

Gender Differences in Mathematics

An Integrative Psychological Approach

Edited by

ANN M. GALLAGHER

Law School Admission Council

JAMES C. KAUFMAN

California State University, San Bernadino

2005



CAMBRIDGE
UNIVERSITY PRESS

"Math is hard!" (Barbie™, 1994)

Responses of Threat vs. Challenge-Mediated Arousal to Stereotypes Alleging Intellectual Inferiority

Talia Ben-Zeev, Cristina M. Carrasquillo, Alison M. L. Ching, Tattiya J. Kliengklom, Kristen L. McDonald, Daniel C. Newhall, Gillian E. Patton, Tiffany D. Stewart, Tonya Stoddard, Michael Inzlicht, and Steven Fein

In 1994, Mattel created a Barbie™ doll that said, "Math is hard." *The Barbie Liberation Organization*, a group composed of activists and media personalities, among others, protested against Barbie's perpetuation of gender-based stereotyping. The media publicized the case and discussions on gender stereotyping in children's toys ensued on and off the air, leading Mattel to withdraw the "math is hard" Barbie from the market.

However, did Barbie's frustration with math represent a reality in which girls and women, more than boys and men, find math to be hard? Benbow and Stanley (1980, 1983) found gender differences in performance on the mathematical section of the SAT (SAT-M) in boys and girls under the age of fourteen who were high in math achievement. The boys outperformed the girls by about half a standard deviation and were over-represented by a ratio of 13:1 among students who scored above 700. Similarly, in a meta-analysis involving over three million participants, Hyde, Fennema, and Lamon (1990) found a gender difference favoring males that emerged from high school ($d = 0.29$) through college ($d = 0.41$), and into adulthood ($d = 0.59$). Finally, Brown and Josephs (1999) reported that the two most widely used standardized tests of mathematics in the United States, the SAT-M and the quantitative portion of the GRE (GRE-Q), revealed a gender difference in the order of half a standard deviation.

This gender difference can also be seen in the types of activities that females vs. males tend to pursue. Young girls are less likely than boys to be interested in playing with scientific toys, participating in mathematical games, and reading mathematical books (Eccles & Jacobs, 1987). In advanced high school math courses, males outnumber females by a margin of 2:1 (see Geary, 1996, for a review). As far as higher education is

concerned, in a 20-year follow up to their research on mathematically gifted students, Benbow and her colleagues (Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000) found that males were about twice as likely as their female counterparts to attain a bachelor's degree in math or in the physical sciences, and twice as likely to gain employment in such fields.

It is disheartening that even females with strong math backgrounds and interests tend to underperform relative to equally prepared males in the math domain. In this chapter, we explore how *stereotype threat* – a situational phenomenon that occurs when individuals who are targets of stereotypes alleging intellectual inferiority are reminded of the possibility of confirming these stereotypes – may be an explanatory framework for understanding why females and other stigmatized students succumb to this underperformance (e.g., Aronson, Lustina, Good, Keough, Steele, & Brown, 1999; Aronson, Quinn, & Spencer, 1998; Shih, Pittinsky, & Ambady, 1999; Spencer, Steele, & Quinn, 1999; Steele, 1997; Steele & Aronson, 1995).

STEREOTYPE THREAT

The social context has been strongly implicated in creating, perpetuating, or eliminating the underperformance that has hindered stigmatized groups, such as females in math. In particular, stereotype threat has been hypothesized to occur as a result of a distress that an individual feels when she faces the possibility of confirming a negative stereotype about her group. It is the fear that her behaviors may substantiate disparaging stereotypes (Aronson et al., 1999; Aronson et al., 1998; Spencer et al., 1999; Steele 1997; Steele & Aronson, 1995). This situational distress can then hinder intellectual performance. For example, Spencer et al. (1999) showed that female college students performed significantly worse than males on a standardized math test when the stereotype about their math ability was made relevant (by informing the students that males performed better than females on this test in the past), but that this gender gap was eliminated simply by changing the wording used for introducing the test so the stereotype would seem to be inapplicable in this situation (by informing the students that males and females have performed equally well on this test in the past).

A counterintuitive finding is that only individuals who are *highly identified* with success and achievement in given stereotyped domains are the ones who show performance deficits under threat (Steele, 1997). Thus, it is the people in the vanguard of their group who are the most vulnerable to situations in which stereotypes become salient.

Spencer et al. (1999) had highly math-identified males and females take a difficult math test. Half was told that the math test had shown sex differences in the past, whereas the other half was told that it there had never been sex-differences on the test. Results showed that females and males

did equally well on the test in the latter, stereotype removed condition, whereas in the former condition in which the stereotype was made relevant, females performed significantly worse than males.

Steele and Aronson (1995) examined the effects of negative stereotypes about intelligence on the academic performance of African Americans. African American and white undergraduates took a difficult test of verbal ability. Half was told that the test was diagnostic of intelligence, whereas the other half was told that the test was nondiagnostic of intelligence. Although every participant took the same test, African American participants performed as well as white participants in the nondiagnostic condition, but performed more poorly than whites in the diagnostic condition.

Stereotype threat can be triggered in even subtler and yet frequently experienced ways, such as by the composition of the individuals in the immediate setting. Inzlicht and Ben-Zeev (2000) found that females' math performance could be undermined significantly simply by being outnumbered by males in the room as they take a math test. Inzlicht and Ben-Zeev demonstrated that undergraduate females who took portions of a GRE-math test with two males in the room performed more poorly than females who took the same test with two other females present. Furthermore, females' deficits were proportional to the number of males in their environment. Even females who were placed in a testing environment with more females than males experienced moderate but significant problem-solving deficits. (For a discussion of the properties of an intellectually threatening environment, see Inzlicht, 2001.)

Does the explicit or implicit reminder of a stereotype cause a desire to disconfirm it? Steele and Aronson (1995) had African American and white participants take a verbal test that was described as either diagnostic or nondiagnostic of intelligence. Before taking the test, however, participants were asked to complete measures of stereotype activation and stereotype avoidance. The stereotype activation measure consisted of eighty word fragments, eleven of which could be completed with, among other words, words associated with African American stereotypes (e.g., - - C E [RACE], B R - - - - [BROTHER], or W E L - - - - [WELFARE]). On the stereotype avoidance measure, participants were asked to rate their preferences on various things, including music and sports, some of which were associated with African American life or culture (e.g., hip-hop music or basketball). If describing a test as diagnostic of intelligence introduces stereotype threat, then racial stereotypes should be activated in the minds of African American participants. Moreover, African American participants should be motivated to disconfirm them.

Results showed that after controlling for self-reported SAT scores, African American participants in the diagnostic condition resolved more word fragments with stereotype-related words and exhibited lower

preferences for things related to African American culture than African Americans in the nondiagnostic condition. White participants' responses did not vary across test description.

Stereotype threat generalizes to any individual belonging to an ability-stereotyped group. Croizet and Claire (1998), for example, found that students of low socioeconomic status (SES) attained lower accuracy, raw, and adjusted scores than high SES students when a GRE verbal test was framed as a test of intelligence. When the test was framed as a test of attention, however, low SES students performed just as well as high SES students. Similarly, Aronson and Salinas (1997) found that highly identified Latino students received lower scores than whites on a GRE verbal test, but only when the stereotype about Latinos' alleged lower intelligence was made salient.

Stereotype threat is a situational predicament that can affect a member of any group targeted by specific negative stereotypes – disadvantaged minority group or otherwise (Aronson et al., 1999). Thus, even advantaged white males may be susceptible to stereotype threat effects. Aronson et al. tested this hypothesis by giving highly math-identified white males a difficult math test. Half was told that the purpose was to gauge their math ability, whereas the other half was told that the purpose was to better understand why Asians outperformed whites in math tests. As in studies looking at other groups, the participants confronted with the white inferiority stereotype performed worse than other participants. Similarly, Stone, Lynch, Sjomeling, and Darley (1999) found that white participants underperformed on a golf task when they were reminded of the African American athletic superiority stereotype. Likewise, Leyens, Desert, Croizet, and Darcis (2000) found that males made more commission mistakes on an affective processing task when they were prompted with the stereotype that men are less sensitive to emotions than women. Negative stereotypes, then, appear to undermine performance of individuals belonging to any group about which negative stereotypes exist, even if the group is not generally disadvantaged in society.

Despite the enthusiasm that stereotype threat research has triggered, little is known about the specific mechanisms that underlie the effects of stereotype threat. Identifying these mechanisms would be informative with respect to theories of how stereotypes about problem-solving ability affect the intellectual processing of stigmatized high-achieving individuals and to educational practice.

POSSIBLE MEDIATORS OF STEREOTYPE THREAT

Steele and Aronson (1995) offered several mediators for the phenomenon of stereotype threat (also see Baumeister & Showers, 1986). The first is distraction due to evaluation apprehension. Instead of focusing on the

task, a stigmatized person may become concerned with others' judgment of his or her ability. A second mediator is lowered self-efficacy, which in turn may moderate performance expectations and effort. Stangor, Carr, and Kiang (1998), for example, suggested that the activation of negative stereotypes lowers task performance expectations, which then undermine initial task confidence. Lowered expectations can undermine motivation and effort (Pyszczynski & Greenberg, 1983), which may then result in lower performance. Hence, negative stereotypes may lead to lower performance because they may undermine confidence and self-efficacy, which, in turn, can lead to a withdrawal of effort.

Empirical research, however, has found little evidence in support of the distraction or expectation mechanisms. For example, in an ambitious study, Spencer et al. (study 3; 1999) examined the potential mediation of both distraction (performance evaluation) and expectation (confidence and self-efficacy). As in previous studies, highly math-identified men and women were instructed to take a math test that was either described as showing gender differences in the past (threat condition) or not (no threat condition). Once all the participants read the test instructions, they filled out a questionnaire assessing evaluation apprehension (e.g., "people will think I have less ability if I do not do well on this test") and self-efficacy (e.g., "I am concerned about whether I have enough mathematical ability to do well on the test"). After filling out this questionnaire, all participants were asked to complete a difficult math test.

Results replicated the basic stereotype threat effect. Women attained lower scores than men in the threat condition, but performed equally to men in the no threat condition. Again, when stereotypes were relevant to performance, women underperformed. Mediation of this effect was tested through a series of mediational analyses for each potential mechanism (Baron & Kenny, 1986), but neither self-efficacy nor evaluation apprehension mediated stereotype threat. Self-efficacy was both unrelated to the stereotype threat manipulation and to test performance, and evaluation apprehension did not appear to mediate the relationship between stereotype threat and performance. Although evaluation apprehension was related to test performance, it was not related to the stereotype threat manipulation, nor did it reduce the direct relationship between the manipulation and performance (Spencer et al., 1999).

Aronson and his colleagues (1999) also investigated the cause of threat, but were unable to reach any definite conclusions. As in previous studies, they found that highly identified white students who were faced with the Asian math superiority stereotype attained fewer number correct than white students who were not made aware of the stereotype. The stereotyped participants also wondered more often what the experimenter would think of them in the threat condition than in the control condition. That is, threatened participants experienced evaluative worries that may have

distracted their performance. However, when this evaluation apprehension was partialled out, the relationship between threat and performance was unchanged. Therefore, worries about being negatively evaluated did not seem to cause threat-induced performance deficits.

Finally, Brown and Steele (1999, as reported in Marx, Brown, & Steele, 1999) directly explored the role of performance expectations. As in previous experiments, white and African American undergraduates took a difficult GRE test of verbal ability. Half of the participants were informed that the test was diagnostic of ability (threat condition), whereas the other half were told that it was diagnostic, but racially fair (no threat condition). Prior to taking the main GRE test, participants took either an easy or difficult verbal task that resulted in complete success or clear failure. The idea was that performance on this task would set either high or low performance expectations for the second task. Brown and Steele, however, only assumed that performance expectations were manipulated. That is, they assumed that prior success or failure without feedback was capable of manipulating expectations, but did not explicitly measure them. Nonetheless, results demonstrated that African American participants performed worse than their white peers, even when they presumably had high performance expectations. Because high performance expectations could not overcome the hindering qualities of racial stereotypes, the authors concluded that stereotype threat was not mediated by expectations.

A factor that has not been examined sufficiently thus far but that has the potential to be an important mediator is physiological arousal (e.g., Steele & Aronson, 1995).

AROUSAL AS A SUGGESTED MEDIATOR OF STEREOTYPE THREAT

According to the classic Yerkes and Dodson (1908) law of physiological arousal, performance is optimal at intermediate levels of arousal and decreases when arousal is either low or high, resulting in an inverted U-shaped function. The data that were the basis for this finding came from a study of animal behavior. Yerkes and Dodson had white mice perform a learning task of finding compartments. During learning, the mice were given electric shocks of low, medium, or high intensity. The results showed that medium-intensity shocks were the most efficient in promoting learning as compared with low- and high-intensity shocks. This finding implies that cognitive processes, such as learning and memory, are executed most efficiently under conditions of medium than of low or high arousal.

More recent neurophysiological research has corroborated and expanded on the classic findings of the Yerkes-Dodson law in animals and in humans. For example, Lupien and McEwen (1997) provided evidence for an inverted U-shaped relationship between level of corticosteroids

(hormones that get released as a result of high arousal) and cognitive processes in a variety of animal and human studies.

We hypothesize that high-achieving individuals in stigmatized domains approach a problem-solving task in the given domains with an optimal level of arousal for performing well (a medium level of arousal conducive to performance). An exposure to a situation of stereotype threat may interfere with these individuals' performance by leading to arousal that exceeds an optimal level, causing performance deficits (also see O'Brien, Crandall, 2001).

Mathematical problem solving relies heavily on controlled (vs. automatic) processing, which operates serially, requires attention, and is capacity (resource) limited (e.g., Shiffrin & Schneider, 1977). Increased levels of arousal have been shown to interfere with the ability to allocate mental resources during controlled processing (e.g., Kahneman, 1973). Therefore, if the activation of stereotypes results in levels of arousal that exceed the intermediate levels necessary for optimal performance, then the controlled processing required for successful problem solving will be hindered.

What cognitive factors may mediate heightened arousal? The literature points to the cognitive appraisals of threat vs. challenge.

EFFECTS OF COGNITIVE APPRAISAL, THREAT VS. CHALLENGE, ON AROUSAL AND PERFORMANCE

The interpretation of an event as either posing a potential threat to the well-being of a person or as a challenge that can lead to a positive growth experience has differential effects on physiological arousal and, in turn, on intellectual performance. When an environment is perceived as a threatening one, animals and humans show an increase in pituitary-adrenal-cortical arousal, in which the pituitary gland releases adrenocorticotropin (ACTH), which in turn, stimulates the release of the glucocorticoid, cortisol. When an environment is perceived as a challenging one, the hypothalamus, acting through the sympathetic nervous system, tends to stimulate adrenaline release from the medullae (Dienstbier, 1989).

Increased arousal is associated with cognitive and behavioral changes. For example, Ursin, Baade, & Levine (1978) found that Norwegian Army paratroopers with increased adrenaline levels had better success in jumps from airplanes than did paratroopers with increased cortisol levels. The latter also tended to show an increased defensiveness. In animal studies, using a range of subjects, from rats to primates, exposure to stressful stimuli that the animal has no control over (e.g., separating an infant from its mother) tends to result in elevation of cortisol.

The main construct that emerges, psychologically, is that of having perceived resources available to cope with situational demands. The situation becomes challenging when there is a perception of adequate resources,

such as sufficient skills or a chance to practice the task at hand. In contrast, situations are appraised as threatening when the resources are perceived to be insufficient (Folkman & Lazarus, 1986). Therefore, situational beliefs about intellectual resources to perform well enough on a math test could play an important role in whether the test is interpreted as a challenge or as a threat.

OUR WORKING STEREOTYPE THREAT MODEL

Our working model of stereotype threat is based on cognitively mediated arousal as the main causal factor (Fig. 9.1).

Spencer et al. (1999) showed that, although generalized arousal (as measured explicitly by asking participants to rate their degree of nervousness and anxiety) was related to both performance and threat, arousal did not significantly reduce the direct relationship between threat and test performance. Nonetheless, these results do not rule out arousal as a mediator of stereotype threat. As Spencer et al. noted, the number of participants in their study was too small to perform an adequate factor analysis. In addition, because the measures of arousal were overt and self-reported, participants' ratings of arousal may have been unreliable or affected by demand characteristics.

To avoid the pitfalls of self-report, correlational analysis, and some of the ecological validity concerns related to directly measuring arousal, we carried out a study to examine the role of arousal using indirect measures (Ben-Zeev, Inzlicht, & Fein, in press). In Experiment 1, high math-identified female college students were divided randomly into two conditions. The experimental group was placed under stereotype threat condition

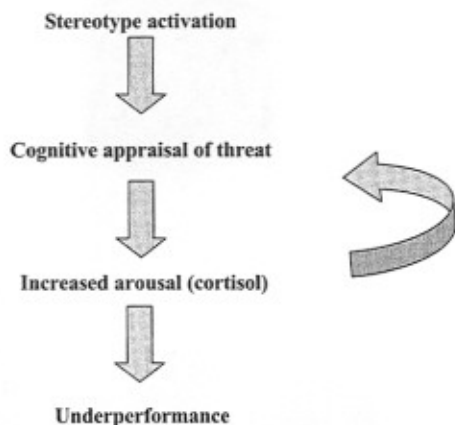


FIGURE 9.1. Performance deficits mediated by threat appraisal and heightened cortisol.

(i.e., they were told that in the past, females have performed more poorly on the math test than males). Participants in the no threat condition were told that males and females have performed equally well on this test. All participants then completed a low-arousal easy task (typing their name forward as many times as they could in a given time frame) and a high-arousal difficult task (typing their name backward as many times as they could in a given time frame) (following the procedure used by Schmitt, Gilovitch, Goore, & Joseph, 1986).

We predicted that if arousal plays an important role in stereotype threat, then the individuals subjected to stereotype threat should do *better* on the easy task but *worse* on the difficult task in comparison to the individuals in the no threat condition. This prediction follows from the arousal and social facilitation literature, which shows that individuals who experience high levels of arousal do better on easy than on difficult tasks (Schmitt et al., 1986; Zajonc, 1965). The data confirmed this prediction. Females in the threat group performed significantly better on the easy ($M = 8.72$) than on the hard task ($M = 5.2$), whereas females in the nothreat condition performed significantly better on the difficult ($M = 8.2$) than on the hard task ($M = 6.2$). As predicted, the analysis of variance (ANOVA) yielded no significant main effects (both $F_s < 1$), but the interaction was significant, $F(1, 15) = 5.96, p < 0.03$.

A second experiment was done to examine whether arousal is implicated in stereotype threat by using a misattribution paradigm. The stereotype threat manipulation consisted of assigning female college students randomly to either same-sex (one female student with two female confederates) or minority (one female student with two male confederates) conditions. We then carried out an indirect assessment of arousal by using a misattribution manipulation. As part of a general cover story about examining performance on standardized tests, participants were told that one of the factors being studied was the effects of subliminal noise on test performance. All participants were seated in front of a large machine that was introduced as the subliminal noise generator. To illustrate how the audio generator worked, participants were exposed to a series of audible tones that increased in frequency, culminating in a subliminal tone of 20,000 Hz – a "silent tone" well beyond the range of human hearing. Participants were then informed that they would be exposed to this tone for the duration of the math test. Participants in the *control* group were told that the subliminal noise would have no discernible physical effects on them. In contrast, those in the *misattribution* condition were told that the noise was associated with a number of side effects, and that previous participants had noted an increase in arousal, nervousness, and heart rate. They were told not to be alarmed if they felt these side effects and were assured that the any such side effects would be temporary. Next, all participants were given a difficult math test with GRE items to solve.

If the environment leads to performance deficits through the process of arousal, then giving stereotype threatened female students the chance to misattribute their arousal to the subliminal noise should have spared them the performance deficits triggered by stereotype threat. The performance deficit for females in the minority (threat) condition should be attenuated for females in the misattribution condition. The results of the pilot study supported the arousal prediction. A 2×2 analysis of covariance (ANCOVA) was performed on the number of math problems answered correctly, using SAT scores as a covariate, and revealed a significant interaction between the manipulations of sex composition and misattribution, $F(1, 31) = 6.06, p < 0.02$. Simple effects analyses revealed that among women not given the misattribution information, the typical stereotype threat effects emerged. Females in the minority group did significantly worse ($M = 6.34, SD = 2.43$) than did females in the same-sex group ($M = 8.82, SD = 2.40$), $F(1, 31) = 5.44, p < 0.03$. This constitutes a large effect size, Cohen's $d = 1.03$. In contrast, there were no significant differences between minority participants ($M = 7.97, SD = 2.33$) and same-sex participants ($M = 6.61, SD = 2.43$) in the misattribution condition, $F(1, 31) = 1.44, ns$, and no other effects approached significance.

The results of this study are consistent with the hypothesis that the activation of stereotype threat may cause arousal to exceed an optimal level, resulting in performance deficits (i.e., controls in the minority condition performed worse than controls in the same-sex condition). However, performance deficits were eliminated when high-achieving females in the threat condition were given an opportunity to misattribute their elevated state of arousal to a benign external source. Females in the misattribution condition who were in the minority performed as well as controls in the same-sex condition. Thus, it appears that the misattribution manipulation may have reduced arousal closer to optimal levels.

Blascovich, Spencer, Quinn, and Steele (2001) found that, when African Americans experienced stereotype threat on a cognitive task, their blood pressure rose faster and reached a higher level than it did in Caucasians. This heightened blood pressure continued even during a five-minute rest period and during an additional task. When stereotype threat was reduced, however, this effect was not apparent. These results, together with the arousal studies by Ben-Zeev et al. (in press), support the view that physiological arousal has a central role in stereotype threat.

We are currently in the process of examining arousal levels by assessing whether individuals under threat secrete higher levels of the corticosteroid cortisol, which has been implicated in a variety of cognitive and health deficits (e.g., Dienstbier, 1989). These findings would imply that intellectual and health risks might become associated with frequent experience of stereotype threat.

We postulate that in order to uncover how stereotype threat is mediated, we must examine the physiological arousal associated with threat as the cause of performance deficits in women's math performance. If a stress response (i.e., cortisol) inhibits cognitive performance, could it be that the physiological response of stress an individual feels under stereotype threat conditions is the mechanism that causes their underperformance? It is our hypothesis that when under stereotype threat, high math-identifying females will perform more poorly than when they are not under threat as a result of this physiological response. A heightened state of arousal due to the stress of possibly confirming the negative stereotype about their group will lead to physiological effects of stress on cognition, namely, subpar performance. Furthermore, women who do not identify strongly with math and/or math-oriented domains (e.g., "low identifiers") will initially exhibit low arousal due to their initial state of apathy toward taking a math test and will subsequently perform better on a difficult math test when under stereotype threat due to an increase in arousal that positively correlates with cognitive functioning. The mediation of physiological arousal, we conjecture, is based on the cognitive appraisals of challenge vs. threat. As can be seen in Fig. 9.1 (as well as in Fig. 9.2), the hypothesized relation between appraisals and arousal is bidirectional. That is, an appraisal can cause a physiological reaction, which, in turn, can affect further appraisals.

PERFORMANCE DEFICITS AND BOOSTS INTERPRETED THROUGH THE LENS OF THREAT VS. CHALLENGE AROUSAL

Traditionally, stereotype threat work has focused on how the activation of stereotypes alleging inferior abilities can cause performance deficits in the given domain. However, it appears that the social context can prime particular aspects of one's self-identity, which, in turn, may hinder or *facilitate* intellectual performance. A particularly compelling illustration of this reality comes from the work of Shih et al. (1999), who examined the effects of stereotype activation on Asian American females' math performance. In the experimental conditions, participants were primed with being Asian American or were primed with being female before taking a difficult math test. The priming was conducted by using pretest questionnaires, which asked participants to answer questions regarding their ethnicity or their gender. A control group of Asian American females was asked to take the math test without completing any demographic questionnaire. Shih et al. found that participants in the Asian American identity priming condition scored significantly higher on the standardized math test than controls, whereas participants in the female identity priming condition scored lower than controls. These results are striking, given that a

subtle priming of one aspect of identity (ethnicity) vs. another (gender) was sufficient for either enhancing or undermining performance.

These results conform to our cognitive appraisal framework. More specifically, we hypothesize that the Asian American females who were primed as females engaged in a threat appraisal, which led them to evaluate themselves as having insufficient resources (i.e., math ability) to successfully complete the math test. In contrast, the Asian American females in the Asian priming condition engaged in a challenge appraisal, which helped them to trust that they had the ability to complete the task successfully despite acknowledging that the task was difficult.

Another effect is the tendency of males to perform better on a math test when they are reminded of the stereotype regarding female inferiority in math or to "worsen" in performance when the stereotype threat condition is removed. Our challenge vs. threat appraisal framework is also useful for examining this trend over different studies. Spencer et al. (1999) were the first to report males' tendency to perform better in math under female threat conditions than under nonthreat conditions. Spencer et al. speculated that the (nonsignificantly) enhanced performance across multiple studies could be explained in one of two ways. First, males exhibit their baseline math performance under female threat conditions because stereotypes alleging math inferiority for females are present in most real world environments. Thus, it is the nonthreat situation that hinders males' performance. Second, males perform at baseline in the reduced stereotype or no threat condition, whereas the introduction of the threat gives men a "boost" because men are reminded indirectly of the stereotype that they are superior to females in math. Whatever the explanation may be, the trend for males to enhance their math performance under threat can be seen across multiple studies (e.g., Inzlicht & Ben-Zeev, 2000; Walsh, Hickey, & Duffy, 1999).

If our arousal model is correct, then high math-identified males and females experience near peak levels of arousal before being asked to take a math test. Under the female threat conditions, females would be more likely to engage in threat appraisal followed by elevated cortisol levels leading to underperformance, whereas males under the female threat condition may engage in a challenge response that would boost their performance. We expect the magnitude of this "boost" to be small because males would already be performing near ceiling.

The challenge that a nonstigmatized person may feel under threat to a stigmatized person may be captured by the following model (see Fig. 9.2):

The male boost in performance that occurs under female threat is not limited to math domains; it can also be found in the domain of negotiation (e.g., Kray, Thompson, & Galinsky, 2001) and for white males, under African American threat (Steele & Aronson, 1995). This boost in performance for the nonstigmatized population under conditions that elicit the alleged inferiority of stigmatized populations is apparent in the majority

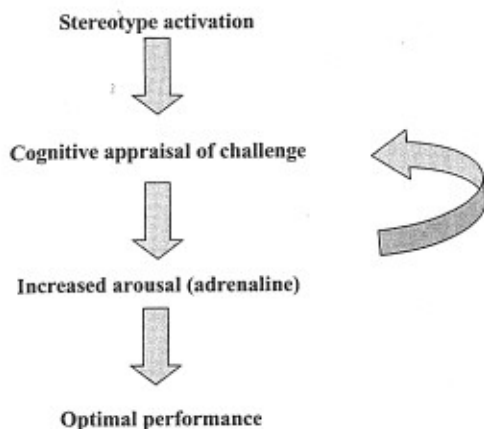


FIGURE 9.2. Facilitation effects of nonstigmatized individuals mediated by challenge and heightened catecholamines.

of research on stereotype threat¹ and can provide a fruitful area of investigation for understanding stereotype threat better.

DISCUSSION AND FUTURE DIRECTIONS

In this chapter, we propose a new framework for examining stereotype threat-induced underperformance, in which stigmatized people's responses to stereotypes alleging their intellectual inferiority engage in a threat appraisal that heightens cortisol levels and hurts intellectual performance. A further prediction is that if people learn to adopt a challenge appraisal instead of a threat appraisal, then performance deficits may be averted.

Wheeler and Petty (2001) propose an alternative account to a cognitive appraisal-mediated arousal for explaining stereotype-threat phenomena. They propose that stereotypes may result from an automatic activation of a network of associations that may hurt performance without the elicitation of higher-level cognitive processing, such as cognitive appraisals. For example, ideomotor theory (e.g., Bargh, Chen, & Burrows, 1996) suggests that particular cues in the environment can automatically affect motivation and performance, without conscious awareness. As an example, a woman entering a room in which a television set is playing an advertisement related to cleaning products might automatically shift into an associated

¹ Usually, the nonstigmatized population is not the focus of such research in stereotype threat because this population tends to always perform better than the stigmatized population. However, when data are reported on the nonstigmatized population, the trend of the nonsignificant difference becomes apparent across different studies.

motivational state where domestic activities are a priority and other tasks, such as academic or career-oriented tasks, are deemed less important.

The problem with ideomotor theory as an explanation for stereotype threat is its inability to account for the fact that low-domain identifiers do not show performance deficits. Identification with the threatened domain (the degree to which one's self-concept is linked to a domain) is central to stereotype threat theory. The more a person cares about being successful in the domain, the more he or she is predicted to be affected by stereotypes alleging a lack of ability in that domain (Steele, 1997). Research by Aronson et al. (1998) indicates that the stereotype threat manipulation disrupts only the performance of the high identifiers. Thus, people who are at risk from stereotype threat are those stigmatized individuals who are at the vanguard of their group in a given domain.

Wheeler and Petty (2001) also discuss other potential mediators of stereotype threat that are in agreement with our model. For example, they propose that stereotype activation changes people's perceptions of their environment, including aspects of other people in it. They discuss a study done by Herr (1986), which showed that participants primed with hostility cues began perceiving their neutral partner as hostile, and behaved more competitively, than those primed with nonhostile cues. This explanation is consistent with the idea that if people engage in threat vs. challenge appraisals, then they focus on potential failure rather than on potential success.

Wheeler and Petty (2001) argued that another explanatory variable may be found in literature on multiple selves. For example, Ruvolo and Markus (1992) asked participants to imagine their own futures as being successful or unsuccessful. Those who imagined positive visions of their future selves showed greater effort and persistence in the task that followed. Perhaps the types of primes that have been used in stereotype threat research, such as an Asian prime or a female prime, could also be activating possible selves. A person could go through an automatic shift, into a particular "self," as well as engage in a more conscious, cognitive appraisal of threat or challenge at the same time.

Another question is, what are the roles of beliefs in producing threat? As we discuss in the section on alternative mediators to arousal, self-efficacy does not appear to mediate stereotype threat. Instead, based on previous evidence that only high achievers are vulnerable to the debilitating effects of threat (e.g., Steele, 1997), it follows that a high-domain self-efficacy may be a *moderator* of stereotype threat. That is, stigmatized individuals with a high domain self-efficacy would be most prone to experiencing problem-solving deficits in threatening environments.

Situationally, after the activation of a stereotype, and as a possible result of threat, people may endorse maladaptive beliefs. Schoenfeld (1983) found that students who thought less of their mathematical abilities were more

likely to attribute their successes to luck and their failures to lack of mathematical ability, whereas students who thought of themselves as high in math ability were more likely to attribute their successes to ability. It would be informative to examine whether, under threat, stigmatized students who identify highly with ability and success in a domain would nevertheless tend to rate themselves lower on ability as a result of experiencing threat situationally.

For example, beliefs about the malleability of intelligence may be implicated in stereotype threat. Dweck and colleagues (Dweck, 1999; Dweck & Leggett, 1988) found that individuals' implicit theories about intelligence have a direct impact on how they think and feel about – as well as approach and engage in – performance tasks. Individuals who believe intelligence to be a malleable and dynamic process ("incremental" theorists) often approach cognitive and performance tasks with a "mastery-oriented" style, which leads them to successful outcomes, whereas those who believe that intelligence is more of a fixed, innate quality ("entity" theorists) are more vulnerable to a "helpless" attitude toward problem solving, which renders them prone to disengage from tasks and to underperform, even in domains in which they have achieved success and enjoy (Dweck, 1999).

However, Dweck (1999) also found that these same, apparently embedded beliefs about intelligence can be changed (if only temporarily) when exposed to different models of intelligence. Notably, when given a lecture on the incremental theory of intelligence, college students who had done poorly on a test were more apt to express interest in taking a tutorial to improve their score, whereas students who had been lectured on an entity theory of intelligence (Hong et al., 1999, in Dweck, 1999) were not.

In addition, Aronson, Fried and Good (2002) found that using an incremental framework of intelligence helped to reduce the impact of stereotype threat among college students. In other words, teaching students that intelligence is not a fixed entity, but is something that can be improved upon both with sustained effort and by varying one's problem-solving strategies, can effectively lessen students' vulnerability to stereotype threat (because stereotypes are based on the idea of innate, fixed traits) and subsequently improve their performance (Dweck, 1999). These findings are encouraging because they give credence to the idea that an individual's abilities may not be predetermined and that stereotype threat (as well as stereotypes themselves) can be diminished with appropriate and meaningful interventions.

The reality of stereotype threat is disconcerting. Students who suffer from stereotypes about their intellectual abilities are vulnerable to situational cues, which activate negative stereotypes and cause performance deficits. We hypothesize that the priming of a stereotype results in threat appraisal that causes heightened arousal and results in more algorithmic rather than meaningful problem solving. By trying to uncover the role of physiological arousal in mediating stereotype threat and the cognitive

nature of the resultant deficits, we hope to contribute to the mitigation of stereotype threat. Thus, our work is designed to help understand the theoretical underpinnings of stereotype threat, as well as to help stigmatized students overcome its effects by enhancing resilience to stereotypes and turning threat into challenge.

AUTHORS' NOTE

The writing of this chapter and some of the research described in it was enabled, in part, by a National Science Foundation ROLE grant, Award # 0207946, given to Talia Ben-Zeev and Steven Fein in September 2002.

We thank Elissa Epel and James Greeno for their invaluable insights. Correspondence concerning this chapter should be addressed to Talia Ben-Zeev, Department of Psychology, 1600 Holloway Ave., San Francisco, California 94132; E-mail: tbenzeev@sfsu.edu.

References

- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113-125.
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*, 29-46.
- Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 83-103). San Diego, CA: Academic Press.
- Aronson, J., & Salinas, M. F. (1997). *Stereotype threat, attributional ambiguity, and Latino underperformance*. Unpublished manuscript, University of Texas.
- Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality & Social Psychology, 71*, 230-244.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Baumeister, R. F., & Showers, C. J. (1986). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology, 16*, 361-383.
- 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*, 474-479.
- Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210*, 1262-1264.
- Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029-1031.

- Ben-Zeev, T., Fein, S., & Inzlicht, M. (in press). Arousal and stereotype threat. *Journal of Experimental Social Psychology*.
- Blascovich, J., Spencer, S. J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*, 225-229.
- Brown, R. P., & Josephs, R. A. (1999). Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology, 76*, 246-257.
- Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588-594.
- Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review, 96*, 84-100.
- Dweck, C. S. (1999). *Self-Theories: Their role in motivation, personality and development*. Philadelphia: Psychology Press.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*, 256-273.
- Eccles, J. S., & Jacobs, J. E. (1987). In M. R. Walsh (Ed.), *The Psychology of Women* (question 10, pp. 333-354). New Haven, CT: Yale University Press.
- Folkman, S., & Lazarus, R. (1986). Stress process and depressive symptomatology. *Journal of Abnormal Psychology, 95*, 107-113.
- Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19*, 229-284.
- Hong, Y., Chiu, C., & Dweck, C. S. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality & Social Psychology, 77*, 588-599.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta analysis. *Psychological Bulletin, 107*, 139-155.
- Inzlicht, M. (2001). *Threatening intellectual environments: When and why females perform worse in the presence of males*. Unpublished doctoral dissertation, Brown University.
- 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.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality & Social Psychology, 80*, 942-958.
- 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.
- Lupien, S. J., & McEwen, B. S. (1997). The acute effects of corticosteroids on cognition: Integration of animal and human model studies. *Brain Research Reviews, 24*, 1-27.
- Marx, D. M., Brown, J. L., & Steele, C. M. (1999). Allport's legacy and the situational press of stereotypes. *Journal of Social Issues, 55*, 491-502.
- O'Brien, L. T., & Crandall, C. S. (2001). *Stereotype threat and arousal: Effects on women's math performance*. Manuscript submitted for publication.

- Pyszczynski, T., & Greenberg, J. (1983). Determinants of reduction in effort as a strategy for coping with anticipated failure. *Journal of Research in Personality, 17*, 412-422.
- Ruvolo, A. P., & Markus, H. R. (1992). Possible selves and performance: The power of self-relevant imagery. *Social Cognition, 10*, 95-124.
- Schmitt, B. H., Gilovitch, T., Goore, N., & Joseph, L. (1986). Mere presence and social facilitation: One more time. *Journal of Experimental Social Psychology, 22*, 242-248.
- Schoenfeld, A. H. (1983). Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. *Cognitive Science, 7*, 329-363.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.
- Spencer, S. J., Steele, C. M., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28.
- Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75*, 1191-1197.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613-629.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.
- Steele, C. M., Spencer, S. J., & Lynch, M. (1993). Self-image resilience and dissonance: The role of affirmational resources. *Journal of Personality and Social Psychology, 64*, 885-896.
- 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.
- Ursin, H., Baade, E., & Levine, S. (Eds.). (1978). *Psychobiology of stress: A study of coping men*. New York: Academic Press.
- Walsh, M., Hickey, C., & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles, 41*, 219-240.
- Wheeler, S. C., & Petty, R. E. (2001). The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin, 127*, 797-826.
- Yerkes, R. M., & Dodson, J. D. (1908). The relationship of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology of Psychology, 18*, 459-482.
- Zajonc, R. B. (1965). Social facilitation. *Science, 149*, 269-274.