validity as a test of math ability, feelings of rushing through the exam, or ability to perform on math exams (all $ps > .25$).

Correlations of Process Measures and Performance

In an attempt to better understand the process measure outcomes, we looked post-hoc at the simple correlations between them and the ratio of correct problems solved for each experimental condition. Only the significant correlations are reported. For participants in control condition, better performance was associated with lower pre-exam anxiety ($r = -.40, p < .05$), anticipated difficulty of exam ($r = -.83, p < .01$), and post-exam fairness of math tests ($r = -.42, p < .05$). Better performance was associated with higher pre-exam perceived competence ($r = .43, p < .05$) and expectation to perform well ($r = .59, p < .05$). For participants threatened by an ability stereotype, better performance was associated with higher pre-exam motivation to do well on the exam ($r = .46, p < .05$), expectation to perform well ($r = .60, p < .01$), and post-exam ratings of ability to perform on a math exam ($r = .61, p < .05$). No significant correlations were found for participants threatened by an effort stereotype.

Discussion

Results of our experimental study demonstrate that participants exposed to the effort stereotype completed fewer problems but correctly solved a higher percentage of problems than those exposed to the ability stereotype or not explicitly exposed to the stereotype at all. This pattern suggests that those in the threat effort condition exerted more effort on each problem, and therefore took more time to complete each problem but did so at a more successful rate than participants in other conditions. Although they took more time on each problem than those in other groups, this strategy did not impair overall performance, as those in the threat effort condition answered the most problems correct (though not statistically different). In addition, the similarity between the control group and those threatened with the ability stereotype is congruent with previous findings that suggest US females may typically be aware of the negative ability stereotypes they face in difficult math exams even when no one explicitly mentions it (Smith and White 2002). Thus, for students who are members of a culture that stresses the importance of ability in education and stigmatizes their sex for lack of ability in math, negative stereotype threat effects on performance seem to have been relatively alleviated when the stereotype content was reframed in terms of effort rather than ability.

The finding that females' performance on a math exam is better when the gender–math stereotype is framed in terms of effort, rather than ability, provides lab-based experimental support for Aronson and colleagues' intervention work that focus on changing students' implicit theories of intelligence. Rather than reframing the stereotype content, Aronson and colleagues (Aronson et al. 2002; Good et al. 2003) teach students that intelligence is based more on effort than natural ability. This strategy of reshaping students' implicit theories of intelligence has proven effective, as students who were taught a more malleable view of intelligence performed better on standardized exams than students who had not received such instructions (and who held a relatively more fixed-entity belief) (Aronson et al. 2002). Indeed, changing views of individuals students' theories of intelligence may be easier (and thus more effective for interventions) than changing socio-cultural views of stereotypes, but theoretically these two findings can be merged to shed light on processes underlying stereotype threat and the person-culture fit that must exist for individuals to be vulnerable to these effects. As demonstrated both by our threat–effort manipulation in the lab and Aronson et al.'s (2002) interventions, stereotype threat effects are alleviated when a mismatch occurs between what is stigmatized in terms of stereotype content and how diagnostic the exam situation is of this stigma. A

mismatch from either variable, then, reduces threat effects on performance in the immediate situation.

While no differences were found on the pre- or post-exam subjective measures of motivation and evaluation of the exam, we did find some significant simple correlations with the ratio of correct problems. Not surprisingly, for control participants, feelings of lower anxiety and higher knowledge were associated with better performance. For control and ability participants, the more they expected to do well, the better they performed, which did not occur for participants in the effort condition. It may be that those in the former two conditions felt more of a threat from the stereotype compared to the latter condition, and thus, believing that they expect to perform well had a stronger relationship to their actual performance.

Implications

The current results not only inform theory for how socio-cultural beliefs can directly affect performance, but they also have implications for prevention of stereotype effects in the US As mentioned earlier, many stereotype threat studies suggest that negative stereotype threat effects on performance can be alleviated with the stereotype is situationally nullified, e.g., when exam proctors state that the exam is not diagnostic of ability (e.g., Steele and Aronson 1995) or that there have been no gender differences on this exam (Spencer et al. 1999). However, as we mentioned above, because gender stereotypes are so prevalent in the socialization of students (see Deaux and LaFrance 1998), students may not believe exam proctors who attempt to nullify the threat on an important actual exam (e.g., ACT, SAT, or GRE math exams). If educators cannot nullify the threat directly, it may be more likely that they could make salient some other, less harmful, component of the stereotype, such as effort (see Steele et al. 2002 for a review of other methods shown to remedy the detrimental effects of stereotype and social identity threats). This less direct method of nullifying the stereotype threat effects may allow educators to shift the emphasis of harmful gender stereotypes away from ability, and on other factors such as effort or interest, that students may feel more power to change. This strategy is similar to those employed by Aronson, Good, and colleagues of shifting the pressure away from the ability stereotype by changing students' beliefs about the nature of intelligence, and perhaps coupling this strategy of shifting beliefs about the nature of intelligence with shifting beliefs about the nature of the stereotype could be even more effective. Again, although the ultimate goal, from a social justice perspective, would be to completely alleviate these negative stereotypes it may be more plausible to take smaller steps toward this goal with students than to achieve the goal in one large leap.

Future Directions

In both the setup and discussion of this study we have emphasized that differences in socio-cultural beliefs about the gender–math stereotype content can influence whether the salience of the stereotype negatively affects performance on a math exam. We manipulated the stereotype content in the lab study presented here, but recently we also conducted a pilot study in Japan, where the education system more strongly emphasizes effort over ability in education (Chen and Stevenson 1995; Hess and Azuma 1991). We assumed that these differences in socio-cultural beliefs about education might translate to differences in the underlying structure of the gender–math stereotype (i.e., the US gender–math stereotype may emphasize ability whereas the Japanese stereotype may emphasize other components). Using the same lab design as the study presented above, we found no significant differences in math exam performance as a function of whether Japanese females were made aware of the ability stereotype, the effort stereotype, or when no stereotype was made salient. In fact, a non-significant trend in these data suggest that women who were made aware of either the ability or effort stereotype tended to outperform those in the no stereotype control condition. From these pilot data it is left unclear whether (and how) stereotype threat effects on performance impact Japanese women or not, and future work is necessary to understand how the salience of stereotypes may affect individuals from places where the socio-cultural belief structures of education and of the gender–math stereotype may influence women.

Another important future direction for this research is to examine dependent variables other than performance, which are likely to reflect more enduring values and choices related to academics and careers. Recent research has begun to shift the sole focus in the stereotype and identity threat literature away from performance to include other academic issues such as one's sense of belonging (Good et al. 2005; Inzlicht and Good 2006; Murphy et al. 2007), feelings of interest and motivation for the activity (see also Sansone and Thoman 2005), and choosing leadership roles (Davies et al. 2005). Coupling this new research with ongoing efforts to understand indirect effects of stereotypes on choice and motivation (e.g., Eccles 1994) may eventually help us understand how stereotypes affect multiple variables that are important to academic and career success, in addition to performance, and how socio-cultural variables (such as the situational or cultural salience of different stereotype content) may influence whether and how stereotypes directly and indirectly affect students.

Conclusion

To truly understand the social psychological theory of stereotype threat, and how to use this theory toward social progress, we must understand why some individuals are vulnerable to stereotype threat and others are not. Our data suggest one possibility for explaining differences in stereotype threat vulnerability for women in mathematics exam situations: the socio-cultural beliefs underlying the gender–math stereotype. Others have identified differences in stereotype threat vulnerability for African Americans at historically Black colleges and universities (Sloan et al. 2004). In either case, the explanation for these differences cannot focus solely on the immediate social context or intra-individual psychological processes. The understanding of stereotype threat, therefore, necessarily emphasizes the interdependence of multiple levels of analysis, and not to be forgotten among these levels of analysis is the socio-cultural structure of beliefs in which individuals are psychologically embedded.

References

Aronson, J., Fried, C., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38,* 113–125.

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

Chen, C., & Stevenson, H. W. (1995). Motivation and mathematics achievement: A comparative study of Asian-American, Caucasian-American, and East Asian high school students. *Child Development, 66,* 1215–1234.

Crocker, J., Major, B., & Steele, C. (1998). Social stigma. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 504–533, 4th ed.). New York: McGraw-Hill.

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotypes threat on women's leadership aspirations. *Journal of Personality and Social Psychology, 88,* 276–287.

Deaux, K., & LaFrance, M. (1998). Gender. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 504–533, 4th ed.). New York: McGraw-Hill.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95,* 256–273.

Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of Women Quarterly, 18,* 585–609.

Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53,* 109–132.

Educational Testing Service (1995). *Practicing to take the general test: Big book.* Princeton, NJ: Educational Testing Service.

Fazio, R. H., Zanna, M. P., & Cooper, J. (1977). Dissonance and self perception: An integrated view of each theory's proper domain of application. *Journal of Experimental Social Psychology, 13,* 464–479.

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology, 24,* 645–662.

Good, C., Dweck, C. S., & Rattan, A. (2005). *Perceiving a malleable-ability versus a fixed-ability environment: The effect of women's sense of belonging to math.* Poster presented at the annual meeting of the Society for Personality and Social Psychology. New Orleans, LA, January.

Hess, R. D., & Azuma, H. (1991). Cultural support for schooling: Contrasts between Japan and the United States. *Educational Researcher, 20,* 2–8.

Inzlicht, M., & Good, C. (2006). How environments can threaten academic performance, self-knowledge, and sense of belonging. In S. Levin & C. van Laar (Eds.), *Stigma and group inequality: Social psychological perspectives* (pp. 129–150). Mahwah, NJ: Lawrence Erlbaum.

Kurtz-Costez, B., McCall, R. J., Kinlaw, C. R., Wiesen, C. A., & Joyner, H. (2005). What does it mean to be smart? The development about children's beliefs about intelligence in Germany and the United States. *Journal of Applied Developmental Psychology, 26,* 217–233.

Murphy, M., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science, 18,* 879–885.

National Science Foundation (1996). *Women, minorities, and persons with disabilities in science and engineering: 1996* (National Science Foundation Publication No. 96–311). Arlington, VA: Author.

Sansone, C., & Thoman, D. B. (2005). Interest as the missing motivator in self-regulation. *European Psychologist, 10,* 175–186.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences.* Boulder, CO: Westview Press.

Sloan, L. R., Wilburn, G., Fenton, T., Craig-Henderson, K., & Starr, B. J. (2004). *Out-group presence, but not overt negative stereotype reminders, produce stereotype threat performance decrements in in-group settings.* Paper presented at the Annual Meeting of the Midwestern Psychological Association, Chicago, IL, April.

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

Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement, 61,* 1040–1057.

Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles, 47,* 179–191.

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

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

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

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype threat and social identity threat. *Advances in Experimental Social Psychology, 34,* 379–440.

Swim, J. K., & Stangor, C. (1998). *Prejudice: The target's perspective.* San Diego, CA: Academic.