Educational Psychology Review, Vol. 16, No. 3, September 2004 (© 2004)

Understanding the Process of Stereotype Threat: A Review of Mediational Variables and New Performance Goal Directions

Jessi L. Smith[1,2]

Stereotype threat is a situational experience in which an individual feels vulnerable and pressured by the possibility of confirming or being judged by a stereotype. This threatening experience leads to performance decrements, even among highly skilled individuals. This article chronicles empirically tested mechanisms for how stereotype threat negatively impacts performance outcomes. A review of relevant published investigations illustrate that a number of intuitive mediators have been suggested and tested, often with discouraging results. Thus, one objective of this article is to provide researchers with a comprehensive and straightforward account of such tested mechanisms to assist with future works. Indeed, there is much room for research in this area considering that to date, as measured, no individual mediator has completely explained the stereotype threat–poor performance relationship. As such, the second objective of this article is to propose a multiple mediator approach drawing from achievement goal theory. The Stereotyped Task Engagement Process Model is presented. This model hypothesizes that performance goal adoption can offer insights into the potential multiple processes involved in stereotype-threat effects on performance.

KEY WORDS: stereotype threat; performance; goal; achievement; approach-avoidance.

Over lunch, a computer science student discussed her classroom experiences. She reflected on how upset she felt when the other (all male) students ignored her when it came time to partner up for assignments. And they refused to look at her programs. It was as if they did not want to waste their

[1]Department of Psychology, The Ohio State University, Newark, Ohio.
[2]Correspondence should be addressed to Jessi L. Smith, Ph.D., Department of Psychology, 1179 University Drive, The Ohio State University, Newark, Ohio 43055; e-mail: smith.4612@osu.edu.

1040-726X/04/0900-0177/0 © 2004 Plenum Publishing Corporation

time with her because they thought she was not capable. Yet, she knew her skills were as good as or better than the other students. If only they would stop seeing her as a woman, and start seeing her as a peer . . . (author's files).

In the mid-1990s, Claude Steele and Joshua Aronson introduced the concept of stereotype threat (Steele and Aronson, 1995). Stereotype threat is the predicament created when an individual, who cares about a given domain, knows that a stereotype about her or his social group can provide a potential explanation for poor performance in that domain. That is, when an individual is placed in a stereotype-relevant domain, there exists a "threat in the air" propelling the possibility of the individual being treated or judged in terms of the stereotype (Steele, 1997).

Stereotype threat has become a "hot topic" at numerous professional conferences and seminars (e.g., annual meetings of the American Educational Research Association, 1998; American Psychological Society, 1999) and media outlets (e.g., *Newsweek* report by Begley, 2000). Interest in researching stereotype-threat effects is growing everyday. So far, these investigations have produced a fairly consistent picture of stereotype threat's deleterious effects on performance outcomes. *How* stereotype threat does its damage, however, has not been studied as systematically and is thus less clear. As Osborne (2001) pointed out, stereotype threat is "an effect without a clearly explicated mechanism" (p. 296).

The purpose of this article is twofold. First, it provides the first in-depth review of all of the various stereotype-threat mechanisms that have been empirically tested (noting how they were measured) and published: anxiety, evaluation apprehension, performance confidence, effort, self-handicapping, perceptions of a test's fairness, stereotype endorsement, and a mix of individual differences. Few systematic investigations have been undertaken to determine the mediators of the stereotype-threat–performance relationship (cf. Spencer *et al.*, 1999). Rather, the majority of stereotype-threat investigations have been carried out to replicate and generalize performance outcome effects. A number of these outcome studies have included some supplementary measures of possible mediators, yet no one single mediator has relished strong empirical support. However, two advantages of reviewing each tested mechanism are to draw attention to the assessments that were employed and to highlight the commonalities among the mediators. It is argued that stereotype-threat process analyses that are conceptualized as including multiple, related, mediators (as suggested by achievement goal theory) might be fruitful for future research.

The stereotype-threat mechanism field is still open to assessment, in part because no theoretical framework has emerged to direct testing. As such, the second purpose of this article is to introduce achievement goals as an illuminating part of the stereotype-performance mechanism process. *Achievement*

goals are broadly defined as wanting to develop or demonstrate task competence (Harackiewicz and Elliot, 1993). Achievement goals have emerged from a broader need for achievement framework that distinguished between approaching success and avoiding failure (e.g., Atkinson and Litwin, 1973). Considering the role of achievement goals in the stereotype-threat–performance relationship supplies an expectancy value and achievement motivation foundation to predict performance processes and outcomes (Atkinson, 1964). Indeed, performance-related achievement goals have proven useful in understanding motivational and performance outcomes in several competence-related contexts. Pulling from this literature, I present the Stereotyped Task Engagement Process (STEP) model as a starting point for programmatic research exploring the question: How does stereotype threat do its damage to performance?

STEREOTYPE THREAT DEFINED AND DETERMINED

Steele and others (e.g., Smith and White, 2002; Steele and Aronson, 1995) demonstrated that when a stereotype about expected performance is made salient, the feelings of threat that this creates leads individuals who are targets of that competence-related stereotype to actually perform more poorly. For example, when sixth-grade girls were reminded of the stereotype that girls are not good at math, their subsequent performance on a math test was impaired (Keller and Dauenheimer, 2003). However, it is not necessarily the case that the girls believed or endorsed the math stereotype. Rather, stereotype-threat effects are found among those individuals who are "identified" with the domain, that is those individuals who seem to care most about succeeding in the domain and are motivated to dispel any stereotypes (e.g., Steele, 1997).

Stereotypes are a part of the social milieu that affects both a stereotyped individual's expectations and values. As such, stereotype threat might be thought of as the situational performance result of a stereotyped individual's expectations for success/failure combined with the importance the individual places on doing well (cf. Hyde and Kling, 2001). Conceptually, stereotype threat is a "hot" motivational example (Wheeler and Petty, 2001) of classic expectancy value formulations (Atkinson 1964) applied to specific competence-related contexts (Eccles, 1994; Weiner, 1992).

Stereotype threat is couched in how others' expectations about performance success/failure affect the individual who is the target of those expectations. For some theorists, expectancies for performance success/failure are discussed in terms of the perceiver (the group or person who holds the expectations) and how the perceiver affects the target (the individual who is being stereotyped). For example, imagine an instructor who holds the stereotype

that African American students are unintelligent. An African American student in this instructor's class may not be called on to perform challenging cognitive tasks because of this perception. As such the student foregoes practice, performs poorly on tests, and thus appears to validate (self-fulfill) the instructor's stereotype. Research on the self-fulfilling prophecy (Jussim *et al.*, 2000; Rosenthal and Jacobson, 1966) has documented this effect, showing how a perceiver's perception (e.g., stereotypes) of members in a given social category determines in part how the individual is evaluated and treated (Jacobs and Eccles, 1992, see also Hebl *et al.*, 2002). In turn, the individual behaves in ways that actually confirm the perceiver's stereotype (Snyder *et al.*, 1977).

Stereotype-threat research has demonstrated that individuals who are targets of a negative competence-related stereotype do behave (perform) in a stereotype-consistent manner. Indeed, Steele and colleagues (e.g., Aronson *et al.*, 1999; Spencer *et al.*, 1999; Steele and Aronson, 1995) and others (e.g., Croizet and Claire, 1998; Keller, 2002; Stone *et al.*, 1999) demonstrated that when a stereotype about expected performance is made salient, the threat that this creates leads individuals who are targets of that stereotype to actually perform more poorly. This finding has been replicated for various stereotypes (e.g., those associated with race, SES status, age, sex) using a wide variety of tasks (e.g., sporting activities, verbal tasks, memory tests, math exams; e.g., Aronson *et al.*, 1998; Croizet and Claire, 1998; Levy and Langer, 1994; Rahhal *et al.*, 2001; Stone *et al.*, 1999) for students at different developmental stages (e.g., elementary school students, junior high students, high school students, college students; e.g., Ambady *et al.*, 2001; Keller, 2002; McKown and Weinstein, 2003; Osborne, 2001; Quinn and Spencer, 2001; Walsh *et al.*, 1999). See Wheeler and Petty (2001) for a detailed review of the effect of stereotype threat on performance.

In contrast to research that examines the perceiver's potentially negative impact on an individual's performance, other lines of expectancy value research expectancies for success/failure are discussed in terms of how the individual sees his- or herself. For example, research on self-efficacy has suggested that insofar as expectancies about success/failure affect an individual's confidence in his or her ability and skills to perform a task, task selection, task effort, task value, and the meaning of task performance can change (e.g., Bandura, 1997; Betz and Hackett, 1983; Bong and Skaalvik, 2003; Eccles, 1994; Molden and Dweck, 2000).

Thus, performance expectancy effects can differ across or within individuals over time. (See Eccles and Wigfield, 2002; for a detailed review of theories that integrate expectancies and value.) Research on stereotype threat is specific to individuals who value the task and outcome (e.g., high achievers, highly domain-identified people; e.g., Steele, 1997). In fact, stereotype-threat

effects are most detrimental to individuals who are identified with or who value the stereotype-relevant domain (i.e., see the domain as highly self-relevant, important, and as having favorable outcomes; Aronson *et al.*, 1999; Smith and White, 2001; Steele, 1997) or value the social group identity (Schmader, 2002).

The basic paradigm for demonstrating stereotype-threat effects is such that group members are made aware (or not) of a domain-relevant group-stereotype, and are then asked to take a (usually standardized) test in that domain. Stereotype threat has been manipulated in many different ways, including the use of techniques that are extremely subtle (having participants indicate their stereotyped demographic (race or SES) on a personal demographic sheet, e.g., Croizet and Claire, 1998), moderately subtle (characterizing the test as diagnostic of ability, e.g., Steele and Aronson, 1995, Study 1; having participants watch gender-stereotypic television commercials, Davies *et al.*, 2002), and not-so-subtle (having participants read a packet of articles claiming that a specific out-group is superior in the to-be-tested domain and then verbally reminded that in that domain, the out-group "outperforms" the in-group, e.g., Aronson *et al.*, 1999, Study 1).

One example of how stereotype threat is operationalized is provided by Inzlicht and Ben-Zeev (2000), who manipulated a gender-stereotype threat for females by requiring them to work on math problems in an environment where males outnumbered them. In a more elaborate inducement of stereotype threat, Blascovich *et al.* (2001) manipulated racial stereotype threat by having participants watch a video recording prior to taking a verbal test. In the stereotype-threat condition, Black participants viewed the video instructions delivered by a White professor. The professor discussed the debate on whether standardized tests were ethnically biased. The professor then claimed that researchers at Princeton, Stanford, and Michigan developed the new IQ test they were about to take and that data were being collected to achieve a "nationally representative sample." In comparison, Black participants in the nullified stereotype-threat condition watched a similar video, presented instead by a Black professor who claimed the IQ test was developed at Tulane University, Howard, and Michigan, and that the test was "the first step in an attempt to develop a culturally unbiased test." In line with past stereotype-threat research, Black participants in the stereotype-threat condition performed more poorly on the difficult verbal test items compared to all other participants (i.e., White participants in either condition and Black participants in the nullified-stereotype-threat condition).

A landmark demonstration of stereotype threat's effect on performance for high-achieving individuals was conducted by Spencer *et al.* (1999, Study 2). They found that high-math-achieving females (e.g., females who

scored at or above the 85th percentile on the SAT-math subsection, or received a grade of B or better in college calculus), who were told prior to taking a math exam that the test had shown stereotypical gender differences in the past, underperformed compared to their male counterparts. However, it is possible that the negative performance effects are simply due to real ability deficiencies (e.g., Benbow and Stanley, 1980). In a follow-up study, therefore, Spencer *et al.* (Study 3) set out to demonstrate that the results were due to the effect of the stereotype and did not simply reflect actual ability. Prior to taking the math exam, highly math proficient male and female students were either told that the math test showed no gender differences (nullifying the stereotype for that specific math test), or were told no information about gender differences on the test. When the gender stereotype was explicitly made irrelevant to performance, males and females performed equally well. Similar to Study 2, females who were given no explicit reminder of the stereotype underperformed compared to males. Presumably this occurred because the gender stereotype is so pervasive that just being in the situation is enough to trigger threat (e.g., Brown and Josephs, 1999; Quinn and Spencer, 2001; Smith and White, 2002).

Fortunately, research has also demonstrated ways to reduce the negative effects of stereotype threat. For example, rendering the stereotype information irrelevant to the specific task can experimentally nullify stereotype-threat effects and engender good performance outcomes for the individual (e.g., Spencer *et al.*, 1999). A decrease in reported feelings of dejection (e.g., frustration) is one mechanism shown to account for how rendering a task as impervious to the stereotype improves performance (Keller and Dauenheimer, 2003). Other research has suggested that interventions that include shaping stereotyped individuals' theories of intelligence (i.e., IQ as fixed vs. malleable) or spotlight other stereotyped people who are successful (i.e., role models who debunk the stereotype) might also help individuals cope with, and potentially overcome, stereotype threat (Aronson *et al.*, 2002; Blanton *et al.*, 2002; McIntyre *et al.*, 2003).

What accounts for stereotype threat's effect on performance? In an early synopsis of the mediators tested to explain stereotype-threat effects on performance, Aronson *et al.* (1998) recounted four studies that suggested "anxiety" was a partial mediator of the stereotype-threat–performance relationship. In the years since their summary, a plethora of studies have tested the anxiety hypothesis and several other related mediators. In the following section, I provide a detailed review of the various stereotype-threat mediators examined to date. As noted, a majority of the mediation examinations come from research that did not set out to test mediation specifically. Nevertheless, there are generally two categories of tested mechanisms: self-reported behavior and self-reported phenomenology.

STEREOTYPE-THREAT BEHAVIORAL MECHANISMS

Effort

Two competing "effort" hypotheses have been put forward as potential explanations for how stereotype-threat effects damage performance outcomes. First is the notion that stereotype-threatened individuals experience something similar to learned helplessness; they shut down and put forth less effort than their nonthreatened counterparts thus leading to poor performance. Second is the notion that stereotype-threatened individuals have an "I'll show them" experience and put forth more effort than their nonthreatened counterparts, thus essentially trying too hard leading to poor performance. However, as is shown below, the evidence for effort as a behavioral mediator is lacking.

Operationalizing effort as the amount of time spent on test items or decision time latencies per item has yielded no significant stereotype-threat effects across several studies (Aronson *et al.*, 1999, Study 1; Leyens *et al.*, 2000; Spencer *et al.*, 1999, Study 2). It should be noted that Steele and Aronson (1995, Study 2) did find marginal evidence that stereotype threat affected time spent on the first five items of their verbal exam (with threatened participants taking longer to answer), however the authors did not test time spent on the items as a mediator between stereotype threat and performance.

Similarly, Steele and Aronson (1995, Study 4) demonstrated that threatened participants completed fewer items than other participants, but the researchers did not test effort (operationalized as number of items attempted) as a mediator for performance. Several other studies, however, showed that effort, as number of items attempted, does not mediate the stereotype-threat–performance relationship (Ambady *et al.*, 2001, Study 1, Study 2; Shih *et al.*, 1999, Study 1, Steele and Aronson, 1995, Study 1, Study 2). Interestingly, although James and Greenberg (1997) also found that this index of effort did not mediate the relationship between their threat conditions and spatial ability scores, results did indicate the possibility of effort acting as a suppressor variable. This last finding has yet to be followed up.

In addition to operationalizing effort as time spent on the test items and as number of items attempted, self-report estimates of how much effort a participant exerts have been examined as well. Of the nine studies that have used this measure, eight did not find any sterotype threat experimental effects (Aronson *et al.*, 1999, Study 2; Gonzales *et al.*, 2002; Keller, 2002; Keller and Dauenheimer, 2003; Smith and White, 2002, Study 1 and Study 2; Steele and Aronson, 1995, Study 2; Stone *et al.*, 1999, Study 1, Study 2) and the one study that did find that threatened participants reported exerting more effort than nonthreatened participants found no statistical

evidence for mediation (Aronson *et al.*, 1999, Study 1). As measured across these 18 studies and 3 different measures (1 self-report, 2 behavioral), effort was not shown to completely or partially mediate the stereotype-threat–performance relationship when tested.

SELF-HANDICAPPING

The concept of self-handicapping was described by Leary and Shepperd (1986) as an individual's tendency to protect the self by either actually behaving or simply claiming to have behaved in a way that would provide a believable external attribution or reason (e.g., I feel under the weather more often than most people) for failure on an important task, such as an exam (Jones and Berglas, 1978; Hirt *et al.*, 1991; Urdan and Midgley, 2001). A number of claimed self-handicaps are also negatively correlated with actual test performance (Rhodewalt and Hill, 1995). Applying this concept to the stereotype-threat literature, then, it is possible that threatened individuals engaging in a stereotype-relevant task claim more self-handicaps than those not threatened, and that these claims mediate the relationship between stereotype-threat conditions and performance. As a first step in examining this self-handicapping hypothesis, Steele and Aronson (1995, Study 3) demonstrated that Black participants under stereotype-threat conditions reported getting less sleep and had less ability to focus than any of the other participants. Because participants did not complete any performance measures, it was impossible to test for mediation. Most researchers who have tested the self-handicapping mediation hypothesis (using self-report measures ranging from amount of sleep the night before to how focused and prepared the participant felt) have found no evidence for either complete or partial mediation (Croizet and Claire, 1998; Keller, and Dauenheimer, 2003; Kray *et al.*, 2001, Study 4; Shih *et al.*, 1999, Study 1; Stone *et al.*, 1999, Study 1). One study has found evidence that high school girls claim self-handicaps (e.g., being stressed) more under stereotype threat and that these handicapping claims mediated the girls' (poor) math performance (Keller, 2002). Thus, although in most studies claimed self-handicapping was not shown to completely or partially mediate the stereotype-threat–performance relationship, evidence was presented by Keller (2002) that claiming more self-handicaps mediated the effects of stereotype threat on task performance.

One potential index of behavioral self-handicapping comes from research by Stone (2002, Study 1 and Study 2), who found evidence that stereotype threatened participants do self-handicap by opting not to practice for an ostensibly upcoming golf task. Participants in theses golf studies, however, did not actually complete the golf task so a stereotype-threat

golf-performance mediation test was not possible. Research by Quinn and Spencer (2001) found that compared to all other participants, females under stereotype-threat conditions were less frequently able to formulate a strategy for solving difficult mathematical word problems. Although Quinn and Spencer also did not test study strategy use as a possible mediator between stereotype threat and performance, their initial results for strategy use are encouraging and inline with the prediction that stereotype-threatened individuals may be doing something behaviorally that affects their performance. To date, behavioral self-handicapping has not been adequately tested as a potential stereotype-threat–performance mediator.

STEREOTYPE-THREAT PHENOMENOLOGICAL MECHANISMS

Anxiety

Drawing from the test anxiety literature (e.g., Sarason, 1972), one possible explanation for how stereotype-threat effects negatively impact performance is that the stereotype creates a generalized sense of anxiety that disrupts the individual's ability to perform. This intuitive hypothesis was tested a number of times with various measures of anxiety. Initially tested by Steele and Aronson (1995, Study 1) and later by other researchers using a variation of Sarason's (1972) measure of cognitive interference, no effects were found (Gonzales *et al.*, 2002; Keller and Dauenheimer, 2003; McKown and Weinstein, 2003). That is, regardless of the participant's race and/or the stereotype-threat test-description condition, participants reported the same degree of "disruptive thoughts." Following up these findings, Steele and Aronson (Study 2) again tested the anxiety hypothesis, using the cognitive interference measure, but added a different measure of anxiety—Spielberger *et al.*'s (1970) state-trait anxiety instrument (STAI)—but no effects were found. That is, regardless of the participant's race and/or the stereotype-threat test-description condition, participants reported the same degree of anxiety and disruptive thoughts while taking the test.

Despite these initial null findings, Spencer *et al.* (1999, Study 3) set out to reevaluate the anxiety hypothesis, using the STAI but with a slightly more elaborated and less subtle stereotype-threat manipulation. Although a marginal experimental effect on STAI was found, anxiety was not found to be a mediator between stereotype-threat conditions and performance. These mixed results (condition effects without mediation) were interpreted by Spencer *et al.* (1999) to suggest that anxiety is a "plausible" mediator. Unfortunately, future tests of the anxiety hypothesis using STAI demonstrated null results (Aronson *et al.*, 1999, Study 1 and Study 2; Hausdorff *et al.*,

1999 Study 1; Oswald and Harvey, 2001; Schmader, 2002; Stone *et al.*, 1999, Study 2).

With respect to other measures of anxiety, research has not demonstrated anxiety as a clear mediator of the stereotype-threat–performance relationship (e.g., Keller and Dauenheimer, 2003). For example, Brown and Josephs (1999, Study 1) implemented a six-item word fragment test (e.g., idiot, dumb) designed to capture anxiety by measuring the ease that performance-anxious-related words were accessible, but they found no effects. In a similar vein, Osborne (2001) examined data from over 28,000 high school students. Like Brown and Josephs, Osborne did not manipulate stereotype-threat exposure. Nevertheless, he examined the naturally occurring relationship between anxiety and performance on three high school achievement tests. All students were asked how they "felt" while taking the exams. Using a "yes/no" response format, they were provided with a list of items to select from to answer the question (tense, under pressure, under strain, nervous, uneasy, calm, afraid, and uncomfortable). The number of "yes" items was then summed together to form the anxiety construct. Osborne found evidence that students of different races reported different levels of anxiety. Although the study's power was quite high, results indicated that anxiety only partially mediated the relationship between race and performance for two (White vs. African Americans, and White vs. Latino) of three race comparisons; no mediation was found for the White versus Native American comparison. Osborne repeated the analyses for male versus female students and similarly found that anxiety was a partial mediator of the gender-performance relationship. Although results are somewhat encouraging and in line with the anxiety hypothesis, the author was quick to point out that the magnitude of the differences was "very small."

Other studies have used more situationally specific measures of anxiety. In Stone *et al.* (1999), for instance, Black and White participants were asked to play a game of golf under varying conditions of stereotype threat. In Study 1, participants completed a five-item self-reported pre- and postperformance measure of situational anxiety (Mattsson, 1960, as cited in Stone *et al.*, 1999). In Study 2, participants completed a six-item postperformance measure of "Competitive State Anxiety Inventory" (e.g., my hands are clammy; Martens *et al.*, 1990, as cited in Stone *et al.*, 1999). Stone *et al.* found no evidence for mediation, using any of the anxiety measures (see also Stone, 2002). Thus, results for anxiety as a phenomenological mediator are mixed: although 2 studies using two different self-report measures found small but reliable evidence for anxiety as a partial mediator across 15 other studies and six measures (four self-report, one cognitive interference, one word fragment), anxiety did not emerge as a complete or partial mediator of the stereotype-threat–performance relationship when tested.

Evaluation Apprehension

A similar, but alternative, phenomenological mediational hypothesis is that stereotype threat negatively affects an individual's performance by creating concern for how other people are evaluating him or her and feelings of self-consciousness. This experience in turn is expected to undermine performance. This evaluation apprehension (e.g., Mullen, 1986) hypothesis, although less popular than the anxiety hypothesis, has received some attention by stereotype-threat researchers. To start, Steele and Aronson (1995, Study 4) operationalized self-conscious concern by asking participants to indicate the extent to which listing their race (the manipulation of stereotype threat in this study) "bothered them" and found no experimental effects. Spencer *et al.* (1999, Study 3) constructed their own four-item self-report measure of evaluation apprehension (e.g., If I do poorly on this test, people will look down on me) and found no evidence for mediation. This measure was also administered in a stereotype-threat study conducted by O'Brien and Crandall (2003), but evaluation apprehension did not mediate the stereotype-threat–performance relationship.

To test the hypothesis that evaluation apprehension would mediate the relationship between stereotype-threat effects and performance, Aronson *et al.* (1999, Study 2) examined race stereotypes in the math domain. In this study, White males were asked to take the GRE math subject test under varying conditions of an "Asians are better at math than Whites" stereotype threat. Following performance, participants completed a self-report item assessing the extent to which they wondered if the experimenter was judging them. Although participants who were highly identified with math and experimentally threatened by the stereotype were found to "wonder" what the experimenter thought about them more often, this measure of evaluation apprehension did not mediate the relationship between stereotype threat and performance. Thus, across these four studies and three self-report measures, evaluation apprehension did not emerge as a complete or partial mediator of the stereotype-threat–performance relationship.

Performance Confidence

Although the measured self-report results for generalized anxiety and evaluation apprehension do not appear to account convincingly for the damaging effect of stereotype threat on performance, another related possibility is that the stereotype threat impacts an individual's self-confidence to perform the specific test that then, in an almost self-fulfilling nature, disrupts performance (Rosenthal and Jacobson, 1966). This performance confidence

hypothesis has been tested using the most diverse set of operationalizations of all the tested mediators. Steele and Aronson (1995, Study 1) provided the first test by conceptualizing performance confidence as feelings of "academic competence" (e.g., I feel confident about my abilities), but they found no experimental effects. Following up their findings, Steele and Aronson then employed a "self-doubt" implicit (i.e., word-fragment completion) self-report measure (Study 3) and found some indication of stereotype-threat effects (with Blacks in the stereotype-threat condition producing the most self-doubt-related word fragments), but because participants in Study 3 did not actually take any performance measures, a test of mediation was not possible. No effects were found in Steele and Aronson's fourth study, which used a self-report item to measure how well the participants' thought they had performed. Stone (2002, Study 2) also found that compared with non-threatened golfers (i.e., Blacks told golf success is related to natural athletic ability) threatened golfers (i.e., Blacks told golf success is related to intelligence) expressed more implicit self-doubt (spontaneously filling in the word "awkward" on a word fragment completion test). This word fragment completion result partially mediated how much participants practiced for an upcoming golf task.

A number of stereotype-threat studies have asked participants to estimate their performance following task completion, but performance estimates did not mediate the stereotype-threat–poor performance relationship (Aronson *et al.*, 1999, Study 2; Keller, 2002; Kray *et al.*, 2001, Study 4; Shih *et al.*, 1999, Study 1). Performance estimates assessed prior to task performance have shown mixed results. In one study, participants were asked to draw a bar graph to illustrate how well they thought they would perform on an upcoming math test, and this estimate partially mediated actual math performance (Cadinu *et al.*, 2003). However, mediation was not found for a study that used a one-item self-report preassessment of performance expectancies (Sekaquaptewa and Thompson, 2003).

Rather than asking participants to estimate performance, Kray *et al.* (2001, Study 1) had participants in their negotiation stereotype-threat study estimate how successful they would be in a classroom demonstration of a business negotiation, compared to their partner, before beginning the task. Stereotype-threatened females (i.e., told that the negotiation task was diagnostic of their ability) expected to perform worse, but this measure of performance expectancy did not mediate the stereotype-threat–performance relationship.

Similarly, Stone *et al.* (1999, Study 2) had participants in their golf study predict the number of strokes they expected to need to complete the course prior to each round. These predictions were then summed together as an index of performance expectancies. Results indicated that although

performance expectancies did explain some of the moderating variables in their study (i.e., athletic disengagement) and were correlated with performance outcomes across all stereotype-threat conditions, performance expectancies did not clearly mediate the stereotype-threat–performance relationship. The authors note a possible "reciprocal relationship" between their assessment of expectancies and actual golf performance that may have clouded their mediational test, such that prior performance on each round of golf might have also influenced performance expectancy ratings on the next round (see Stone *et al.*, 1999, footnote 4). This notion suggests another conceptualization of performance confidence; namely, the self-reported confidence that a stereotype-threatened participant has in his or her performance. This "confidence" conceptualization was also tested, but did not mediate the stereotype-threat effects on performance (Aronson *et al.*, 1999, Study 1 and Study 2; Kray *et al.*, 2001, Study 2; Gonzales *et al.*, 2002).

Another variation of the performance confidence hypothesis is assessing perceptions of personal power within the testing situation. Kray *et al.* (2001, Study 4) demonstrated that within mixed-sex dyads only, males perceived more power (with scale endpoints of "0–other person had the advantage," "50%–equal power," to "100%–I had the advantage") than females in the stereotype-threat condition, whereas amount of power experienced during the negotiation was perceived equally for males and females under non-stereotype-threat conditions. In this same study, an inconsistent pattern of results emerged for a different measure of performance expectancies. Participants were asked to rate how "challenged" they felt during their negotiations. However, unlike the perceptions of power assessment, Kray *et al.* (2001, Study 4) found no gender differences in perceptions of challenge under stereotype-threat conditions or under nonstereotype conditions. Given these counter results for their two items of performance confidence, it is not surprising that the authors did not report the relationships between perceptions of power and performance, or between perceptions of challenge and performance, and, hence, did not test either as a mediator of stereotype effects on performance.

Related to perceptions of challenge, performance confidence has also been operationalized as the perceived difficulty of specific test items or the test overall. Although task difficulty has moderated stereotype-threat effects (only difficult tests show stereotype-threat effects; O'brien and Crandall, 2003), task difficulty did not emerge as a mediator (Aronson *et al.*, 1999, Study 1, Study 2; Keller, 2002; Shih *et al.*, 1999, Study 1; Smith and White, 2002, Study 1 and Study 2). Indeed, even conceptualizing performance expectancies as self-reported self-efficacy (e.g., I can handle this test) has demonstrated no relationship between performance expectancies and stereotype-threat manipulations or performance outcomes (Oswald and

Harvey, 2001; Spencer *et al.*, 1999, Study 3). Across 18 studies and 11 measures (10 self-report, 1 thought listing), performance expectancies did not emerge as a clear mediator when tested. For two studies using word fragment implicit measures (Steele and Aronson, 1995; Stone, 2002) and for one study assessing performance expectancies prior to task engagement (Cadinu *et al.*, 2003), some evidence for partial mediation of the stereotype-threat–performance relationship was found.

Stereotype Endorsement

In a study by Leyens *et al.* (2000), the authors investigated whether stereotype-threat effects could be produced with a stereotype that was only situationally induced versus a stereotype perpetuated by mainstream society. Leyens *et al.* tested the extent to which endorsing the stereotype that males are not as good as females at processing affective information accounted for the relationship between stereotype-threat and number of errors males made on an affective test. Using assessments of stereotype endorsement at both the group level (i.e., gender groups) and acceptance of the stereotype at the individual level yielded no evidence for mediation.

Perceptions of the Test

External attributions, namely that a test is biased or unfair, was also suggested as a potential mediator for stereotype-threat effects on performance (Steele, 1999). This test perception hypothesis is based on Steele's earlier work showing that Blacks under stereotype threat reported perceiving an upcoming exam as unfair to a greater extent than the other participants (Steele and Aronson, 1995, Study 3). However, because participants did not actually take the exam in that study, testing for mediation was not possible (see also Stone 2002, Study 1 and Study 2). In two of the earlier golf studies conducted by Stone *et al.* (1999, Study 1 and Study 2), however, performance measures were collected in addition to asking participants to rate their perceptions of the golf test as biased. Results demonstrated no evidence for the test perception hypothesis. Measuring test perceptions via a more positive frame (ratings of liking the test) have also yielded no evidence for test perception mediating the relationship between stereotype-threat effects and performance (Shih *et al.*, 1999, Study 1). Thus, as measured in these studies, test perceptions did not emerge as a complete or partial mediator of the stereotype-threat–performance relationship.

Feelings About the Self

A handful of intuitive measures of changes in feelings about the self were also tested to account for stereotype-threat effects on performance. Specifically, Levy and Langer (1994) found that self-esteem (as measured by the Rosenberg Self-Esteem Scale) did not mediate the influence of negative aging stereotypes on cognitive performance of the elderly (see also Stone 2002, Study 1 and Study 2). Similarly, Oswald and Harvey (2001) did not find evidence that state self-esteem (using the scale by Heatherton and Polivy, 1991) mediated the relationship between a participant's sex and math performance. Finally, Hausdorff *et al.* (1999) employed the geriatric depression scale and found no evidence that depression mediated the relationship between stereotype-threat conditions and walking gait speed and swing time in the elderly.

The value placed on the stereotype-relevant social group (e.g., social identity) and the value placed on the domain in relation to the self (i.e., domain identification) have also been investigated as possible constructs mediating the stereotype-threat–performance relationship. To examine the role of social identity, Schmader (2002) administered the four-item importance subscale of the Collective Self-Esteem Scale (Luhtanen and Crocker, 1992) to participants. The scale was reworded to assess the value placed on gender identity. Participants then took a 20-item multiple-choice math test under different levels of gender salience. Participants in the gender-salience condition were told their individual performance on the test would be used as an indicator of women's and men's math ability in general. The typical pattern of results expected for stereotype threat would be a two-way interaction between sex of the participant and gender salience, such that compared to males, females would perform the worst when told their performance would be indicative of all women. Results indicated that gender identity moderated this typical stereotype-threat interaction, with only females who were strongly identified with their gender performing poorly on the math test when gender was made a salient part of the task. Schmader thus found evidence for social group identity as an important moderator of stereotype-threat effects, but did not conceptualize it as a possible mechanism.

The extent to which an individual cares about a given stereotyped domain was also examined (e.g., Cadinu *et al.*, 2003; Josephs *et al.*, 2003). For example, Aronson *et al.* (1999, Study 2) found some evidence that their one-item measure of academic identification (i.e., participants' ratings of the importance of math abilities to their self-concept) moderated the stereotype-threat–performance relationship, but they did not test for mediation. Smith and White (2002) found that their nine-item measure of Domain Identification (Smith and White, 2001) was a covariate of math performance for both

men and women in a stereotype-threat paradigm, but they did not test for mediation. Nevertheless, Leyens *et al.* (2000) did test if their two-item measure of identification mediated the relationship between stereotype threat and performance but found no supporting evidence.

SUMMARY AND LIMITATIONS

Studies examining stereotype-threat effects on performance outcomes have often included behavioral and phenomenological variables to examine possible mechanisms accounting for the stereotype-threat–performance relationship. No one construct emerged as a strong reliable candidate for a complete mediator, however some evidence for anxiety and performance confidence as partial mediators did emerge in some studies. Other variables were not always adequately tested or replicable as mediators (e.g., self-handicapping and test-bias perceptions). Thus, how stereotype threat does its damage to performance remains unclear.

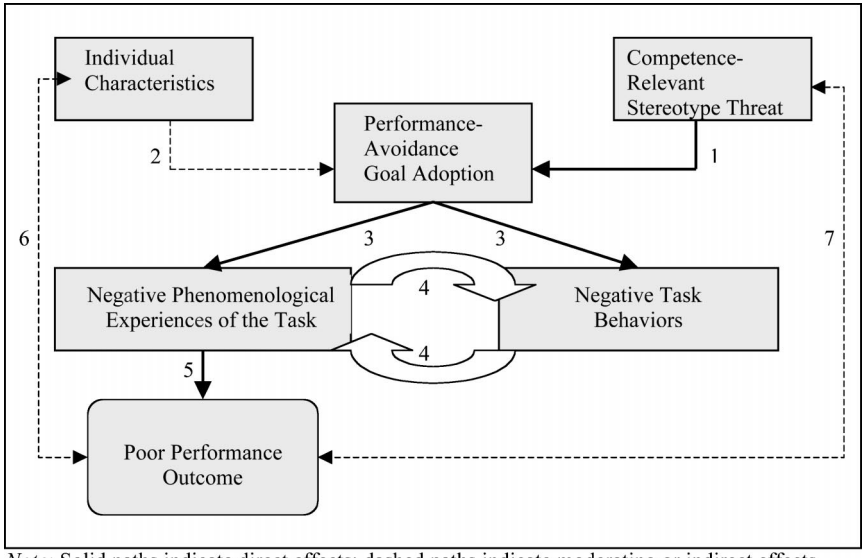
No single mediator should be prematurely dismissed. Few of the tested mediators were assessed with non-self-report methods, and self-report may not be the most ideal method to capture the stereotype-performance process (e.g., Nisbett and Wilson, 1977), particularly if the stereotype-threat mechanism is operating at a nonconscious level (as discussed by Croizet and Claire, 1998; see also Oswald and Harvey, 2001). For those studies that did include non-self-report measures, results were mixed: response-latency indexes of effort, cognitive interference indexes of anxiety, word-fragment indexes of anxiety, and thought listing indexes of self-efficacy yielded similar null mediation results. On the other hand, a number of the other non-self-report measures were not subjected to a mediation test (usually because of methodological limitations). For example, evidence for word-fragment indexes of performance confidence (self-doubt) as a potential mediator emerged but because participants did not complete a performance measure, mediation could not be tested. Different types of measures might help to illuminate the stereotype-threat–performance mechanism.

Measures that can assess multiply determined and/or implicit (latent) mechanisms might also lead to better understanding. For example, Blascovich *et al.* (2001) examined if stereotype threat leads to an increase in mean arterial blood pressure (MAP) to test the hypothesis that stereotype threat triggers something (effort, anxiety, expectancies, or apprehension) or some combination of factors that engenders increased blood pressure (BP; cf. Cacioppo and Tassinary, 1990). This hypothesis might be viewed as a holistic test of the effort, anxiety, expectancies, and apprehension hypotheses

because general BP responses are often an index of all four factors (e.g., Gendolla, 1998; Larkin *et al.*, 1998; Tennyson and Woolley, 1971; Wright *et al.*, 1986). Participants in the Blascovich *et al.* study worked on a verbal task during which their MAP was assessed once per minute. Results indicated that after controlling for baseline levels of MAP, African American participants under stereotype threat had greater MAP responses than all other participants. Furthermore, stereotype-threat participants performed more poorly on the difficult test items. Blascovich *et al.* tested if performance accounted for the stereotype-threat–BP relationship and found that it did not. Inopportunely, however, the authors did not test for the alternate (and of interest here) mediation. That is, the authors did not test if BP (as an index) mediated the stereotype-threat–performance relationship. This promising hypothesis should be tested in future investigations.

The Blascovich *et al.* (2001) study also suggests that there might not be a sole stereotype-threat mechanism. There may be different processes for different stereotypes and groups (Aronson *et al.*, 1998; Cadinu *et al.*, 2003) rendering it difficult to find evidence for any one mediator. It is possible that stereotype threat influences both behavioral actions and phenomenological experiences during performance in a reciprocal feedback loop making it necessary to examine the relationship among the process variables (e.g., Cury *et al.*, 2002; Harackiewicz and Sansone, 2000). Indeed, a single explanatory process analysis might not only be incomplete but also account for why stereotype-threat mediation tests results are mixed. If the stereotype-threat–performance mediating chain of process variables is long and interconnected, single mediation tests could be statistically unpowerful and thus difficult to document (Judd and Kenny, 1981). Power to detect mediation depends on the effect size of stereotype threat on performance, and this effect size is attenuated by the effect sizes of the mediator(s). Thus, if the effect of any one of the (single) mediators tested is small or moderate, then demonstrating mediation is limited by the power of that single mediator (J. Butner, personal communication, September 3, 2003).

Multiple, reciprocal mediators might account for the stereotype-threat–performance relationship. I posit a general *Stereotyped Task Engagement Process* (STEP) model to illustrate what one such multiple mediator chain might look like (see Fig. 1). The STEP model bridges stereotype-threat research with achievement goal research. The achievement goal literature posits that performance expectancies drive the type of achievement goal one adopts (Elliot and Church, 1997). In a situation where a stereotype exists about an individual's competence, determining the nature of the individuals' achievement goal focuses attention on a *pattern* of behavioral and phenomenological variables.



Note: Solid paths indicate direct effects; dashed paths indicate moderating or indirect effects.

Fig. 1. The proposed stereotyped task engagement process.

A MULTIPLE MEDIATOR PROPOSAL: EXTRACTING FROM THE ACHIEVEMENT GOAL LITERATURE

Although they have developed in virtual isolation, research on achievement performance goals links well with the stereotype-threat research. Both literatures focus on competence and performance in achievement settings. Goal research suggests that how an individual performs a task depends on his or her goals (Sansone and Harackiewicz, 1996; Sansone and Smith, 2000). Understanding the role of achievement goals in a stereotype-threat situation (e.g., women working in computer science) is important to consider because extant goal research suggests that individuals of all ability levels will interact with, experience, and perform tasks differently depending on the type of goal they spontaneously adopt or are assigned to adopt (e.g., Barron and Harackiewicz, 2001; Dweck and Leggett, 1988). This potentially adds the missing mechanism piece to the stereotype-threat puzzle. For example, when an individual feels pressured to demonstrate that he or she has high ability, achievement goal research predicts that he or she will become concerned with performance relative to other people, feel less interested in the task, and ultimately perform poorly on difficult tasks (Nicholls, 1984).

In the achievement goal literature, goals were historically dichotomized. For example, a distinction was made between task-involved and ego-involved goals (Nicholls, 1984), task-oriented and performance-oriented goals (Maehr

and Midgley, 1996), and mastery-oriented and performance-oriented goals (Ames, 1992). The goal dichotomies all emphasize the distinction between a focus on the task for the sake of learning (mastery) and a focus on the task for the sake of a reward (performance). Making a distinction between two types of achievement goals is important because mastery goals and performance goals are associated with unique performance processes and outcomes (e.g., Elliot and Thrash, 2001).

Performance goals are characterized as a focus on ability relative to other people (e.g., Harackiewicz and Elliot, 1993). It was traditionally believed that having a performance goal results in poor performance (e.g., Ames, 1992). More recently, however, it has been demonstrated that not all performance goals produce negative outcomes (e.g., Barron and Harackiewicz, 2001; Elliot and Church, 1997; see also Rawsthorne and Elliot, 1999). Rather, it is necessary to distinguish carefully between performance *approach* goals (PAP-goal) and performance *avoidance* goals (PAV-goal). A PAP-goal is defined as wanting to demonstrate competence (e.g., Elliot and Church, 1997; e.g., I want to do the best on the exam). This type of goal is rooted in an individual's overall need for achievement and associated with good performance (Elliot and McGregor, 2001). A PAV-goal, on the other hand, is defined as wanting to avoid demonstrating incompetence (e.g., Pintrich, 2000; e.g., I want to avoid failing the exam). This type of goal is rooted in an individual's fear of failure and is associated with poor performance (Elliot and McGregor, 2001).

The STEP is proposed as a model for linking stereotype threat and achievement goal research. It is expected that stereotype threat triggers a PAV-goal because PAV-goals are conceptually similar to the state in which individuals feel threatened to prove that a competence-stereotype is not valid. One advantage of adding PAV-goals to the stereotype-threat mediation puzzle is extant research that predicts the behavioral and phenomenological processes set in motion by PAV-goal adoption. It is not expected that PAV-goal adoption in and of itself explains the stereotype threat-poor performance relationship. Rather, PAV-goal adoption leads an individual to predictably self-regulate his or her behavior and motivation according to potential negative outcomes (Elliot and Church, 1997).

The Stereotyped Task Engagement Process

The STEP is a model for bridging together the research in stereotype threat and achievement goal theory to better understand the complex experience of stereotyped individuals working on competence-related activities. The STEP model (see Fig. 1) was designed to be predictive across a wide variety of domains that contain negative competence-related stereotypes about a group of individuals engaging in a stereotype-relevant task. As

seen in Fig. 1, *Arrow 1* depicts the first part of the hypothesized mediation chain. The model suggests that negative stereotype information in the situation can lead to the adoption of PAV-goals while performing the stereotype-relevant activity (depending on whether the stereotype applies to the individual's group). PAV-goal adoption should be moderated by several individual characteristic differences. Unlike the relatively new area of stereotype-threat research, achievement goal theory has a longer history of investigation, and as such has clearly identified central individual differences in the study of achievement goals, namely an individual's achievement motivation orientation. Achievement goal research shows that in order to understand achievement goal adoption effects on performance and motivation, it is necessary to consider an individual's general reaction to achievement goals (e.g., Harackiewicz and Elliot, 1993). For example, students who are higher in achievement motivation experience a "match" between external performance-approach goals and their individual characteristic desire to do well. This individual-goal match leads to positive outcomes (e.g., more interest in the task), whereas a mismatch leads to negative outcomes (Durik and Harackiewicz, 2003). Achievement motivation, then, is an important moderator of goal effects, and identification with the domain is an important moderator of stereotype-threat effects (Aronson *et al.*, 1999; Smith and White, 2001). As such, *Arrow 2* depicts the moderating influences of such individual differences.

The mediation chain continues as the stereotype-threat-induced PAV-goal sets in motion a pattern of behavioral and phenomenological processes. Research has not yet determined the sequential nature of the process categories; therefore, *Arrows 3* represents the joint influence of PAV-goal adoption (cf. Elliott and Dweck, 1998). Beginning the discussion with behavioral processes, the STEP model predicts that depending on the adopted achievement goal, the individual will be inclined to use a certain set of self-regulation strategies (e.g., Pintrich, 2000). In addition to the previously considered stereotype-threat behavioral processes (i.e., effort, self-handicapping, study strategies), the achievement goal literature adds that PAV-goal adoption leads to minimal risk taking (e.g., Linnenbrink and Pintrich, 2000).

PAV-goal adoption also influences the individual's phenomenological experience. From the previously considered stereotype-threat mediators and other known goal correlates, this experience is likely to manifest as subjective feelings (i.e., feelings of interest, feelings of anxiety, feelings of efficacy and confidence) and physiological activity (i.e., cardiac and vascular activity, as suggested by the Blascovich *et al.*, 2001, findings reviewed earlier). In turn, the phenomenological experience influences additional behavioral interactions with the task, which then affect the phenomenological experience. *Arrows 4* represents this feedback loop.

Finally, it is hypothesized that performance outcomes are affected by the overall experience (Elliot and McGregor, 1999; Sansone and Smith, 2000). As illustrated in the STEP model, performance on the task is likely influenced most directly by the individual's phenomenological experience of the task (e.g., Deci, 1975; Schiefele, 1991), as shown by *Arrow 5*.

Although the mediation chain is hypothesized as complete at this point, *Arrow 6* is included to illustrate that various individual characteristics as mentioned above can directly moderate performance, and that performance, in turn, influences the status of those characteristics. Additionally, the individual's performance on the stereotype-relevant task is likely to bolster or undermine the individual's and others' belief in the (in)validity of the stereotype. This is shown by *Arrow 7*.

As illustrated in the STEP model, there is much overlap between the process variables hypothesized and tested in the stereotype-threat literature and the achievement goal literature. If stereotype threat leads to differential performance goal adoption, then the achievement goal literature also suggests that other variables such as interest are important to consider. If, as expected, goals are closer to performance outcomes than stereotype-threat information in the situation, then it is not surprising that past stereotype-threat–performance mediator tests were obscured.

Does the presence of a competence-related stereotype threat make salient different types of achievement performance goals? The STEP model rests on the assumption that stereotype threat leads to differential achievement goal adoption, depending on the relevancy of the stereotype to the individual. The STEP model hypothesizes that stereotype threat engenders a PAV-goal, whereas a nullified stereotype engenders a PAP-goal. These hypotheses are based on parallel performance findings from the achievement goal and the stereotype-threat literature. For example, Elliot and McGregor (1999; see also Elliot and Church, 1997) demonstrated that psychology students' performance on a multiple-choice exam and a short-essay exam was associated with both types of performance goals; PAV-goals were negatively related to test performance, whereas PAP-goals were positively related to test performance. In the stereotype-threat literature an explicit reminder of a stereotype impairs performance, whereas a reminder of the stereotype that is subsequently nullified enhances performance (e.g., Smith and White, 2002). Merging these performance results together, one possibility is that achievement goals are differentially triggered by stereotype-threat information.

As an initial test of this fundamental hypothesis, Smith (2004) reminded stereotyped (females) and nonstereotyped (males) students of the math-gender stereotype (i.e., that males are better at math than females). Students were next asked to look over a math exam (several multiple-choice

GRE items). Students then completed a 12-item achievement goal questionnaire modeled after past research (see Elliot and McGregor, 2001) before estimating their performance on the exam (similar to Stangor *et al.*, 1998).

Results showed the expected gender difference with females reporting lower math performance expectancies compared to males. Results also showed females endorsing PAV-goals to a greater extent than males. Importantly, PAV-goal endorsement mediated the relationship between participants' gender and performance expectancies. The results of Smith (2004) offer preliminary support for the hypothesis that PAV-goals are involved in the stereotype–performance relationship. Nevertheless, given the nature of the study, participants did not actually perform the math exam. Therefore, one avenue for future research is to more clearly document the relationship between stereotype threat and achievement goals.

CONVERGENT EVIDENCE FOR THE STEP MODEL

Convergent support for the notion that stereotypes lead to the adoption of performance-avoidance achievement goals comes from stereotype-threat anecdotes and research. For example, in a 2003 *Newsweek* interview with Beyonce Knowles, her comments on Black women in America reflect stereotype threat, but her statement can also be taken as a good example of the worry about performance failure that stereotyped individuals may face. The African American popular music artist was quoted as saying, "It's like you have something to prove, and you don't want to mess it up and be a negative reflection on black women" (Samuels, 2003, p. 54).

Research by Brown and Josephs (1999) also provides supportive evidence that stereotype threat triggers different achievement goals. Brown and Josephs (1999) investigated the different "concerns" students have when taking a multiple-choice math exam. In their studies, males and females were subjected to a "concern" manipulation before taking a math exam. A portion of the participants were led to believe that the math exam was designed to identify people whose math reasoning abilities are "exceptionally strong," whereas the other participants were led to believe that the math exam was designed to identify people who are "exceptionally weak" in their math reasoning abilities. Results indicated that females who thought the test would identify them as weak in their math skills performed worse than females who thought the test would identify them as having exceptionally strong math skills. The reverse results were found for males.

Stereotype threat was not directly manipulated or measured by Brown and Josephs (1999). However, the authors discussed their moderation results in terms of stereotype threat, noting that the gender math-stereotype is experienced differently for males and females. The manipulation of concern used

by Brown and Josephs' (1999) is similar to the manipulation of PAP- and PAV-goals used by Elliot and colleagues. Elliot and Harackiewicz (1996), for example, obtained evidence that giving participants a PAV-goal prior to task engagement (the puzzle task they were about to work on provides "the opportunity to demonstrate that you are not a poor puzzle solver") undermined motivation compared to giving participants a PAP-goal (the task provides "the opportunity to demonstrate that you are a good puzzle solver"). Thus, although Brown and Josephs (1999) did not make any explicit mention of or link to achievement goals, their results suggest that individuals faced with a stereotype might spontaneously adopt "concerns" or achievement goals that, in turn, influence performance outcomes.

IMPLICATIONS FOR FUTURE STEREOTYPE-THREAT MECHANISM RESEARCH

The novel hypothesis of the Stereotyped Task Engagement Process is that stereotype threat triggers the adoption of PAV-goals, preliminary support for this hypothesis was provided by Smith (2004). However, future research is needed to fully explicate the PAV-goal adoption hypothesis and the role of the process variables on performance. For instance, one important question is whether achievement goals are the only stereotype-threat–performance mediator, or if other more proximal performance mediators are also triggered by goal adoption as suggested in the STEP model. If stereotype threat does influence goal adoption, then according to goal theory it can be predicted that a stereotyped induced goal adoption next gives way to several behavioral and phenomenological processes that in turn are linked to important performance and motivational outcomes (e.g., Cury *et al.*, 2002). In fact, goal adoption is related to most of the previously considered stereotype-threat mediators. Some examples of behavioral variables identified as important to goal effects include effort and study strategy use. For instance, research on study strategy use has shown that PAV-goal adoption predicts disorganized study strategy use (measured by self-report, example item: "I am not sure how to study for this course"; Elliot *et al.*, 1999). The finding that PAV-goals are associated with disorganized study strategy use is similar to the finding that under stereotype threat, students have a hard time generating a problem-solving strategy (Quinn and Spencer, 2001).

Some examples of phenomenological variables identified as important to goal effects include the finding that endorsing PAV-goals is associated with increased anxiety (as measured by STAI) and the finding that endorsing PAV-goals is associated with decreased task involvement (as measured by self-report, example item: "While [working on the task] I was totally absorbed in the [task]" (Elliot and Harackiewicz, 1996; Elliot and McGregor, 2001). In addition, PAV-goal adoption predicts heightened feelings of

evaluation apprehension (as measured by the "Worry" scale, Elliot and McGregor, 2001). In short, PAV-goals trigger similar phenomenological experiences hypothesized to mediate the stereotype-threat performance relationship.

Although conceptually similar to the process variables already discussed, the achievement goal research has further demonstrated that individuals who adopt PAV-goals take minimal risks and report low levels of interest in the task (e.g., Elliot and McGregor, 2001; Pintrich, 2000). Individuals with PAP-goals, in contrast, take higher risks, have higher self-reported task persistence (effort), and higher levels of task interest (e.g., Elliot *et al.*, 1999; Elliot and McGregor, 1999; Linnenbrink and Pintrich, 2000). Feelings of interest and risk taking have not yet been examined in the stereotype-threat literature but may prove useful in understanding stereotype-threat effects. For instance, examining feelings of interest as a potential mediator of stereotype-threat effects on performance is an important possibility (i.e., stereotype threat leads to students not experiencing interest while engaged in the stereotype-relevant task; cf. Aronson *et al.*, 2002; Davies *et al.*, 2002). After all, one potential consequence of being the victim of a stereotype is the finding that a stereotyped individual often perceives the stereotype-relevant domain as "uninteresting" (Seymour and Hewitt, 1997), and lack of interest is often related to poor performance (e.g., Deci and Ryan, 1985). If stereotype threat influences the types of achievement goal adopted, then variables such as interest and risk taking may help to further elucidate the stereotype-threat–performance relationship.

Given the link between performance outcomes and motivation (e.g., Deci and Ryan, 1985), research examining the relationship among achievement goals, motivation, and performance in a stereotype-threat paradigm may also be important. For example, in a study by Cury *et al.* (2002), intrinsic motivation for performing a nonacademic task (basketball dribbling) was shown to differ as a function of the manipulated performance goal. Specifically, children who were given a PAV-goal dribbled the basketball less (free play is a common index of intrinsic motivation) compared to children given a PAP-goal. Multiple mediators mediated this goal-motivation relationship: competence valuation, state anxiety, and task absorption (i.e., low cognitive interference). Moreover, the mediation path was sequentially demonstrated such that focusing on PAV-goals lead to lower competence valuation, which in turn interfered with absorption in the task. Lower task absorption predicted lower levels of intrinsic motivation. Understanding the processes that achievement goals set in motion suggests that in addition to looking at goal adoption in and of itself, stereotype-threat research should consider the subsequently goal-triggered process variables (behavioral and phenomenological).

In sum, combining the previously considered stereotype-threat mediators with the findings from achievement goal theory, it is predicted that if stereotype threat induces certain performance goals, these goals, in turn, affect task behavior (effort, self-handicapping, study strategies, and risk taking) which, in turn, influence the individual's phenomenological experience while working on the task (anxiety, evaluation apprehension, efficacy/confidence, and interest), which then feeds back into the individual's task behavior in a self-regulatory manner (e.g., Pintrich, 2000). The individual's overall experience then is predicted to be the most proximal determinate of task performance (Sansone and Smith, 2000).

IN CLOSING

The role of achievement goals is important to consider within stereotype-threat research. Extant achievement goal research suggests that individuals of all ability levels interact with, experience, and perform tasks differently depending on the type of goal they adopt. Although some of the behavioral and phenomenological indexes may differ, results from the achievement goal research overlaps with much of the stereotype-threat findings. Recall that the tested stereotype-threat mediators were effort, self-handicapping, anxiety, evaluation apprehension, performance confidence, stereotype endorsement, perceptions of a test's fairness, and feelings about the self.

Although the majority of tested mechanisms in the stereotype-threat literature have not yielded convincing individual results, it is premature to dismiss them entirely. Little systematic research has examined stereotype-threat mechanisms specifically (cf. Keller and Dauenheimer, 2003; Spencer *et al.*, 1999) and thus it is unclear if the primary method of measurement (i.e., self-report) and/or the timing of measurement (i.e., postexam) are adequate to detect a working process variable. Moreover, past research examined individual mechanisms by examining the hypothesis that stereotype threat causes "X" which in turn causes poor performance. It is possible that past research has not found convincing evidence for any one mechanism because the process is more multidimensional (Aronson *et al.*, 1998) with a long-mediating chain (Judd and Kenny, 1981) as depicted by the STEP model. The STEP model is complex but perhaps more realistic and ecologically valid than the single mediator model.

The Stereotyped Task Engagement Process was presented as a multidimensional model that may prove useful in guiding methodical mechanism testing of stereotype-threat effects on performance, via the influence of performance goal adoption. This model includes both behavioral and phenomenological process variables from the stereotype-threat and

achievement goal literatures and predicts that the overall task experience predicts performance outcomes. Preliminary evidence for the first stereotype-threat–PAV goal link in the mediation chain was provided by Smith (2003). This study showed that under conditions of stereotype threat, beliefs about math performance suffers for females as a result of their self-reported performance–avoidance goal adoption. Future research is needed to fully test, clarify, and refine the STEP model.

The process by which stereotype-threat effects do their damage to performance outcomes remains unclear. This review, however, has directed future research by identifying the tested mechanisms and by specifying how they have been tested and how they might fit under a multidimensional umbrella. Hopefully, this review might stimulate researchers in other domains to consider ways that those domains might speak to stereotype-threat effects. To begin this process, I have extracted from the rich achievement goal literature to posit that a competence-threatening stereotype triggers a PAV-goal for the threatened individual even if he or she feels competent at the task. Achievement goal adoption engenders specific behavioral and phenomenological processes that ultimately lead to poorer performance. I have further presented the STEP model as a framework for merging these literatures and to illustrate the importance of bridging stereotype-threat research with other potentially related literatures.

ACKNOWLEDGMENTS

This research was supported in part by an American Education Research Association Spencer Fellowship awarded to the author. I am grateful to Dr. Carol Sansone, Dr. Paul H. White and Dr Gale Sinatra for their guidance while synthesizing this review. I would also like to thank Dr. Carol Werner, Dustin Thoman, Camile Johnson, Dr. Diane Mackie and the reviewers for their helpful comments on earlier versions of this manuscript.

REFERENCES

Ambady, N., Shih, M., Kim, A., and Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychol. Sci.* 12: 385–390.

American Psychological Society (1999). Denver, Colorado.

American Educational Research Association (1998). Chicago, Illinois.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *J. Educ. Psychol.* 84: 261–271.

Aronson, J., Fried, C. B., and Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *J. Exp. Soc. Psychol.* 38: 113–125.

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

Aronson, J., Quinn, D. M., and Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In Swim, J. K., and Stangor, C. (eds.), *Prejudice: The Target's Perspective*, Academic Press, New York, pp. 83–103.

Atkinson, J. W. (1964). *An Introduction to Motivation*, Van Nostrand, Princeton, NJ.

Atkinson, J. W., and Litwin, G. H. (1973). Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. In McClelland, D. C., and Steele, R. S. (eds.), *Human Motivation: A Book of Readings*, General Learning Press, Morristown, NJ, pp. 145–230.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*, W. H. Freeman, New York.

Barron, K. E., and Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *J. Person. Soc. Psychol.* 80: 706–720.

Begley, S. (2000, November 6). The stereotype trap. *Newsweek* pp. 66–68.

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

Betz, N. E., and Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *J. Vocat. Behav.* 23: 329–345.

Blanton, H., Crocker, J., and Miller, D. T. (2000). The effects of in-group versus out-group social comparison on self-esteem in the context of a negative stereotype. *J. Exp. Soc. Psychol.* 36: 519–530.

Blascovich, J., Spencer, S., Quinn, D., and Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychol. Sci.* 12: 225–229.

Bong, M., and Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they? *Educ. Psychol. Rev.* 15: 1–40.

Brown, R. P., and Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *J. Person. Soc. Psychol.* 76: 246–257.

Cacioppo, J. T., and Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *Am. Psychol.* 45: 16–28

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., and Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *Eur. J. Soc. Psychol.* 33: 267–285.

Croizet, J. C., and Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Person. Soc. Psychol. Bull.* 24: 588–594.

Cury, F., Elliot, A., Sarrazin, P., Fonseca, D. D., and Rufo, M. (2002). The trichotomous achievement goal model and intrinsic motivation: A sequential mediational analysis. *J. Exp. Soc. Psychol.* 38: 473–481.

Davies, P. G., Spencer, S. J., Quinn, D. M., and Gerhardstein, R. (2002). Consuming Images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Person. Soc. Psychol. Bull.* 28: 1615–1628.

Deci, E. L. (1975). *Intrinsic Motivation*, Plenum Press, New York.

Deci, E. L., and Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*, Plenum Press, New York.

Durik, A. M., and Harackiewicz, J. M. (2003). Achievement goals and intrinsic motivation: Coherence, concordance, and achievement orientation. *J. Exp. Soc. Psych.* 39: 378–385.

Dweck, C. S., and Leggett, E. L. (1988). A social–cognitive approach to motivation and personality. *Psychol. Rev.* 95: 256–273.

Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychol. Women Q.* 18: 585–609.

Eccles, J. S., and Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annu. Rev. Psychol.* 53: 109–132.

Elliot, A. J., and Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *J. Person. Soc. Psychol.* 72: 218–232.

Elliot, A. J., and Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *J. Person. Soc. Psychol.* 70: 461–475.
Elliot, A. J., and McGregor, H. A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *J. Person. Soc. Psychol.* 76: 628–644.
Elliot, A. J., and McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *J. Person. Soc. Psychol.* 80: 501–519.
Elliot, A. J., McGregor, H. A., and Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *J. Educ. Psychol.* 91: 549–563.
Elliot, A. J., and Thrash, T. M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educ. Psychol. Rev.* 13: 139–156.
Eliott, E., and Dweck, C. (1998). Goals: An approach to motivation and achievement. *J. Person. Soc. Psychol.* 54: 5–12.
Gendolla, G. H. E. (1998). Effort as assessed by motivational arousal in identify-relevant tasks. *Basic Appl. Soc. Psychol.* 20: 111–121.
Gonzales, P. M., Blanton, H., and Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Person. Soc. Psychol. Bull.* 28: 659–670.
Harackiewicz, J. M., and Elliot, A. J. (1993). Achievement goals and intrinsic motivation. *J. Person. Soc. Psychol.* 65: 904–915.
Harackiewicz, J. M., and Sansone, C. (2000). Rewarding competence: The importance of goals in the study of intrinsic motivation. In Sansone, C., and Harackiewicz, J. M. (eds.), *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, Academic Press, San Diego, CA, pp. 82–103.
Hausdorff, J. M., Levy, B. R., and Wei, J. Y. (1999). The power of ageism on physical function of older persons: Reversibility of age-related gait changes. *J. Am. Geriatr. Soc.* 47: 1346–1349.
Hebl, M. R., Foster, J. B., Mannix, L. M., and Dovidio, J. F. (2002). Formal and interpersonal discrimination: A field study of bias toward homosexual applicants. *Person. Soc. Psychol. Bull.* 28: 815–825.
Heatherton, T. F., and Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *J. Person. Soc. Psychol.* 60: 895–910.
Hirt, E. R., Deppe, R. K., and Gordon, L. J. (1991). Self-reported versus behavioral self-handicapping: Empirical evidence for a theoretical distinction. *J. Person. Soc. Psychol.* 61: 981–991.
Hyde, J. S., and Kling, K. C. (2001). Women, motivation, and achievement. *Psychol. Women Q.* 25: 364–378.
Inzlicht, M., and Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem solving deficits in the presence of males. *Psychol. Sci.* 11: 365–371.
Jacobs, J. E., and Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *J. Person. Soc. Psychol.* 63: 932–944.
James, K., and Greenberg, J. (1997). Beliefs about self and about gender groups: Interactive effects on the spatial performance of women. *Basic Appl. Soc. Psychol.* 19: 411–425.
Jones, E. E., and Berglas, S. (1978). Control of attributions about the self through self-handicapping strategies: The appeal of alcohol and the role of underachievement. *Person. Soc. Psychol. Bull.* 4: 200–206
Josephs, R. A., Newman, M. L., Brown, R. P., and Beer, J. M. (2003). Status, testosterone, and human intellectual performance: Stereotype threat as status concern. *Psychol. Sci.* 14: 158–163.
Judd, C. M., and Kenny, D. A. (1981). *Estimating the Effects of Social Interventions*, Cambridge University Press, New York.
Jussim, L., Palumbo, P., Chatman, C., Madon, S., and Smith, A. (2000). Stigma and self-fulfilling prophecies. In Heatherton, T. F., Kleck, R. E., Hebl, M. R., and Hull, J. G. (eds.), *The Social Psychology of Stigma*, The Guilford Press, New York, pp. 374–418.
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., and Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting effect on women's math performance. *Person. Soc. Psychol. Bull.* 29: 371–381.

Kray, L. J., Thompson, A., and Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *J. Person. Soc. Psychol.* 80: 942–958.

Larkin, K. T., Ciano-Federoff, L. M., and Hammel, D. (1998). Effects of gender of observer and fear of negative evaluation on cardiovascular reactivity to mental stress in college men. *Int. J. Psychophys.* 29: 311–318.

Leary, M. R., and Shepperd, J. A. (1986). Behavioral self-handicaps versus self-reported handicaps: A conceptual note. *J. Person. Soc. Psychol.* 51: 1265–1268.

Levy, B., and Langer, E. (1994). Aging free from negative stereotypes: Successful memory in China and among the American deaf. *J. Person. Soc. Psychol.* 66: 989–997.

Leyens, J. P., Desert, M., Croizet, J. C., and Darcis, C. (2000). Stereotype threat: Are lower status and history of stigmatization preconditions of stereotype threat? *Person. Soc. Psychol. Bull.* 26: 1189–1199.

Linnenbrink, E. A., and Pintrich, P. R. (2000). Multiple pathways to learning and achievement: The role of goal orientation in fostering adaptive motivation, affect, and cognition. In Sansone, C., and Harackiewicz, J. M. (eds.), *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, Academic Press, San Diego, CA, pp. 195–227.

Luhtanen, R., and Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Person. Soc. Psychol. Bull.* 18: 302–318.

Maehr, M. L., and Midgley, C. (1996). *Transforming School Cultures*, Westview Press, Boulder, CO.

McIntyre, R. B., Paulson, R. M., and Lord, C. G. (2003). Alleviating women's mathematics stereotype threat though salience of group achievements. *J. Exp. Soc. Psychol.* 39: 83–90.

McKown, C., and Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child Dev.* 74(2): 498–515.

Molden, D. C., and Dweck, C. S. (2000). Meaning and motivation. In Sansone, C., and Harackiewicz, J. M. (eds.), *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*, Academic Press, San Diego, CA, pp. 131–159.

Mullen, B. (1986). Stuttering, audience size, and the other-total ratio: A self-attention perspective. *Person. Soc. Psychol. Bull.* 12: 187–197.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychol. Rev.* 91: 328–346.

Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychol. Rev.* 84: 231–258.

O'brien, L. T., and Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Person. Soc. Psychol. Bull.* 29: 782–789.

Osborne, J. W. (2001). Testing stereotype threat effects: Does anxiety explain race and sex differences in achievement? *Contemp. Educ. Psychol.* 26: 291–310.

Oswald, D. L., and Harvey, R. D. (2001). Hostile environments, stereotype threat, and math performance among undergraduate women. *Curr. Psychol.* 19(4): 338–356.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In Boekaerts, M., Pintrich, P. R., and Zeidner, M. (eds.), *Handbook of Self-Regulation*, Academic Press, San Diego, CA, pp. 451–502.

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

Rahhal, T. A., Hasher, L., and Colcombe, S. J. (2001). Instructional manipulations and age differences in memory: Now you see them, now you don't. *Psychol. Aging* 16: 697–706.

Rawsthorne, R., and Elliot, A. J. (1999). Achievement goals and intrinsic motivation: A meta-analytical review. *Personal. Soc. Psychol. Rev.* 3: 326–344.

Rhodewalt, F., and Hill, S. K. (1995). Self-handicapping in the classroom: The effects of claimed self-handicaps on responses to academic failure. *Basic Appl. Soc. Psychol.* 16: 397–416.

Rosenthal, R., and Jacobson, L. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychol. Rep.* 19: 115–118.

Samuels, A. (2003, March 3). Time to tell it like it is. *Newsweek* pp. 52–55.
Sansone, C., and Harackiewicz, J. M. (1996). "I don't feel like it": The function of interest in self-regulation. In Martin, L. L., and Tesser, A. (eds.), *Striving and Feeling: Interactions Among Goals, Affect, and Self-Regulation*, Erlbaum, Mahwah, NJ, pp. 203–228.
Sansone, C., and Smith, J. L. (2000). Interest and self-regulation: The relation between having to and wanting to. In Sansone, C., and Harackiewicz, J. M. (eds.), *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and performance*, Academic Press, San Diego, CA, pp. 341–372.
Sarason, I. G. (1972). Experimental approaches to test anxiety: Attention and the uses of information. In Spielberger, C. D. (ed.), *Anxiety: Current Trends in Theory and Research Vol. 2*, Academic Press, New York, pp. 381–403.
Schiefele, U. (1991). Interest, learning, and motivation. *J. Educ. Psychol.* 26: 299–323.
Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *J. Exp. Soc. Psychol.* 38: 194–201.
Sekaquaptewa, D., and Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *J Exp. Soc. Psychol.* 39: 68–74.
Seymour, E., and Hewitt, N. M. (1997). *Talking About Leaving: Why Undergraduates Leave the Sciences*, Westview Press, Boulder, CO.
Shih, M., Pittinsky, T. L., and Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychol. Sci.* 10: 80–83.
Smith, J. L. (2004). *The Interplay Among Stereotype Threat, Performance–Avoidance Goals, and Math Performance Expectancies*. Manuscript submitted for publication.
Smith, J. L., and White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educ. Psychol. Measure.* 61: 1040–1057.
Smith, J. L., and 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.
Snyder, M., Tanke, E. D., and Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *J. Person. Soc. Psychol.* 35: 656–666.
Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *J. Exp. Soc. Psychol.* 35: 4–28.
Spielberger, C. D., Gorsuch, R. L., and Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*, Consulting Psychologist Press, Palo Alto, CA.
Stangor, C., Carr, C., and Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *J. Personal. Soc. Psychol.* 75: 1191–1197.
Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *Am. Psychol.* 52: 613–629.
Steele, C. M. (1999, August). Thin ice: Stereotype threat and Black college students. *Atlantic Monthly* pp. 44–54.
Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *J. Person. Soc. Psychol.* 69: 797–811.
Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in White athletes. *Person. Soc. Psychol. Bull.* 28: 1667–1678.
Stone, J., Lynch, C. I., Sjomeling, M., and Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *J. Person. Soc. Psychol.* 77: 1213–1227.
Tennyson, R. D., and Woolley, F. R. (1971). Interaction of anxiety with performance on two levels of task difficulty. *J. Educ. Psychol.* 62: 463–467.
Urdan, T., and Midgley, C. (2001). Academic self-handicapping: What we know, what more there is to learn. *Educ. Psychol. Rev.* 13: 115–138.
Walsh, M., Hickey, C., and Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles* 41: 219–240.
Weiner, B. (1992). *Human Motivation: Metaphors, Theories, and Research*, Sage, Newbury Park, CA.
Wheeler, S. C., and Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychol. Bull.* 127: 797–826.
Wright, R., Contrada, R., and Patane, M. J. (1986). Task difficulty, cardiovascular response, and the magnitude of goal valence. *J. Person. Soc. Psychol.* 51: 837–843.