Sex Roles, Vol. 54, Nos. 1/2, January 2006 (© 2006)
DOI: 10.1007/s11199-005-8873-2

Discounting the Difficult: How High Math-Identified Women Respond to Stereotype Threat

Alexandra C. Lesko[1] and Jennifer Henderlong Corpus[1,2]

In this study, we examined how math identity moderates women's response to gender-related stereotypes in the domain of mathematics. Male and female college students with varying degrees of math identification took a challenging math test with a gender-related stereotype either activated (i.e., stereotype threat) or nullified. Consistent with previous research, women performed worse than men in the stereotype threat condition, but equal to men in the stereotype nullification condition when performance was adjusted for math SAT scores. Moreover, when faced with stereotype threat, high math-identified women discounted the validity of the test more than did less math-identified women or men in general. We discuss potential benefits and drawbacks of a discounting strategy for women who are highly identified with math.

KEY WORDS: stereotype threat; gender; identity; mathematics.

The underperformance of women and girls on standardized tests of mathematical aptitude has been well documented (American Association of University Women, 1995; Gonzalez et al., 2004). Women perform worse than men by approximately half of a standard deviation on the math section of the Scholastic Aptitude Test (SAT; Wilder & Powell, 1989; see also Coley, 2001), and approximately two-thirds of a standard deviation on the quantitative section of the Graduate Record Exam (GRE; Educational Testing Service, 2002). Cultural stereotypes prevail even beyond these documented gender differences in math performance. The average American is not only familiar with the stereotype of women as inferior mathematicians, but also believes that the stereotype has some truth to it (Eccles, Jacobs, & Harold, 1990; Swim, 1994; see also Devine, 1989).

Gender differences in both genetically endowed capabilities (Benbow & Stanley, 1980) and socialization experiences (e.g., Eccles, Barber, Jozefowicz, Malenchuk, & Vida, 1999) have been proposed to account for these discrepancies in math performance. For example, researchers have argued that gender differences exist with respect to spatial abilities (Fennema & Sherman, 1977), approaches to solving math problems (Harris & Carlton, 1993), treatment in the classroom (Eccles et al., 1999), perceptions of math ability (Meece, Parsons, Kaczala, Goff, & Futterman, 1982), and interest in math (Eccles et al., 1999). Although each of these explanations provides possible insight into test score differences, none of them conclusively explains why women—including those who have high math GPAs and/or who are highly identified with math and/or who want to pursue math as a career—consistently perform lower than men on standardized tests. Further, none of these explanations account for the replicable finding that women perform equally as well as men on challenging tests of mathematical aptitude under certain conditions. For this, we must turn to another theoretical account: *stereotype threat*.

Stereotype Threat

Stereotype threat is a situational pressure that stigmatized individuals experience when they are in

[1]Department of Psychology, Reed College, Portland, Oregon.
[2]To whom correspondence should be addressed at Department of Psychology, Reed College, 3203 SE Woodstock Blvd., Portland, Oregon 97202; e-mail: jennifer.henderlong.corpus@reed.edu.

jeopardy of confirming a negative stereotype about themselves and their in-group (Steele, 1997). In evaluative situations in the mathematical domain, for example, women must contend not only with the stressful nature of the evaluation per se, but also with the threatening possibility of confirming the cultural belief that they are inherently less competent than men are at mathematics. Because of this additional pressure, the performance of stigmatized individuals suffers when they are reminded of their group membership (e.g., female) or of the relevant stereotype (e.g., women are inferior mathematicians), but not when the stereotype is removed or nullified (Keller & Dauenheimer, 2003; Quinn & Spencer, 2001; Spencer, Steele, & Quinn, 1999; Steele, 1997; Steele & Aronson, 1995; Wheeler & Petty, 2001). Moreover, these threat-induced performance deficits are most likely to occur when there is at least some identification with the domain (Spencer et al., 1999; Steele, 1997), which is almost certainly the case for women who are pursuing math-related careers.

Numerous studies have documented women's struggle with stereotype threat in the domain of mathematics (e.g., O'Brien & Crandall, 2003; Schmader, 2002; Schmader & Johns, 2003; Spencer et al., 1999), some of which have elicited lower performance through seemingly subtle cues. For instance, women who took a test in a group with a majority of male participants did worse than those tested with a majority of female participants (Inzlicht & Ben Zeev, 2000). The cues to elicit stereotype threat, therefore, range from the explicit, such as statements about typical gender differences (Keller, 2002; Smith & White, 2002), to the implicit, such as testing conditions with peers of the other sex.

Before considering further the consequences of stereotype threat for women in the domain of math, it is important to note that stereotype threat can affect any individual or group provided that the situation evaluates ability in a domain for which a relevant negative stereotype exists. A compelling set of studies has shown that both African American (e.g., Osborne, 1995, 1997; Steele, 1997; Steele & Aronson, 1995) and Latino/a American (Gonzalez, Blanton, & Williams, 2002) students perform worse on "intelligence" tests than do their European American peers when race or the diagnostic nature of the exam is made salient. Moreover, European American men are not immune to stereotype threat. Their performance suffers when stereotypes of inferiority are made salient, such as when they are subtly told that African Americans have more natural athletic ability (Stone, Lynch, Sjomeling, & Darley, 1999), that women are better at affective processing tasks (Leyens, Desert, Croizet, & Darcis, 2000), or that Asian Americans are superior mathematicians (Smith & White, 2002).

Finally, the effects of stereotype threat have been documented in children as well as in adults. Ambady, Shih, Kim, and Pittinsky (2001) found that lower elementary and middle school Asian girls performed worse on math tests when gender was made salient but better when race was made salient. Likewise, McKown and Weinstein (2003) found that children from stigmatized ethnic groups were not only more aware of broadly held stereotypes than nonstigmatized children were, but also that their awareness heightened the impact stereotype threat activation had on their performance.

Ego-Protective Responses to Stereotype Threat

Given the well-documented detrimental effects of stereotype threat on women's performance in math testing situations, it is important to consider how women may cope effectively with this threat. Previous researchers have considered a number of response strategies that individuals use to protect the self when faced with stereotype threat (for a discussion, see Pronin, Steele, & Ross, 2004).

Unfortunately, many women adopt an extreme ego-protective response of leaving the stereotyped domain altogether through a process of *disidentification* (Schmader, Johns, & Barquissau, 2004; see also Osborne, 1995; Steele, 1997). Disidentification occurs when, after having experienced repeated failures or threats of failure in a stigmatized domain, individuals remove the centrality of that domain from their self-concept (Steele, 1997). This process is consistent with the broader literature on self and social identity, which suggests that shifts in personal values often occur when individuals perceive that success is unlikely in a domain central to their identity (Crocker & Major, 1989; Major, Spencer, Schmader, Wolfe, & Crocker, 1998; Tesser & Campbell, 1980). Indeed, the coping response of disidentification, in conjunction with stereotype threat, could help to explain why African Americans have lower achievement outcomes (e.g., grades and test scores) than European Americans and why they show a pattern of weakening correlations between academic outcomes and self-esteem from the 8th to 12th grades (see Osborne, 1995, 1997). It could also help to explain why girls show better computational mathematical ability than boys until the 10th grade, at which point they begin to decline in math performance and continue to perform worse than male peers, even if they choose to pursue math as a profession (Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000; Meece et al., 1982).

Although disidentification may explain the process by which some women leave the domain of mathematics entirely, it does not explain how women respond to stereotype threat while operating within the domain. One important antecedent to disidentification may be *psychological disengagement*, which is the "defensive detachment of self-esteem from outcomes in a particular domain such that self-esteem is not contingent upon one's successes or failures in that domain" (Major et al., 1998, p. 35). For example, Osborne (1995, 1997) found that African Americans reported their global self-esteem to be higher than the reports of European Americans, even though their school performance was worse. In a similar vein, Major et al. (1998) found that African Americans' state-specific self-esteem did not change significantly as a result of positive or negative feedback that directly followed an "intelligence" test, whereas European Americans' self-esteem either increased or decreased depending upon their receipt of positive or negative feedback. Major et al. hypothesized that, because African Americans faced a negative stereotype about their intellectual ability, they had defensively removed any potentially harmful effects to their self-esteem by ceasing to care about their performance in that particular instance. Although this may be an effective response to one particular instance of negative feedback, over time it may lead individuals increasingly to view the domain as unimportant or not central to their self-concept (Osborne, 1999).

Another ego-protective strategy women may use when faced with stereotype threat is self-handicapping, that is, purposely lowering their performance in order to protect their self-esteem in case they should fail (Keller, 2002; Smith, 2004; for research on other stigmatized groups, see Steele & Aronson, 1995; Stone, 2002). There is a large literature that demonstrates both behavioral (e.g., choosing not to study) and self-reported (e.g., claiming not to have studied) handicaps when individuals feel pressure to defend their ability (e.g., Arkin & Oleson, 1998; Covington, 1992; Leary & Shepperd, 1986), and these techniques may also be applicable when responding to stereotype threat. Thus, when faced with stereotype threat, women may sabotage their own performance by failing to study, failing to try, or creating other barriers to success. Although they almost ensure failure, these techniques do allow poor performance to be blamed on controllable or external causes rather than on insufficient intelligence or ability.

Ensuring failure, however, is not an effective response for women who are highly identified with mathematics and wish to remain—and excel—in the domain. Although a necessary condition for stereotype threat is that individuals must be somewhat identified with the domain in order to experience lower performance, very little research has addressed the extent to which levels of domain identification may influence different responses to stereotype threat (cf. Pronin et al., 2004; Stone, 2002). Therefore, the present study examined how high math-identified women respond to stereotype threat differently than do low math-identified women or men. The responses of women who plan to remain in the domain of mathematics despite negative stereotypes about their ability could reveal potentially effective strategies for coping with stereotype threat.

One promising model for understanding the responses of high math-identified women to stereotype threat comes from research that specifies the processes by which individuals disengage their self-esteem from negative feedback. According to Schmader, Major, and Gramzow (2001), individuals can either *discount* the validity of a given evaluation by reporting that it is an inaccurate measure of ability or they can *devalue* the domain by reporting that performing well is not a personal value. Unlike devaluing, which is a domain-general response that is likely to lead to disidentification, discounting is a situation-specific response that may allow individuals to remain identified with a domain. Schmader et al. (2001) found that African Americans employed the strategy of discounting when they perceived racial discrimination, which arguably would serve to buffer self-esteem and allow for continued domain identification (for an attributional analysis of a similar phenomenon, see van Laar, 2000). In a similar vein, discounting may be employed by high math-identified women when they are faced with stereotype threat during challenging assessments of their mathematical ability. Like Schmader et al.'s (2001) African American participants, high math-identified women likely perceive sex discrimination (Steele, James, & Barnett, 2002), and, therefore, they may discount the validity of a given evaluation when they feel the pressure of stereotype threat.

The Present Study

This study was focused on women's use of discounting as it may be moderated by their personal identification with the domain of mathematics. Women and men who varied in their level of math identity took a challenging math test under either a stereotype threat or a stereotype nullification condition. We hypothesized that high math-identified women would be more likely to respond by discounting when faced with stereotype threat than would any other comparison group of men or women. High math-identified women are presumably heavily invested in the domain, and, consequently, they are likely to be affected by the pressure to avoid confirming a negative stereotype about female mathematicians. These women would reasonably expect to encounter math-related tasks and situations in the future, and, therefore, they may best be served not by disengaging from the domain but rather by discounting the validity of the specific test in question. In this case, discounting could potentially preserve their identity with the domain, protect their self-esteem, and offer them a means to return to the domain following a negative experience.

Before examining women's use of discounting, we first sought to document the standard threat-induced performance deficit. After adjusting for math SAT scores, we expected that women and men would perform equally well on a challenging math test when the negative stereotype was nullified, but that women would perform worse than men when the stereotype was invoked. Finally, we included measures of self-esteem and domain identification both before and after the difficult math test to explore the possibility that the discounting response could serve as an effective strategy for coping with stereotype threat.

METHOD

Participants

Participants were 121 undergraduates (68 women and 53 men) at a selective liberal arts college, most of whom were European Americans. They represented a broad range of years in school (33 first-year students, 28 sophomores, 29 juniors, and 31 seniors) and academic majors (38 math/science, 39 social science, 30 humanities, and 14 other). Recruitment took place through introductory psychology courses and fliers posted on campus. Because a highly selective admissions process ensured that participants would generally have high math SAT scores, all students interested in a "study of academic attitudes" were invited to participate. As an incentive, participants were offered a raffle ticket that made them eligible to win a cash prize.

Design

The experiment took the form of a 2 (women and men) × 2 (stereotype condition: stereotype threat and stereotype nullification) between-participants design. Dependent variables included test performance, discounting, posttest self-esteem, and posttest domain identification. In addition, participants' self-reported math identity was examined as a moderator of the effects of stereotype threat on the use of discounting.

Materials and Measures

Math Performance

Previous math performance was assessed through participants' math SAT scores, which were reported by all but five participants. This method of assessment has been used in previous stereotype threat research (e.g., Spencer et al., 1999) and allows for a standardized comparison across participants with different math-related backgrounds. Consistent with our assumption, math SAT scores were quite high in the present sample ($M = 662.72$, $SD = 83.37$), and women ($M = 641.67$, $SD = 81.23$) had significantly lower scores than those of men ($M = 685.57$, $SD = 80.24$), $t(114) = 2.92$, $p < .01$.

Math performance in this study was assessed with a 15-item multiple-choice test. Questions were taken from the math section of the general GRE and the advanced GRE in mathematics, as has been done in previous stereotype threat research (e.g., Spencer et al., 1999). Difficult GRE questions were used to mirror real-world testing environments, to ensure that participants took the test seriously, and to elicit the type of anxiety that is associated with standardized tests. All participants were given the same 15 problems to solve in a 15-min period. They were encouraged to guess if they were not certain of the correct answer.

Discounting

The discounting measure consisted of four statements to which participants could indicate their level of agreement on a 7-point Likert-type scale that ranged from 1 (*strongly disagree*) to 7 (*strongly agree*). The specific items were as follows: "This test is *not* an accurate measurement of my math ability"; "I feel that I am better at math outside of this test"; "Tests like this one are *not* biased against certain groups of people"; "My score on this test will likely be an accurate measure of my math ability." The third statement was taken from the work of Schmader et al. (2001), and the remaining three statements were created for the present study. The first and third statement were reverse-scored and all four items were combined to form a single scale that has reasonable internal consistency ($\alpha = .69$).

Math Identity

Participants' math identity was determined on the basis of their response to two statements: "I am good at math" and "It is very important to me that I am good at math" (Spencer et al., 1999). Participants indicated their agreement with each statement using an 11-point Likert-type scale that ranged from 1 (*very bad/not important*) to 11 (*very good/very important*). Responses to these two items were highly correlated in the present sample, $r = .47$, $p < .01$, and were therefore averaged together for all analyses. Overall, the present sample had a substantial range in math identity, and participants were—on average—moderately identified with math ($M = 6.50$, $SD = 2.12$).

Posttest domain identification was assessed with the following four items to which participants responded on a 7-point Likert-type scale: "Being good at math is *not* an important part of who I am"; "Doing well on mathematical tasks is very important to me"; "Success in math is very valuable to me"; "It usually *doesn't* matter one way or the other how I do in math." These items were based on the work of Schmader et al. (2001). The first and fourth items were reverse-scored, and all items were subsequently combined to form one internally consistent scale ($\alpha = .85$).

Self-Esteem

Baseline self-esteem was assessed with the Rosenberg Self-Esteem Inventory (RSEI; Rosenberg, 1965). The RSEI comprises 10 statements (e.g., "I feel that I have a number of good qualities") with which participants indicate their agreement using a 6-point Likert-type scale that ranges from 1 (*disagree*) to 6 (*agree*). This scale has been used widely with college populations, and reliability was good in the present sample ($\alpha = .89$).

Posttest self-esteem was assessed with the performance (seven items) and social (seven items) subscales from the State Self-Esteem Scale (SSES; Heatherton & Polivy, 1991). Items on the performance subscale assess participants' beliefs regarding their abilities (e.g., "I feel confident about my abilities"), and items on the social subscale assess participants' concerns about how others might view their abilities (e.g., "I am worried about whether I will be regarded as a success or failure"). Participants indicated their agreement with each statement using a 7-point Likert-type scale that ranged from 1 (*strongly disagree*) to 7 (*strongly agree*). Responses to the 14 items were combined to form one internally consistent scale ($\alpha = .80$).

Procedure

The procedure was identical for all participants with the exception of the stereotype manipulation, as described later. Participants were tested by the first author in mixed-sex groups of 6–10 in an environment that mirrored real-world testing situations (e.g., appropriately spaced desks and standardized answer sheets). After the experimenter had obtained informed consent from the participants, she stated that all instructions for the study would be given via audio tape in order to ensure consistency across sessions.[3] She then randomly selected a tape to insert into the recorder. In both conditions, a male voice explained that the goal of the study was to examine students' attitudes about various academic subjects. The voice explained that participants

[3]The person whose voice was recorded for the instructions and crucial manipulation was blind to the hypotheses.

would be asked to complete a number of questionnaires and to take a test that was considered to be an accurate measure of ability. Directly following these instructions, the experimenter, who was blind to the experimental condition of the particular session, distributed questionnaires that contained the measures that assessed baseline self-esteem, previous math performance, and math identity. Questionnaires also contained items about previous performance and identification with a number of other academic subjects so that the mathematical nature of the experiment and the content area of the test they were about to take were not yet apparent. Once the questionnaires were completed and collected by the experimenter, the tests and answer sheets were distributed face down. At this point, the experimenter played the taped instructions that delivered the crucial stereotype manipulation:

> Today you will be taking a math test. The math test you are about to take is an evaluative test used to predict the likelihood of an individual's success in a mathematical field. Research has examined age differences and found that this test predicts mathematical success equally well for all levels of college students—freshmen through seniors. And, as you may know already, there has been some controversy over whether there are gender differences in math ability. [Stereotype threat condition: *Although we don't yet fully understand this phenomenon, studies suggest that males and females DO perform differently on this test.* Stereotype nullification condition: *The math test you will be taking today has NOT shown gender differences in previous research.*]

The taped instructions reminded participants of the 15-min time limit, explained that there was no penalty for guessing, and asked participants to raise their hands if there were any questions at that point. Usually there were none. The experimenter then instructed the participants to begin working on the test.

During the test, the experimenter quietly monitored the room and gave participants two time-related warnings: one when 5 min remained, and one when 1 min remained. At the end of the 15-min testing period, the experimenter collected all of the tests and answer sheets and distributed the second questionnaire that contained the measures of posttest self-esteem, discounting, and posttest domain identification. The experimenter reminded participants that she would be leaving the classroom for a few minutes to run the tests through the scoring machine and that, upon her return, participants would see their scores and complete one final questionnaire. With the answer sheets in hand, she walked out of the door.[4]

Although the tests were not actually scored at this point, the experimenter waited outside the room for approximately 5 min. Upon returning, she sat quietly until all participants had completed the measures and then asked them to indicate their gender and any ideas they had about the goals of the study on the back of their questionnaires. No participants correctly guessed that the study manipulated test instructions, though a small minority suspected that gender differences were being investigated. Following this, the experimenter explained that the study was over and that participants' answer sheets had not been scored, but that she had been interested in their responses to the questionnaire items while they anticipated receiving a score. Finally, participants were thoroughly debriefed about the purposes and procedures of the study, given the opportunity to ask questions, and thanked for their participation.

RESULTS

Test Performance

The first hypothesis was that women in the stereotype threat condition would perform worse than men in the stereotype threat condition and both men and women in the stereotype nullification condition. Although some researchers have examined accuracy as an indicator of performance (e.g., Inzlicht & Ben-Zeev, 2000; Johns, Schmader, & Martens, 2005), we focused on the total number of test items correct as the primary dependent variable because participants had been encouraged to guess if they were not certain of the correct answer. Therefore, we examined differences in the number correct on the math test with a 2 (gender) × 2 (stereotype manipulation) ANCOVA; we used math SAT score as a covariate because it was highly correlated with math test performance, $r = .42$, $p < .01$. Means reported below have been adjusted to reflect this covariate.

[4]Because the experimenter was not blind to the hypotheses of the study, it is possible that her behaviors during the test and administration of the second questionnaire may have differed in subtle ways across experimental conditions. This would be a matter of concern if we were primarily interested in main effects. Because the key analyses of interest involve two- and three-way interactions, it is difficult to imagine how demand characteristics could account for significant effects.

Although there was no main effect of the stereotype manipulation, $F(1, 111) = 1.09$, *ns*, there was a main effect of gender such that women ($M = 6.14$, $SD = 2.02$) performed worse than men ($M = 7.10$, $SD = 2.04$), $F(1, 111) = 6.28$, $p < .05$, $\eta_p^2 = .05$. As predicted, there was a significant interaction between gender and stereotype manipulation on test performance, $F(1, 111) = 5.09$, $p < .05$, $\eta_p^2 = .04$. In the stereotype threat condition, women ($M = 5.52$, $SD = 1.99$) performed worse than men ($M = 7.33$, $SD = 1.98$), but in the stereotype nullification condition, women ($M = 6.76$, $SD = 2.01$) and men ($M = 6.88$, $SD = 2.04$) performed equally well, as shown in Fig. 1. A planned contrast that compared test performance of women in the stereotype threat condition to the other three conditions supported the hypothesis of a performance decrement, $F(1, 111) = 12.22$, $p < .01$, $\eta_p^2 = .10$. These findings replicate previous research that shows that women's performance in mathematics suffers under conditions of stereotype threat, but not when the threat is removed.

It is possible, however, that the stereotype manipulation might hinder performance especially for individuals who tend to perform worse in math, and has little to do with gender per se. As Yzerbyt, Muller, and Judd (2004) demonstrated, when a measured independent variable (i.e., gender) is crossed with a manipulated independent variable (i.e., stereotype manipulation) and adjusted for a covariate (i.e., math SAT score) that is related to the measured variable, the estimate of the interaction term is biased. Although in our case it might appropriately reflect a gender by stereotype manipulation interaction, it could also reflect a less theoretically interesting math SAT score × stereotype manipulation interaction. Therefore, we analyzed the test performance data using a regression approach that explicitly tested the preferred interaction and controlled for the alternative interaction. Consistent with our interpretation of the ANCOVA reported above, the interaction between math SAT score and stereotype manipulation was not significant, $\beta = .02$, $p > .80$, but the interaction between gender and stereotype manipulation was only slightly reduced, $\beta = .28$, $p < .07$. This analysis supports the conclusion of previous researchers that stereotype threat produces a specific performance deficit for the threatened group.

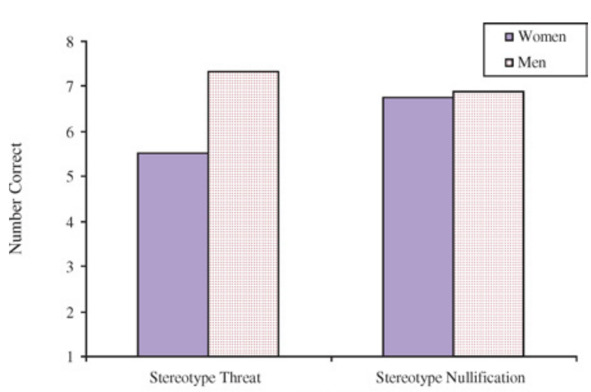

Fig. 1. Number of items correct by gender and stereotype condition. Means have been adjusted for math SAT score.

Discounting

The hypothesis of greatest interest was that high math-identified women in the stereotype threat condition would respond by discounting more than would any other group. To test this hypothesis, high and low math-identified groups were created on the basis of a median split of scores on the math identification measure (*Median* = 7). Differences in discounting were then tested using a 2 (gender) × 2 (stereotype manipulation) × 2 (math identification) ANOVA. There were no main effects or two-way interactions. As predicted, the three-way interaction was significant, $F(1, 113) = 5.73$, $p < .05$, $\eta_p^2 = .05$, and it revealed that the high math-identified women in the stereotype threat condition discounted more than did any other group. A planned contrast that compared high math-identified women in the stereotype threat condition to the other seven conditions also showed a significant difference, $F(1, 113) = 6.25$, $p < .05$, $\eta_p^2 = .05$. Moreover, Student–Newman–Keuls comparisons of all possible pairs of means showed no difference ($p > .05$) among the remaining seven conditions. Levels of discounting for women across the different conditions are displayed in Fig. 2, and descriptive statistics for all conditions are presented in Table I. It should be noted that, although the average group size in this analysis was 15 participants,[5] there were only

[5]Because of unequal cell sizes, we also created high and low math identity groups based on median splits computed separately for men (*Median* = 7.5) and women (*Median* = 6.0). Although this only slightly helped to equalize cell sizes (at most there was a shift of three participants to any one condition), we used this new domain identification measure to test for differences in discounting. The effects were highly similar to those found on the basis of the overall median split. The critical three-way interaction dropped to marginally significant, $F(1, 113) = 3.30$, $p = .07$, $\eta_p^2 =$

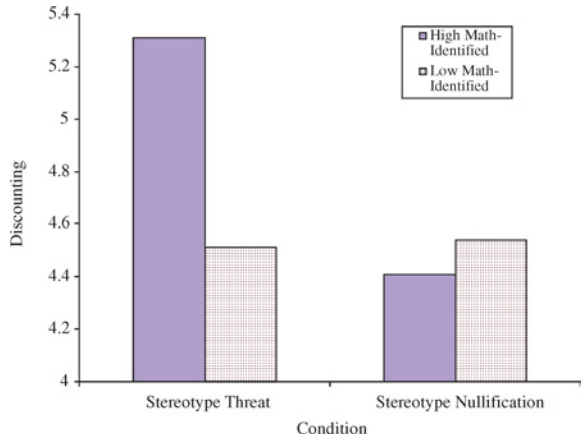

Fig. 2. Levels of discounting for women by math identification and stereotype condition.

6 low math-identified men in the stereotype nullification condition; the mean level of discounting for this group in particular should be interpreted with caution.

To confirm that high math-identified women discounted because of stereotype threat per se rather than simply because of poor performance on the test, we conducted a 2 (gender) $\times$ 2 (stereotype manipulation) $\times$ 2 (math identification) ANCOVA using test performance as a covariate. The three-way interaction pattern remained robust, $F(1, 112) = 5.63$, $p < .05$, $\eta_p^2 = .05$. A similar analysis was conducted using a regression approach to test our preferred interpretation of the three-way interaction and control for a possible three-way interaction between gender, test performance, and math identification (see Yzerbyt et al., 2004). Consistent with our interpretation of the ANCOVA results, the three-way interaction among gender, test performance, and math identification was not significant, $\beta = .05$, $p > .80$, but the preferred three-way interaction among gender, stereotype manipulation, and math identification remained robust, $\beta = -.69$, $p < .05$. Thus, women's use of discounting appears to be driven by stereotype threat per se rather than by poor performance on the test.

.03, but the planned contrast that compared high math-identified women in the stereotype threat condition to the other seven conditions showed a significant difference, $F(1, 113) = 6.51$, $p < .05$, $\eta_p^2 = .05$, and Student–Newman–Keuls comparisons of all possible pairs of means showed no difference ($p > .05$) among the remaining seven conditions.

Table I. Levels of Discounting by Gender, Stereotype Threat Condition, and Math Identification

	Stereotype threat			Stereotype nullification		
Math identity	*M*	*SD*	*n*	*M*	*SD*	*n*
Women						
Low	4.52	1.36	24	4.54	1.34	17
High	5.31	1.02	12	4.41	0.95	15
Men						
Low	4.67	0.83	13	3.75	1.02	6
High	4.31	1.01	16	4.67	1.39	18

Exploratory Analyses

We also examined the data for differences in posttest self-esteem and posttest domain identification. Self-esteem response was examined using a 2 (gender) $\times$ 2 (stereotype manipulation) $\times$ 2 (math identification) ANCOVA with baseline self-esteem as a covariate because it was highly correlated with posttest self-esteem, $r = .60$, $p < .001$. There was a significant main effect of gender such that women reported lower self-esteem ($M = 4.30$, $SD = 0.94$) than men did ($M = 4.64$, $SD = 0.76$), $F(1, 112) = 9.61$, $p < .01$, $\eta_p^2 = .08$. No other main effects or interactions approached significance.

Posttest domain identification was examined using a 2 (gender) $\times$ 2 (stereotype manipulation) $\times$ 2 (math identification) ANOVA. There was a main effect of math identification such that individuals who initially fell into the low math-identified group had lower posttest domain identification ($M = 2.71$, $SD = 1.01$) than did those who initially fell into the high math-identified group ($M = 4.29$, $SD = 1.19$), $F(1, 113) = 56.96$, $p < .001$, $\eta_p^2 = .34$, which verifies that our initial median split into high and low identity groups was effective. No other main effects or interactions approached significance. We also standardized participants' baseline levels of math identification and posttest domain identification to create change scores that could be examined using a 2 (gender) $\times$ 2 (stereotype manipulation) ANOVA. There was a main effect of gender such that posttest domain identification dropped for men ($M = 0.19$, $SD = 0.62$) but not women ($M = -0.15$, $SD = 0.87$), $F(1, 117) = 6.05$, $p < .05$, $\eta_p^2 = .05$. There were no other significant effects. On the basis of these analyses, there does not appear to be a significant loss in posttest domain identification for women relative to men, as might be predicted by some disengagement theories of coping with stereotype threat. Although

Table II. Intercorrelations Between Dependent Variables

Variable	1	2	3	4
1. Test performance	—	−.05	.19*	.27*
2. Discounting		—	−.07	.06
3. Posttest self-esteem			—	−.11
4. Posttest domain identification				—

*$p < .05$.

these findings do not show that high math-identified women's use of discounting prevents psychological disengagement from the domain, they do suggest that women faced with stereotype threat in this study did not suffer a loss to their math identification.

Finally, the relationships among test performance, discounting, posttest self-esteem, and posttest domain identification were examined with a series of bivariate correlations, as shown in Table II. As might be expected, test performance was positively correlated with posttest self-esteem, $r = .19$, $p < .05$, and positively correlated with posttest domain identification, $r = .27$, $p < .01$. No other relationships between variables were significant. Most notably, there was no correlation between discounting and either posttest self-esteem or posttest domain identification.

DISCUSSION

The primary aim of the current research was to explore women's response to stereotype threat on the basis of their level of identification with the domain of mathematics. Next, we first consider our replication of the performance deficits produced by stereotype threat and then discuss the findings with respect to discounting.

Performance

As predicted, women in the stereotype threat condition performed worse on the challenging math test than did any other comparison group of women or men. This finding replicates previous research that shows a decrement in women's performance on math tests in the face of stereotype threat, but not when the threat is eliminated or nullified (e.g., Schmader, 2002; Schmader & Johns, 2003; Spencer et al., 1999). The results of the present study also verify that this performance deficit was driven by membership in a stigmatized group (i.e., women doing math), rather than a tendency to perform poorly in math more generally. Indeed, the performance decrement for women in the stereotype threat condition remained robust even when we used procedures outlined by Yzerbyt et al. (2004) to control for a possible interaction between the stereotype manipulation and math SAT scores. This is a valuable contribution to the literature because the procedure eliminates a hidden confound that is present in many studies of stereotype threat (e.g., Steele & Aronson, 1995). Of course, the use of SAT scores as covariates in stereotype threat research more generally is a matter of some controversy (see Sackett, Hardison, & Cullen, 2004). It is important to emphasize that the "equal" performance of men and women in the stereotype nullification condition does not necessarily mean that they received numerically equivalent scores. Rather, it is the case that men and women in the stereotype nullification condition performed equally as well as their math SAT scores would predict, whereas women in the stereotype threat condition performed significantly worse than their math SAT scores would predict.

The Discounting Response

Our primary goal in examining discounting was to identify a response to stereotype threat that was unique to high math-identified women. Because high math-identified women are heavily invested in the domain, we expected that they would seek a response strategy that would allow them to dismiss a particular negative evaluation and, consequently, to persist in their math-related pursuits. Indeed, the results of this study showed that high math-identified women in the stereotype threat condition discounted the validity of the test more than did any other comparison group.

When interpreting this finding it is important to consider possible alternatives other than stereotype threat to which women may have been responding, such as their poor performance or the difficulty of the test itself. We propose that women in our study responded to stereotype threat. First, the use of discounting by high math-identified women in the stereotype threat condition remained significant even when test performance was used as a covariate in the analysis. Second, although women could have inferred that they had performed poorly on the test, neither test scores nor explicit feedback were given. High math-identified women, therefore, are likely to have responded to the threat of confirming the negative stereotype about women as inferior mathematicians rather than to their own suspected poor performance on the test per se. Third, high math-identified women did not respond exclusively to the difficulty of the test itself because discounting was observed only in the stereotype threat condition and not in the stereotype nullification condition. Thus, this study adds to the literature by identifying a response that is specific to stereotype threat and that is used primarily by those individuals most identified with the domain.

The results of this study also extend the work of previous researchers who have demonstrated that stereotype threat affects stigmatized group members' perceptions of test fairness (for a review, see Smith, 2004). Both African Americans in the domain of intelligence (Steele & Aronson, 1995) and European Americans in the domain of natural athletic ability (Stone et al., 1999) have been shown to discount evaluations as being biased or unfair when faced with stereotype threat. The present study extends the study of discounting processes to women in the domain of mathematics and, importantly, documents the moderating role of domain identification. Although some stereotype threat researchers have included measures of both discounting processes and domain identification (see Pronin et al., 2004; Stone, 2002), the present study is the first to show that the discounting response is unique to those individuals most highly identified with the stereotyped domain. The discounting measure used in the present study is also distinct from those measures that assess only the perceived bias of standardized tests in general (e.g., Steele & Aronson, 1995) as opposed to the perceived bias, unfairness, or inadequacy of a particular test participants have just completed. Future research is needed to determine whether high math-identified women would discount the validity not only of a particular threatening evaluation but also of standardized tests in general. Finally, some previous researchers have tested the extent to which discounting processes might serve as mediators of stereotype threat effects on performance, but no supporting evidence has been found to date (see Smith, 2004; Stone et al., 1999). The results of the present study support the conclusion that perceptions of test unfairness are best understood as a response to stereotype threat and not as a reason why stereotype threat leads to performance decrements.

Identifying the discounting response among high math-identified women also extends Schmader et al.'s (2001) research on discounting to a new stigmatized group and a new domain. Because Schmader et al. found that perceptions of racial discrimination led to discounting among their African American participants, it would be informative to measure women's perceptions of discrimination as well as their use of discounting in future research. Although women have been shown to experience higher levels of discrimination in male-dominated fields (e.g., mathematics) as opposed to more gender-balanced fields (Steele et al., 2002), the relationship between perceived discrimination and level of identification with a domain is unclear. It may be the case that high math-identified women experience high levels of both perceived discrimination and stereotype threat, which leads them to search out a response—such as discounting—that would allow them to maintain their identification with math by assuming that certain assessments are biased and unfair. Consistent with this scenario, previous research has shown that women's self-esteem is higher when they think they are being discriminated against, presumably because they have an external cause to which they can attribute failures in the domain (Crocker, Voelkl, Testa, & Major, 1991, but see Kobrynowicz & Branscombe, 1997). Future researchers should examine whether perceived discrimination leads to discounting and how identity, perceived discrimination, and discounting may influence and protect self-esteem.

Implications of the Discounting Response

The results of this study reveal that high math-identified women respond to stereotype threat by discounting, but what are the consequences of this response? Consistent with the broader literature on ego-protective responses to stereotype threat (e.g., Major et al., 1998; Pronin et al., 2004; Schmader et al., 2004; Steele, 1997), one might imagine that discounting serves as an effective coping strategy that buffers self-esteem and potentially helps women to persist in mathematical fields despite the threat of confirming a negative stereotype. The results of the present study, however, provide no direct evidence that discounting had such a benefit. Discounting was unrelated to both posttest self-esteem and posttest domain identification in correlational analyses. Moreover, high math-identified women in the stereotype threat condition showed no greater levels of either posttest self-esteem or posttest domain identification than did women in the comparison groups.[6]

More refined measurement procedures may be needed in order to test the extent to which discounting may serve as a coping strategy. For example, at this point, it is unclear exactly *when* during the test-taking and responding process a self-esteem deficit is likely to occur. Do women experience a self-esteem deficit following a negative evaluation and then subsequently respond to that deficit, or do they immediately respond to stereotype threat using an ego-protective strategy such that they fail to report a self-esteem deficit at all? Resolving these questions may be an important step toward testing for an esteem-protecting benefit of discounting. Our understanding of the consequences of discounting would also be enhanced by including larger cell sizes than those of the present study, which would allow for within-cell analyses of the correlations between discounting and posttest measures of domain identification and self-esteem. Finally, the inclusion of long-term measures would be informative, especially for examining the relationship between discounting and domain identification.

It is also important to consider the potential negative consequences of the discounting response. Do any possible self-protective benefits of discounting outweigh the potential costs of dismissing all failure feedback as due to external causes? What are the long-term consequences of discounting for domain-specific conceptual understanding? There may be alternative strategies that serve a coping function without some of these drawbacks. For example, the use of humor to cope with stereotype threat eliminates the standard performance deficit shown by stigmatized groups (Ford, Ferguson, Brooks, & Hagadone, 2004); perhaps it would also prevent women from leaving the domain of mathematics. It may also be useful to encourage women to adopt mastery-oriented or approach-oriented goals rather than the failure-avoidant goals that are likely to be invoked as a result of stereotype threat (Smith, 2004; see also Brown & Josephs, 1999). Future researchers might focus on the relationship between domain identification and these alternative coping strategies (see Pronin et al., 2004), as well as the extent to which potentially adaptive strategies could be taught effectively to members of stigmatized groups (see Johns et al., 2005).

[6]Interestingly, after calculating standardized change scores, we found that men showed a decrease in domain identification relative to women. This finding needs replication, but it indicates that women's math identification did not suffer as a result of the stereotype manipulation. The surprising drop in domain identification for men may be due to the fact that they began with higher levels of domain identification and, therefore, there was more room on the scale for scores to decrease. It may also be an artifact of our measures. A more developed and validated measure of domain identification (e.g., Smith & White, 2001) is recommended for future research.

Conclusion

This research contributes to the current stereotype threat literature in two ways. First, we replicated the basic finding that women suffer a performance deficit when facing stereotype threat, but that they perform equally as well as men when the threat is removed. Second—and more important—we began to examine the underlying dynamics of women's responses to stereotype threat, and demonstrated that high math-identified women respond to stereotype threat by discounting the validity of particular evaluations. Indeed, if we are to address the long-term effects of stereotype-induced inequities, we must understand the emotional responses and psychological defenses that accompany stereotype threat, including the ego-protective strategies used by members of stigmatized groups. This may especially be important for women who have a history of high achievement and involvement in the domain of mathematics in order to ensure that they can reach their full potential.

ACKNOWLEDGMENTS

Preparation of this article was supported by a grant from the office of the Dean of Faculty at Reed College. This article was based on research conducted by the first author and advised by the second author in fulfillment of the senior thesis requirement at Reed College. Portions of this research were presented at the February 2003 annual meeting of the Society for Personality and Social Psychology in Los Angeles, California. We are grateful to Kathy Oleson for her insight regarding our experimental design and data analysis, and for her comments on a previous draft of this article. We also thank Jim Holmes, John Huss, and Bob Adams for their assistance with making the audio tapes and Albyn Jones for his help constructing the math test.

REFERENCES

Ambady, N., Shih, M., Kim, A., & Pittinsky, L. T. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science*, *12,* 385–389.

American Association of University Women. (1995). *How schools shortchange girls: The AAUW Report*. New York: Marlowe.

Arkin, R. M., & Oleson, K. C. (1998). Self-handicapping. In J. M. Darley & J. Cooper (Eds.), *Attribution and social interaction: The legacy of Edward E. Jones* (pp. 313–347). Washington, DC: American Psychological Association.

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science*, *11,* 474–480.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, *210,* 1262–1264.

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

Coley, R. J. (2001). *Differences in the gender gap: Comparisons across racial/ethnic groups in education and work*. Retrieved March 29, 2005, from http://www.ets.org/research

Covington, M. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. Cambridge, UK: Cambridge University Press.

Crocker, J., & Major, B. (1989). Social stigma and self esteem: The self-protective properties of stigma. *Psychological Review*, *96,* 608–630.

Crocker, J., Voelkl, K., Testa, M., & Major, B. (1991). Social stigma: The affective consequences of attributional ambiguity. *Journal of Personality and Social Psychology*, *60,* 218–228.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56,* 5–18.

Eccles, J. S., Barber, B., Jozefowicz, D., Malenchuk, O., & Vida, M. (1999). Self-evaluations of competence, task values, and self-esteem. In N. Johnson, M. Roberts, & J. Worell (Eds.), *Beyond appearance: A new look at adolescent girls* (pp. 53–83). Washington, DC: American Psychological Association.

Eccles, J. S., Jacobs, J. E., & Harold, R. E. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, *46,* 183–201.

Educational Testing Service. (2002). *Sex, race, ethnicity, and performance on the GRE General Test*. Retrieved December 1, 2003, from http://www.gre.org/codelist.html

Fennema, E., & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization, and affective factors. *American Education Journal*, *14,* 51–71.

Ford, T. E., Ferguson, M. A., Brooks, J. L., & Hagadone, K. M. (2004). Coping sense of humor reduces effects of stereotype threat on women's math performance. *Personality and Social Psychology Bulletin*, *30,* 643–653.

Gonzalez, P., Guzman, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., et al. (2004). *Highlights from the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES 2005–005). [U.S. Department of Education, National Center for Educational Statistics.] Washington, DC: U.S. Government Printing Office.

Gonzalez, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, *28,* 659–670.

Harris, A. M., & Carlton, S. T. (1993). Patterns of gender differences on mathematics items on the Scholastic Aptitude Test. *Applied Measurement in Education*, *6,* 137–151.

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, *60,* 895–910.

Inzlicht, M., & Ben Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11,* 365–371.

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.

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, *47,* 193–198.

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

Kobrynowicz, D., & Branscombe, N. R. (1997). Who considers themselves victims of discrimination? Individual difference predictors of perceived gender discrimination in women and men. *Psychology of Women Quarterly*, *21,* 347–363.

Leary, M. R., & Shepperd, J. A. (1986). Behavioral self-handicaps versus self-reported handicaps: A conceptual note. *Journal of Personality and Social Psychology*, *51,* 989–997.

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

Major, B., Spencer, S., Schmader, T., Wolfe, C., & Crocker, J. (1998). Coping with negative stereotypes about intellectual performance: The role of psychological disengagement. *Personality and Social Psychology Bulletin*, *24,* 34–50.

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Development*, *74,* 498–515.

Meece, J. L., Parsons, J. E., Kaczala, C. M., Goff, S. B., & Futterman, R. (1982). Sex differences in math achievement: Toward a model of academic choice. *Psychological Bulletin*, *91,* 324–348.

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.

Osborne, J. W. (1995). Academics, self-esteem, and race: A look at the underlying assumptions of the disidentification hypothesis. *Personality and Social Psychology Bulletin*, *21,* 449–455.

Osborne, J. W. (1997). Race and academic disidentification. *Journal of Educational Psychology*, *89,* 728–735.

Osborne, J. W. (1999, April). *Historical trends in academic disidentification by race: 1972–1992*. Paper presented at the meeting of the American Education Research Association, Montreal, Canada.

Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology*, *40,* 152–168.

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.

Rosenberg, M. (1965). *Society and the adolescent self-image*. New Jersey: Princeton University Press.

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. *American Psychologist*, *59,* 7–13.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38,* 194–201.

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., Major, B., & Gramzow, R. H. (2001). Coping with ethnic stereotypes in the academic domain: Perceived injustice and psychological disengagement. *Journal of Social Issues*, *57,* 93–111.

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.

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, 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,* 46–51.

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in White athletes. *Personality and Social Psychology Bulletin*, *28,* 1667–1678.

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.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology*, *66,* 21–36.

Tesser, A., & Campbell, J. (1980). Self-definition: The impact of the relative performance and similarity of others. *Social Psychology Quarterly*, *43,* 341–346.

van Laar, C. (2000). The paradox of low academic achievement but high self-esteem in African American students: An attributional account. *Educational Psychology Review*, *12,* 33–61.

Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin*, *127,* 797–843.

Wilder, G. Z., & Powell, K. (1989). *Sex differences in test performance: A survey of literature* (College Board Report No. 89-3; ETS RR No. 89-4). New York: College Entrance Examination Board.

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.