Creating, Closing, and Reversing the Gender Gap in Test Performance

Frédérique Autin, Nyla Branscombe, Jean-Claude Croizet

Creating, Closing, and Reversing the Gender Gap in Test Performance: How Selection Policies Trigger Social Identity Threat or Safety Among Women and Men

Frédérique Autin, Nyla R. Branscombe and Jean-Claude Croizet

Psychology of Women Quarterly published online 8 November 2013
DOI: 10.1177/0361684313510485

Article

Psychology of Women Quarterly
XX(X) 1-13

DOI: 10.1177/0361684313510485
pwq.sagepub.com
SAGE

Creating, Closing, and Reversing the Gender Gap in Test Performance: How Selection Policies Trigger Social Identity Threat or Safety Among Women and Men

Frédérique Autin[1,2], Nyla R. Branscombe[3], and Jean-Claude Croizet[1]

Abstract

We investigate how selection policies—the rules defining access to a valued position—can act as situational cues signaling social identity threat or safety among women and men. College students took a logic test ostensibly determining their assignment to a position of leader or subordinate for a subsequent task. Study 1 showed that when only the test score determined the selection, women experienced more identity threat and performed worse than men. When the policy allowed the selection of women at a lower level of performance than men to promote diversity, men's performance decreased compared to the merit condition, falling to the level of women's performance and thus closing the gender gap. Study 2 replicated these findings and established that the meaning derived from selection practices affects candidates' performance. A third policy that also preferentially selected women, but to correct for unequal treatment based on gender, leads to a reversed gender gap (i.e., women outperformed men). These findings suggest that structural features of test settings including selection practices can constrain individuals' potential access to opportunities.

Keywords

selection tests, social identity, stereotype threat, personnel selection, human sex differences, test bias, social discrimination

Most admission tests for higher education reveal gender differences in scores favoring men over women. Men averaged 31 points higher than women on the math section of the Scholastic Assessment Test in 2011 (College Board, 2011). From 2003 to 2010, Law School Admission Test scores have been higher for men than women (Law School Admission Council, 2010). Likewise, men have been outperforming women on the Graduate Management Admission Test (GMAT) over the last 10 years (Graduate Management Admission Council, 2011). We investigate whether aspects of the selection procedure itself could play a role in the gender gap in performance on standardized tests.

Men and women can differentially experience test settings. They especially differ in perceived devaluation of their gender identity, which can trigger a social identity threat that has the power to disrupt performance (Steele, Spencer, & Aronson, 2002). Features or cues in the test setting can signal that one's identity is safe or threatened (Murphy, Steele, & Gross, 2007). We propose that features of the selection procedure itself can convey identity threat or safety; by varying the social context—the rules determining who will be selected or not—we can create, close, or reverse the gender gap in performance on a standardized test.

Situational Cues Signal Threat or Safety

When taking a math, science, or logic test, women's awareness of the stereotype about their inferiority in these domains may create an extra burden that can prevent them from performing at their best (Spencer, Steele, & Quinn, 1999). This phenomenon is known as stereotype threat, which is a subtype of a broader social identity threat that arises when people believe they might be devalued because of their membership in a social group (Branscombe, Ellemers, Spears, & Doosje, 1999; Steele et al., 2002). Identity threat has a variety of negative consequences including stress (Townsend, Major, Gangi, & Mendes, 2011), negative thoughts (Cadinu, Maass,

[1] Department of Psychology, CeRCA, University of Poitiers and Centre National de la Recherche Scientifique, Poitiers, France
[2] Institut des Sciences Sociales, University of Lausanne, Lausanne, Switzerland
[3] University of Kansas, Lawrence, KS, USA

Corresponding Author:
Frédérique Autin, Institut des Sciences Sociales, Université de Lausanne–ISS-SSP, Geopolis, CH 1015, Lausanne, Switzerland.
Email: frederique.autin@unil.ch

Rosabianca, & Kiesner, 2005), task-related worries (Beilock, Rydell, & McConnell, 2007), depleted working memory capacity (Schmader & Johns, 2003), and impaired intellectual performance (see Nguyen & Ryan, 2008, for a review). Settings are considered "safe" if group members perceive they might be neutrally or positively judged. When the negative stereotype is not perceived as relevant in the test setting, identity safety preserves women's performance (Davies, Spencer, & Steele, 2005; Spencer et al., 1999). For men, the feeling of worth induced by the awareness of out-group devaluation can even lift their performance (Walton & Cohen, 2003).

Social identity threat has been shown to be triggered by subtle situational cues such as the numerical underrepresentation of one's group (Inzlicht & Ben-Zeev, 2000), exposure to stereotypic commercials (Davies et al., 2005; Davies, Spencer, Quinn, & Gerhardstein, 2002), stereotypical environments (Cheryan, Plaut, Davies, & Steele, 2009), or the suggestion of a biased instructor (Adams, Garcia, Purdie-Vaughns, & Steele, 2006). In these studies, devaluation does not result from overt prejudice and discrimination but rather from subtle and structural cues present in a test setting. Thus, minorities and women are oppressed not only by intentional acts of differentiated treatment but also by cultural values, norms, and practices that structure their environment (Adams, Biernat, Branscombe, Crandall, & Wrightsman, 2008). In the present article, we focus on a specific form of institutional practices: selection policies. We propose that they can convey identity threat or safety in men and women. We consider how the rules that are perceived as relevant to accessing a valued position can influence the test performance of women and men during a selection process. Due to the experimental nature of this work, the selection policies used in our studies are not actual practices used by institutions; rather they are proxies for the types of policies that candidates might face during a real selection process.

Selection Policies and Intellectual Performance

During a selection process, various criteria[1] can be considered when deciding among candidates. Due to the meritocratic ideal, measures of individual merit (e.g., standardized test scores) are often perceived as the best criterion (Bobocel, Son Hing, Davey, Stanley, & Zanna, 1998). However, because selection based on test scores results in outcomes that show systematic bias against women and minority groups, some researchers advocate including group membership in the selection criteria (Walton, Spencer, & Erman, 2013). People's attitudes toward selection policies have been the focus of research attention (Harrison, Kravitz, Mayer, Leslie, & Lev-Arey, 2006), but little is known about how selection criteria might affect intellectual performance.

Some studies have considered how selection policies influence performance-related outcomes by examining women's self-perceptions and performance evaluation after being selected for a leadership position. These studies thus used contexts where women may be targeted by the negative stereotype about their managerial abilities. Indeed, despite some changes over time, leaders are predominantly perceived as masculine (Koenig, Eagly, Mitchell, & Ristikari, 2011). People view managers as similar to men, but not very similar to women, and as more masculine rather than androgynous or feminine; furthermore, they describe them with more agentic than communal traits. Women are viewed as possessing less leadership ability, and leadership enactments by women are less favorably evaluated than when these behaviors are produced by men (Eagly & Karau, 2002). Thus, the consequences of selection policies have been examined in contexts where women may experience social identity threat. This literature focused on women's self-views and performance evaluations after being selected for a leadership position under either a merit-based policy or a gender-based preference policy. When women thought that they had been selected only because of their gender, they evaluated their performance more negatively, doubted their competence, and devalued their accomplishments (Heilman, Battle, Keller, & Lee, 1998; Heilman, Rivero, & Brett, 1991; Heilman, Simon, & Repper, 1987; Major, Feinstein, & Crocker, 1994; see Turner & Pratkanis, 1994, for a review). These results suggest that women who benefited from gender-based preferential selection do experience psychological discomfort.

Three studies have investigated the actual performance of women who benefitted from various selection policies (Brown, Charnsangavej, Keough, Newman, & Rentfrow, 2000; Roberson & Alsua, 2002; Turner & Pratkanis, 1993). In each study, women showed impaired performance when they thought they had been selected only because of their gender, compared to their merit. However, this detrimental effect disappeared if the assignment was based on both gender and merit (i.e., women were given preference only if they were equally qualified; Brown et al., 2000) or when the task was presented as a training task, thus reducing its diagnosticity (Roberson & Alsua, 2002). Previous research therefore suggests that selection criteria can result in performance deficits among people who have benefited from the policy, possibly because these criteria elicited identity threat. However, these studies did not address the central question of how selection criteria affect individuals' performance *at the selection stage*. People rely on selection procedures (e.g., standardized tests) to gain access to resources such as higher education and employment. We propose that performance that determines access to opportunities can be affected precisely by the rules used to define access.

Previous research has only examined the consequences of selection criteria on performance for the group that is usually devalued in the domain (e.g., women in leadership). Nothing is known about the impact of selection criteria on groups that are not usually devalued (e.g., men). Yet, we know that situational cues can elicit identity threat in men. For example, when anticipating that they might lose their high status, men

experience identity threat characterized by a physiological reaction of stress (Scheepers, Ellemers, & Sintemaartensdijk, 2009), which can disrupt performance (Blascovich, Spencer, Quinn, & Steele, 2001). Fully understanding how the selection procedure contributes to the gender gap in standardized test scores requires investigating the impact of selection criteria on the experience of identity threat or safety in men as well as in women.

Giving Meaning to Selection Practices

Prior studies investigating selection criteria manipulated only whether gender was included in the criteria themselves or not. No information was given to help the participants understand why such a selection policy was applied, although research suggests that the features of a setting have different consequences for one's identity depending on the meaning given to those features. For example, a low representation of African Americans in a company can trigger identity threat for those group members if it is perceived as a sign of discrimination. However, if the setting conveys that diversity and mutual respect are valued, then African Americans experience identity safety despite the underrepresentation of their group (Purdie-Vaughns, Steele, Davies, Ditlmann, & Crosby, 2008). Thus, situational cues that help candidates give meaning to selection criteria may contribute to men's and women's experience of identity threat or safety. Three different justifying philosophies emerge in the debate about selection policies which focus on (a) individual merit, (b) diversity, and (c) correction of unequal treatments experienced by each groups. We propose that the rules determining access to opportunities that are defined by these philosophies have differential consequences for women's and men's social identities, which may translate into differences in test performance.

The implementation of selection policies based on test scores is usually justified by the meritocratic ideal. Test scores are presented as an objective measure of individual merit (Sackett & Wilk, 1994). Considering any other factor, including group membership, would violate the principles of fairness and equality. However, research has shown that the mere presentation of a task as diagnostic of intellectual ability is enough to trigger identity threat among women and safety threat among men. The gender gap in performance appeared when participants were merely told that they would be taking a math test. However, the gap was reduced when participants learned that the test yields no gender difference (Quinn & Spencer, 2001). Evaluative situations seemingly based on merit are therefore not free from the influence of group stereotypes on performance. We expect that when scores on a logic test are said to be the only selection criterion for a leadership position, women will be threatened by their supposed incompetence in both logic and leadership and will show impaired performance. In contrast, men's performance should be lifted by the identity safety that results from their awareness of women's supposed incompetence. The usual gender gap in logical reasoning should therefore be confirmed in this meritocratic setting.

The inclusion of group membership in the selection criteria is commonly presented as a way to ensure diversity (Miller, 1997; Tierney, 1997). Such a policy increases the representation of minorities in education (Lempert, Chambers, & Adams, 2000) and in the workplace (Herring & Collins, 1995). Diversity benefits both minority and majority members by enhancing consideration of others' perspectives and bringing complementary skills (Gurin, Dey, Hurtado, & Gurin, 2002). Although this philosophy emphasizes women's complementary skills, it does not question their supposed incompetence in the stereotypical domain. Thus, the salience of the negative stereotype in this setting should trigger social identity threat for women and have a deleterious effect on performance (see Schmader, Johns, & Forbes, 2008). From the perspective of men, increased representation of women could put their higher status in jeopardy, which might trigger identity threat (Scheepers et al., 2009). When the inclusion of gender in the selection criteria is presented as a way to increase diversity, men should experience identity threat, thus lowering their performance to the level of the women who are also threatened by the negative stereotype. We thus hypothesize that the gender gap in performance will be reduced by the selection policy promoting diversity.

A justifying philosophy could signal social identity safety for women by casting doubt both on women's supposed incompetence and on men's supposed greater competence. The correction-of-bias philosophy presents the underrepresentation of women not as a sign of their incompetence but as a result of systematic obstacles; and the overrepresentation of men not as a sign of greater competence but of systematic privilege (Crosby, Iyer, & Sincharoen, 2006). This conception stems from evidence that racial minorities and women face prejudice and discrimination, whereas racial majority members benefit from favoritism (Dovidio & Gaertner, 2000). The standard testing setting could systematically underestimate racial minorities' and women's potential because it triggers identity threat and lower test performance while lifting majorities' and men's performance (Walton & Cohen, 2003; Walton & Spencer, 2009). A way to compensate for this bias is to correct the scores of negatively stereotyped candidates (Walton et al., 2013). The inclusion of gender in selection criteria can thus be framed as a way to counter barriers faced by women and to correct for men's privilege. From the perspective of women, this framing should question the negative stereotype about their incompetence and allow them to experience the test setting as safe for their identity, thereby restoring their performance. From the perspective of men, this policy reminds them of their privilege, which can be distressing (Branscombe, 1998). Such a reminder should elicit identity threat in men and lower their performance. We thus expect a reversed gender gap in performance, such that women's performance will be better than men, under a policy that emphasizes the correction of structural bias between gender groups.

Present Research

We tested the hypothesis that selection policies can act as situational cues signaling identity threat or safety for men and women. Undergraduate men and women were told they would be placed in a work group setting ostensibly involving two groups differing in status: managers and workers. To determine their group assignment, participants were told they would take a test that is diagnostic of logical reasoning ability, which is highly related to leadership and decision making—qualities essential to succeed as a manager.

In both experiments, we investigated whether the selection policy affects men's and women's performance on the selection test. In Experiment 1, there was one condition where selection was based on merit, with admission to be gauged by test scores. In the second condition, both test scores and gender were said to be taken into consideration such that women were to be selected at a lower level of performance than men. This policy was presented as a way to increase diversity in the high-status (manager) group. In Experiment 2, we further investigated the impact of the meaning that candidates give to the selection policy. We replicated the two conditions of Experiment 1 and added a third condition. In this new condition, women again would be accepted at a lower level of performance, but the selection policy was presented as a way to correct for unequal treatment based on gender.

Experiment 1

College men and women completed a logic test after reading text describing the selection procedure that would be used to choose people for a group task. In a first condition, participants learned that the logic test score was the only criterion determining the selection of the managers of the group task. In a second condition, participants were told that both their test score and gender would be considered in the selection process. More specifically, their test score was to be adjusted by gender, with women being selected as managers at a lower level of performance than men. This selection policy was presented as a way to achieve a more balanced representation of both genders in the high-status manager group.

We expected that in the setting emphasizing merit, gender stereotypes would elicit identity threat in women and identity safety in men. Women as a group would thus perform worse than men as a group. Under a policy favoring women to promote diversity, the negative stereotype about women's ability remains and thus should threaten women's identity. We predicted that in the diversity condition men should be threatened by the prospect of status loss so that their performance would be lowered to the level of women's, resulting in a reduced gender gap in performance. To assess the possibility that the changes in men's and women's performance are related to changes in their engagement with the test, we also analyzed the number of items attempted. To investigate their perception of the selection procedure, participants also reported their experience of social identity threat.

Method

Participants

A sample of 117 psychology undergraduates (58 women and 59 men) at a large U.S. Midwestern university participated in exchange for course credit. No additional demographic information was recorded but judging from other research with this population, it is safe to assume that participants were primarily White and young (about 19 or 20 years of age). They were introduced to the experiment by one of two experimenters (one male and one female) in mixed-gender groups composed of 5–10 students.

Procedure

Participants were informed that the study involved first taking a test and then working on a group task. Participants received a booklet containing the instructions implementing the selection policy manipulation, followed by measures of social identity threat and difficult items from the Graduate Management Admission Test (GMAT). Participants read the cover story informing them that the study would investigate group work in a simulated organizational setting and that two groups would be created: a group of managers and a group of workers. Managers would receive interesting and challenging tasks involving supervision and evaluation of the workers who would receive repetitive tasks. Their group assignment would be allegedly based on their performance on the upcoming test that was presented as evaluative of logical reasoning ability and predictive of success as a manager. Within each gender, participants were randomly assigned to one of the two selection conditions: either that group task assignment would be (a) based on test scores only or (b) based on both test scores and gender.

Participants in the *merit condition* read that despite the usually observed underrepresentation of women in the manager group, the test was a reliable tool to estimate individual ability. With these instructions, our goal was to achieve two things. First, we wanted to mimic the usual conditions under which both women and men take tests: Individuals think they are evaluated on their ability while knowing that some groups are known to perform better than others. Indeed, people automatically link merit-based evaluation with group stereotypes about ability (Steele & Aronson, 1995; Walton & Cohen, 2003). As such, the merit condition constitutes a performance baseline for both gender groups. Second, at a methodological level, by explicitly mentioning group membership, we aimed to equate experimental conditions regarding the salience of social identity. This kind of induction is classically used to investigate social identity threat or safety (Beilock et al., 2007; Rydell, McConnell, & Beilock, 2009). The implementation of the policy was illustrated using a race-running metaphor. Participants read that to ensure a fair race to succeed,

the conditions had to be the same for all runners so that individual ability was the only deciding factor. No suggestion that gender would affect group assignment was given in this condition. Participants were simply told they had to get more than 60% of the items correct to be selected as a manager.

Participants in the *diversity condition* read that women would be favored in selection in order to reduce their underrepresentation in the manager group. Specifically, it was noted that because some categories (women) lose the race to succeed more often, they would be given a head start to ensure diversity. To achieve a more balanced representation of both genders, women would be accepted in the managers group at a lower threshold of performance (50% of correct answers on the test) than men (60%).

Before taking the test but after reading the selection policy, participants were asked to complete measures of social identity threat. Social identity threat arises from the perception that one's group might be devalued and unfairly treated in a setting (Steele et al., 2002). We thus assessed social identity threat with a measure of perceived fairness of the selection setting for the in-group. Participants reported on a 6-point scale, from 1 (*extremely unjust) t*o 6 (*extremely just*), how fair they thought the selection procedure was for their gender group. This item was reverse scored so that higher numbers indicate a greater experience of threat.

The test consisted of four problems, each composed of 6 items, which were adapted from the analytical section of past GMAT exams (Adams et al., 2006). For example, in the first problem, participants were given comparative information about the wetness and fatness of six panda bears. To answer the 6 items, they had to deduce the relative position of the panda bears along these two dimensions. The participants were given 25 minutes to complete the 24 questions. Past research has established that the time constraint and the items' difficulty allow for the emergence of social identity threat (Adams et al., 2006).

Results

Self-Reported Social Identity Threat

Scores on the social identity threat measure were analyzed in a 2 (gender) × 2 (selection policy) analysis of variance (ANOVA).[2] The analysis yielded no main effect of selection policy ($F < 1$). A main effect of gender was obtained, indicating that men perceived their in-group as treated more fairly by the selection policy than did women, $F(1, 112) = 7.80$, $p = .006$, $d = .53$. This main effect was, as expected, qualified by the Gender × Selection Policy interaction, $F(1, 112) = 10.95$, $p = .001$, $d = .63$, as can be seen in Table 1(a). Planned comparisons revealed that women perceived the selection policy based on merit ($M = 4.40$) as more threatening than the selection policy promoting diversity ($M = 3.71$), $F(1, 112) = 7.13$, $p = .009$, $d = .51$. In contrast, men perceived the policy based on merit ($M = 2.93$) as less threatening than the policy promoting diversity ($M = 3.84$), $F(1, 112) = 4.04$, $p = .05$, $d = .38$. As a result, women felt more threatened than men in the merit condition, $F(1, 112) = 18.30$, $p < .001$, $d = .81$. However, when the selection policy promoted diversity, women and men did not differ in their perception of how fairly their in-group was treated ($F < 1$).

Table 1. Social Identity Threat and Logic Performance as a Function of Gender and Selection Policy, Experiment 1.

	Selection Policy	
	Merit *M* (*SD*)	Diversity *M* (*SD*)
	(a) Self-Reported Social Identity Threat Scale	
Women	4.40_a (1.35)	3.71_b (1.24)
Men	2.93_c (1.49)	3.84_b (1.09)
	(b) Logic Performance	
Women	$.45_a$ (.17)	$.49_a$ (.19)
Men	$.57_b$ (.20)	$.47_a$ (.19)

Note. GMAT = Graduate Management Admission Test; *SD* = standard error. The 6-point social identity threat item was scored so that higher numbers indicate a greater experience of threat. Logic performance is the proportion of each participant's correct answers across GMAT items. Means within each dependent variable that do not share subscripts across rows and columns differ significantly (*ps* ≤ .05).

Logic Performance

The proportion of correct answers on the logic test was analyzed in a 2 (gender) ×2 (selection policy) ANOVA. There was no main effect of selection policy ($F < 1$) or gender, $F(1, 113) = 2.16$, $p = .15$. As shown in Table 1(b), the expected interaction between gender and selection policy was significant, $F(1, 113) = 4.05$, $p = .05$, $d = .38$. Planned comparisons revealed that under a selection policy based only on merit the usual gender gap was observed: Women ($M = .45$) underperformed relative to men ($M = .57$), $F(1, 113) = 5.91$, $p = .02$, $d = .46$. When participants believed women would be preferentially selected to ensure diversity, men's performance was lower ($M = .47$) than in the condition emphasizing merit, $F(1, 113) = 4.54$, $p = .04$, $d = .40$. Women's performance in the diversity condition ($M = .49$) did not differ from the merit condition ($F < 1$). As a result, in the diversity condition, men did not perform differently from women ($F < 1$).

Further examination revealed no main effects of gender or selection policy and no interaction on the number of items attempted (*F*s < 1), suggesting that the performance effects cannot easily be accounted for by effort withdrawal. The number of items attempted did not correlate with self-reported social identity threat ($r = -.03$, $p = .73$). We also correlated test performance and self-reported social identity threat and found no significant correlation ($r = .004$, $p = .96$).

Discussion

The present findings suggest that selection policies can elicit social identity threat or safety in men and women. A selection

procedure based only on test scores, which best mimics usual test settings, conveyed threat for women and safety for men. In this setting, the usual gender gap was observed (i.e., men outperformed women). The inclusion of gender in the criteria to promote diversity was perceived as equally threatening to both gender groups. The gender gap in performance was closed in this condition, but gender equality was achieved at the expense of men's success. Our results indicate that structural features of a selection setting, such as the rules determining access to opportunities themselves, can work as situational cues signaling social identity threat or safety and affect performance.

Experiment 2

To further investigate the role played by the selection procedure in the gender gap on standardized test performance, we examined whether a different philosophy justifying the inclusion of gender in the selection criteria can moderate its impact on men's and women's test scores. Two philosophies emerged from the debate about why group membership should be included in selection criteria: One emphasizes diversity benefits, whereas the other focuses on the need to correct for unequal treatment between low- and high-status groups (Crosby et al., 2006). This last conception presents the correction of women's scores as a way to compensate for the biases in usual evaluation of individual potential, which promotes men's success and hinder women's (Walton et al., 2013).

Participants were again enrolled in a selection procedure ostensibly to form a high- and a low-status group (managers vs. workers) for a later task. Two conditions were replicated from the first experiment: A standard condition where selection to be a manager was only based on merit and a condition where women could expect to be selected at a lower level of performance to increase diversity. In a third condition, women were selected at a lower level of performance, but the inclusion of gender in the criteria was justified by the need to correct for the obstacles faced by women and the privileges that benefit men. We expected that this policy would release women from the threat of incompetence and restore their performance. Men, by comparison, should be threatened by the prospect of status loss and the reminder of their privilege, resulting in lowered performance similar to the diversity condition. To further examine the alternative account that the differences in men's and women's performance are due to effort withdrawal, we again analyzed the number of items attempted and included a self-report measure of motivation.

Method

Participants

Psychology students ($N = 204$) at a large U.S. Midwestern university took part in the study in exchange for course credit. No additional demographic information was recorded, but we assume that participants were primarily White and young. Participants were met by one of two female experimenters in mixed-sex groups with 4–12 students. Five participants were excluded after they expressed suspicion so that the final sample of 199 comprised 98 women and 101 men.

Table 2. Logic Performance as a Function of Gender and Selection Policy, Experiment 2.

	Selection Policy		
	Merit *M* (*SD*)	Diversity *M* (*SD*)	Correction *M* (*SD*)
Women	$.48_{a}$ (.21)	$.51_{a,\ b}$ (.21)	$.58_{b}$ (.19)
Men	$.59_{b}$ (.22)	$.48_{a}$ (.20)	$.48_{a}$ (.19)

Note. *SD* = standard error. Means across both rows and columns that do not share subscripts differ ($ps \leq .05$), except for the comparison of women in the merit and correction conditions ($p < .07$).

Procedure

The merit and the diversity conditions were identical to Experiment 1. Participants in the new *correction for unequal treatment condition* read that women were underrepresented in the managers group because the test overestimated men's competence and underestimated women's. The metaphorical runners' race to succeed was unfair therefore because the lanes used by some runners (women) had obstacles like stereotyping and biased tests that favor men. Participants were told that equality among test takers would be restored by correcting for the gender bias of the test. Scores would be adjusted and women selected as managers at a lower threshold of performance than men (50% vs. 60%). The logic test was identical to the one used previously. After the test, participants were asked to report on a 6-point scale, from 1 (*not important at all*) to 6 (*extremely important*), how important it was for them to perform well on the test.[3]

Results

The proportion of correct answers on the logic test was analyzed in a 2 (gender) × 3 (selection policy) ANOVA. The analysis yielded no significant main effect of gender or selection policy (Fs < 1). As can be seen in Table 2, the expected interaction between gender and selection policy was significant, $F(2, 193) = 4.47$, $p = .01$, $d = .43$. Replicating Experiment 1, planned comparisons revealed that when the selection policy emphasized merit, women ($M = .48$) underperformed relative to men ($M = .59$), $F(1, 193) = 4.75$, $p = .03$, $d = .31$. Again, men's performance was hindered when they thought women would be preferentially selected on behalf of diversity ($M = .48$) compared to the merit condition, $F(1, 193) = 5.64$, $p = .02$, $d = .34$. Under the diversity policy ($M = .51$), women's performance did not differ from the performance of women in the merit condition ($F < 1$). The gender gap was closed in the diversity condition where men's

performance did not differ from women's ($F < 1$). When gender was included in the criteria to correct for unequal treatment, men's performance ($M = .48$) was also lower in the merit condition, $F(1, 193) = 5.74$, $p = .02$, $d = .35$. In contrast, women's performance tended to be higher in the correction condition ($M = .58$) compared to the merit condition, $F(1, 193) = 3.33$, $p = .07$, $d = .26$. As a result, women performed better than men in the correction condition, $F(1, 193) = 3.84$, $p = .05$, $d = .28$. The gender gap was therefore reversed when women were expected to be preferentially selected to counter both the obstacles they face and men's privileges. Under this selection policy, women's performance did not differ from the performance of men in the merit condition ($F < 1$).

Supplemental analyses were conducted to examine a motivational account of these performance effects. The analysis of the number of items attempted yielded no main effects of gender or selection policy or their interaction (Fs < 1). Participants also did not differ in their reported motivation to succeed on the test (no main effect of gender, selection policy, or their interaction, all ps > .25).[4] Although reported motivation correlated with performance ($r = .23$, $p = .001$), when predicting performance, reported motivation did not interact with participant gender, $F(1, 192) = 1.64$, $p = .20$, selection policy, or both (Fs < 1), and controlling for motivation to do well did not affect the results on performance. These findings indicate that the variation we observed in test performance does not seem to be due to participants' differential engagement in the task.

Discussion

These results confirm that selection policies impact candidates' performance, suggesting that they create threat or safety differentially in women and men. Replicating Experiment 1, expected selection based on merit was deleterious for women and beneficial to men. In contrast, implementing a procedure favoring the selection of women hindered men's performance, and this pattern was observed regardless of the justification given for the selection criteria used. Although women's performance was still hindered by the diversity selection criteria, they benefited from a procedure that favored them to correct for unequal treatment. Such a construal of the test setting restores performance and allows women to perform at the same level as men in the merit condition. These findings demonstrate that the justification given for the selection criteria differentially impacts candidates' performance. The same set of criteria—including both test scores and gender—led to either a closed or a reversed gender gap in performance depending on the justification given for it.

General Discussion

With the present research, we sought to understand the potential role of selection policies in the gender gap observed in standardized test performance. We argued that selection policies can signal social identity threat or safety in men and women and thus affect performance. To test this hypothesis, we first implemented two selection policy conditions: one said to be simply based on test scores as a gauge of merit and the other based on both scores and gender to favor women on behalf of diversity. Results showed that when intellectual ability was said to be the only criterion, women experienced threat and showed lower performance than men. When women were to be preferentially selected to increase diversity, both women and men experienced threat and performed at the same low level. In the second experiment, we implemented a third policy that also included gender in the criteria but was justified, not to increase diversity but to correct for unequal treatment. This selection policy led to a reversed gender gap in performance (i.e., women outperformed men) and women performed at the same level as men in the merit condition, suggesting that women experienced identity safety and men experienced identity threat. The meaning derived from selection procedures can thus play a key role in candidates' performance.

Interestingly, men's performance was highly sensitive to the selection policies perceived to be in place when they took the test. The difference in men's performance on the logic test by threat condition cannot be accounted for by concerns about a negative stereotype of lower leadership skills because such a stereotype does not exist for men. Instead men's lower performance seems to reflect a more general concern about devaluation and unfair treatment of their group. This finding fuels the idea that the experience of threat is not contingent on having a stigmatized identity (Aronson et al., 1999). Group members who are usually advantaged and valued in the society can also experience the deleterious effect of being devalued and disadvantaged in a specific setting. Previous research demonstrated that members of high-status groups, who might lose their favorable status position, show a physiological response typical of identity threat (Scheepers et al., 2009). The present studies extend this research by showing that intellectual performance of members of high-status groups is negatively affected by the prospect of status loss.

Women's performance was positively affected by selection rules that favor them to counter the barriers they face along with men's privilege. Having better chances to be selected is thus not enough to restore women's performance. Rather women need to learn that their group's underrepresentation is due to structural barriers and that the negative stereotype is incorrect. It is probable that the preferential selection of women framed as a way to increase diversity reduces threat for the in-group. More women should get access to a high-status position so the group as a whole is perceived as not too badly treated, as suggested by the low level of social identity threat reported by women in the diversity condition. However, this setting would not restore performance because the self would still be the target of a threat (see Shapiro & Neuberg, 2007). Women may still fear that

they, as individuals, will underachieve and confirm the negative stereotype. Their performance is thus only restored when the stereotype is discounted as in the correction condition where their behavior would no longer be interpreted through the lens of their supposed incompetence.

Our results demonstrate that the gender gap in test performance can be not only closed but also reversed. Indeed, intellectual performance is highly sensitive to sociopsychological factors (Wilson, 2006) as repeatedly shown by the literature on stereotype threat (Quinn & Spencer, 2001; Steele et al., 2002). Our research extends this literature by showing that performance is affected not only by the stereotypes that are prominent in a testing situation but also by aspects of the social structure that are salient. We demonstrated that the rules that define access to opportunity can directly affect individuals' performance. Thus, depending on the way people perceive the testing situation, the gender gap can be created, closed, or reversed. In that sense, our study contributes to a conceptualization of the situational determinants of performance that includes structural factors such as the selection practices to which individuals are exposed.

Limitations and Future Directions

One limitation of our research is that we mentioned the usual underrepresentation of women in the manager group in the merit condition. Our intent was to equalize the salience of social identity among the different selection conditions. We think that this condition is equivalent to standard testing situations where individuals are aware of the stereotypes targeting social groups. However, by mentioning women's underrepresentation, we explicitly activated gender stereotypes. The impact of such explicit activation on social identity threat or safety has been inconsistent. Some research has shown that increasing the salience of the stereotype produces similar effects on performance compared to implicit activation (Campbell & Collaer, 2009; Keller, 2002; Marchand & Taasoobshirazi, 2013; Smith & White, 2002; Walton & Cohen, 2003). Other studies indicate that explicitly referring to the usual gender gap in performance induces reactance, characterized by the engagement in counter-stereotypical responses (Kray, Thompson, & Galinsky, 2001), and reduces the impact of stereotype threat on women's performance (Nguyen & Ryan, 2008). Due to this inconsistency in existing findings, it is difficult to be sure how mentioning women's underrepresentation affected our results. Having participants take the test without mentioning the usual women's underrepresentation in manager groups could either produce the same gender gap as the one we observed or, by avoiding reactance, foster women's underperformance and increase the gender gap.

A second limitation is that our research was conducted with mostly White participants and was confined to gender identity. Future research should investigate how ethnic–racial identity as it intersects with gender identity can be threatened by the rules that define access to a valued position. Studies have shown that leaders of color are perceived more negatively and face more obstacles in exercising their leadership than White leaders, notably because racial stereotypes do not fit with leadership characterization (Ospina & Foldy, 2009), but there is growing evidence that women of color face specific challenges in the domain of leadership (Sanchez-Hucles & Davis, 2010). Due to their multiple stigmatized identities, women of color may suffer multiple disadvantages (Bowleg, 2008; Rosette & Livingston, 2012). We might therefore predict heightened threat and performance decrements when both gender and ethnic–racial stereotypes are relevant in the selection setting (Gonzales, Blanton, & Williams, 2002).

Another limitation is the lack of documentation of the mechanism by which selection policies affect candidates' performance. We proposed that identity threat or safety underlies the performance variations. In our first experiment, we obtained the expected impact of gender and selection policy on identity threat. However, this self-report measure did not correlate with performance. This result seemingly questions the validity of the measure of identity threat. First, we defined identity threat as a perception that one's group might be unfairly treated because men would be unlikely to feel devalued in this context due to the positive stereotype of men's superiority in leadership. By not assessing the feeling of devaluation directly, we may have failed to capture a core aspect of women's experience of threat. Second, threat was measured at the group level but not at the individual level. Because identity threat is a multifaceted concept (Shapiro & Neuberg, 2007), future studies should include items asking participants about their experience of devaluation as well as unfairness—for themselves as well as for their group. Nonetheless, we should note that there is growing evidence that self-report methods often fail to capture reliable mediators of identity threat and safety effects (Bosson, Haymovitz, & Pinel, 2004; Johns, Inzlicht, & Schmader, 2008; Smith, 2004). More indirect measures of the experience of threat have been identified, including blood pressure, cardiovascular responses, and skin conductance (Croizet et al., 2004; Mendes, Blascovich, Lickel, & Hunter, 2002; Murphy et al., 2007; Scheepers & Ellemers, 2005). Such physiological markers of threat would be useful for future research.

The absence of a correlation between self-reported threat and performance questions the theoretical claim that social identity threat or safety contributes to the performance differential. Existing literature on the impact of selection policies on performance proposes alternative mechanisms. Roberson and Alsua (2002) showed that goal orientation moderates the impact of selection policy on performance such that preferential treatment is only deleterious for women when the test was framed as performance-oriented (i.e., an assessment of ability) as opposed to learning-oriented (i.e., learning opportunity). In our studies, the test was always presented as diagnostic of abilities; no mention of opportunity for training or improvement was made, so it is likely that all participants

adopted a performance goal orientation. Goal orientation thus cannot easily account for the observed interaction on performance. Turner and Pratkanis (1994) suggested that selection policies can be viewed as threatening or supporting for the individual self. As mentioned earlier, no measure of threat at the individual level was included. Further studies should investigate whether the selection policies implemented induce self-threat and affect self-efficacy or self-esteem.

Our research only begins to investigate the impact of selection practices on test performance. Future studies should explore potential moderators. In line with the stereotype threat literature, domain identification seems a good candidate. When leadership is central to women's identity, they might worry that their performance would confirm the negative stereotype. A poor performance would also question an important aspect of their self-concept, thus adding an extra pressure (Steele, 1997). Research suggests that settings where a negative stereotype is applicable are more likely to suppress the performance of women who are highly (Keller, 2007) or moderately identified with the testing domain (Nguyen & Ryan, 2008). In contrast, women low in domain identification would feel less performance pressure and suffer less from the threat (Nguyen & Ryan, 2008) or even show boosted performance (Keller, 2007). We thus hypothesize that the identity threat triggered by selection policies based on merit or promoting diversity would especially impair the performance of highly domain-identified women. The same logic can be applied to men. Those who highly identify with the domain would be threatened by the prospect that their group will lose its privileged status and that they will personally not succeed in an important domain. The performance drop observed in the diversity and the correction conditions relative to the merit condition would thus be exacerbated for highly identified men.

Another possible moderator is gender group identification. Although results are somewhat mixed, the literature on stereotype threat suggests that stigmatized individuals for whom group identity is highly important are more susceptible to underperform when the negative stereotype is applicable in the test setting (Armenta, 2010; Clark, Eno, & Guadagno, 2011; Kaiser & Hagiwara, 2011; Martiny, Roth, Jelenec, Steffens, & Croizet, 2012; Schmader, 2002; Wout, Danso, Jackson, & Spencer, 2008 but see Brown & Pinel, 2003; Kiefer & Sekaquaptewa, 2007). This implies that highly gender-identified women would have lower performance in the merit and diversity conditions. At the same time, gender identification is related to perceived discrimination among low-status group members (Branscombe, Schmitt, & Harvey, 1999; Major et al., 2002), so highly gender-identified women may be more responsive to the framing referring to correction for unequal treatment and show higher levels of performance. As for men, the performance lift induced by the negative outgroup reputation would occur in highly identified group members (Armenta, 2010; Walton & Cohen, 2003). Men for whom gender is central to their self-concept would perform at the highest level in the merit condition. Simultaneously, the possibility of losing a high-status position triggers identity threat only in members who highly identify with their group (Scheepers & Ellemers, 2005). Further studies should investigate whether the results obtained here are most likely in highly gender-identified participants.

Practice Implications

The aim of our research was not to mimic real selection policies per se in order to determine which one is the best to apply. In real situations, it has been previously established that perceiving that one has benefited from preferential treatment is associated with lower performance (van Laar, Levin, & Sinclair, 2008). Rather, our research was aimed at establishing a causal link between selection policies and actual performance of both beneficiaries and nonbeneficiaries and to investigate the impact of justifying philosophies. To achieve this goal, we had to (a) explicitly tell the candidates about the standards for who would be selected to be a manager and (b) give justifying information about the ways each policy would address group differences. Real workplace selection policies would not use these standards, nor would they be as explicit in conveying them.

Our results may, though, provide preliminary insight concerning how, in real settings, individuals' beliefs about selection policies could affect their performance. When no specific indication is given, candidates would probably assume that selection is based on merit, which is widely known to disrupt stigmatized candidates (Walton & Spencer, 2009). However, many institutions advertise that they apply equal opportunity and affirmative action policies. People often assume that affirmative action programs involve preferential treatment or quotas aimed at increasing the representation of low-status group members (Eberhardt & Fiske, 1994; Kravitz & Platania, 1993). Our research suggests that this belief could be detrimental for all candidates.

In contrast, if affirmative action policies are framed as useful in compensating for unequal treatment, women could experience the selection process more safely. However, male candidates would probably suffer from that message. A way to induce identity safety in both gender groups could be to mention past discrimination and acknowledge stereotypes to reaffirm women's competence but without mentioning men's privilege. This kind of message might show that the climate is welcoming for women without threatening men. Allusion to preferential selection of women might be best avoided so that men do not think they might lose their status. Although consistent with the present findings and our rationale for them, these suggestions are, however, highly speculative and more research is needed to establish what policy could be beneficial to both gender groups and investigate how the selection policies affect performance in real settings.

In our research, we sought to demonstrate how selection policies constrain candidates' potential access to opportunities where test performance is assessed to determine that access.

We showed that performance on the GMAT is affected by selection policies. In real situations, candidates would have taken this type of test when they apply for admission to education programs—well before they apply for the job. And yet, selection based on interviews or other kinds of tests would also probably be affected by identity threat. Identity threat is known to have a wide range of cognitive and emotional negative consequences (Beilock et al., 2007; Schmader et al., 2008). We could thus expect that if candidates perceive that they might be devalued or unfairly treated in the selection process, they would experience deleterious emotions and cognitions that might prevent them from doing their best at the test or interview.

Conclusion

Although our research was aimed at investigating how selection policies can be a situational determinant of performance, we believe our results contribute to the ongoing debate about how to reconcile the meritocratic principle with increased diversity goals (Crosby, Iyer, Clayton, & Downing, 2003; Walton et al., 2013). Walton et al. (2013) have argued that because test settings systematically underestimate the intellectual potential of negatively stereotyped candidates, correcting their test scores could be a way to establish "real" meritocracy. Our results indicate that candidates' awareness of the implementation of such policy affects their performance. Study 2 showed that a "correction" policy is indeed beneficial to negatively stereotyped candidates (i.e., women) but prevents the advantaged group (i.e., men) from performing at their best. Interestingly, the consequences of the policy favoring women in order to correct for unequal treatment are strikingly symmetrical to the consequences of a meritocratic setting. Both policies produce a gender gap but in opposite directions. Settings emphasizing merit thus seem to offer conditions that favor one group while disfavoring another. Our results suggest such settings can actually work as a preferential treatment policy for advantaged group members. Meritocratic settings should therefore not be considered more "neutral" or more "objective" than those employing other selection policies. Finally, the policy that favored women on behalf of diversity yielded no gender gap but group equality was obtained via lower performance in men relative to the merit condition.

Each selection policy, from assumed meritocracy to selection to expand diversity and correct inequities, can lead to the exclusion of individuals who would have succeeded in a different setting. Our research demonstrates that the rules defining access to opportunity can themselves introduce bias in the selection process. One implication of our work, along with the abundant literature on identity threat (Walton & Spencer, 2009), is that advantaged and disadvantaged group members' test performance cannot be assumed to reflect the individual's "true ability." Depending on the meaning individuals derived from the test setting, both men and women can be prevented from exhibiting their full potential no matter how apparently "objective" performance testing appears. The usual gender gap in test performance where men outperform women may reflect the greater safety men experience in typical test settings rather than the intellectual superiority of one gender group over the other.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:The research was supported in part by a grant from the Association Française des Femmes Diplomées des Universités awarded to Frédérique Autin.

Notes

1. In line with the literature on selection policies (Heilman et al., 1998; Turner & Pratkanis, 1994), the term "criteria" refers here to the kinds of information on which the selection decision is based (i.e., scores only or scores and gender).
2. One male participant did not fill in the scale, so there is one missing value in the analysis.
3. This study did not include a measure of social identity threat. Although a measure of perceived fairness was included, it mistakenly made no mention of the in-group—participants reported on a 6-point scale how fair they thought the selection procedure was from 1 (*extremely unjust*) to 6 (*extremely just*). We believe that the question framed this way does not assess social identity threat but a more global sense of justice. Analysis of this measure revealed a main effect of gender such that women ($M = 2.91$, standard deviation [SD] $= 1.22$) thought that the procedure was less just than men ($M = 3.27$, $SD = 1.14$) did, $F(1, 176) = 4.32$, $p = .04$, $d = .31$. The main effect of selection policy ($F < 1$) and the interaction, $F(1, 176) = 1.33$, $p = .27$, were not significant. This result suggests that the selection policy manipulation does not affect the general sense of justice experienced by participants.
4. One male participant did not fill in the motivation scale so there is one missing value in the analysis.

References

Adams, G., Biernat, M., Branscombe, N. R., Crandall, C. S., & Wrightsman, L. S. (2008). Beyond prejudice: Toward a sociocultural psychology of racism and oppression. In G. Adams, M. Biernat, N. R. Branscombe, C. S. Crandall, & L. S. Wrightsman (Eds.), *Commemorating Brown: The social psychology of racism and discrimination* (pp. 215–246). Washington, DC: American Psychological Association.

Adams, G., Garcia, D. M., Purdie–Vaughns, V., & Steele, C. M. (2006). The detrimental effects of a suggestion of sexism in an instruction situation. *Journal of Experimental Social Psychology*, *42*, 602–615. doi:10.1016/j.jesp.2005.10.004

Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural*

Diversity and Ethnic Minority Psychology, *16*, 94–98. doi:10.1037/a0017564

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.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, *136*, 256–276. doi:10.1037/00963445.136.2.256

Blascovich, J., Spencer, S. J., Quinn, D. M., & Steele, C. M. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, *12*, 225–229. doi:0.1111/1467–9280.00340

Bobocel, D. R., Son Hing, L. S., Davey, L. M., Stanley, D. J., & Zanna, M. P. (1998). Justice–based opposition to social policies: Is it genuine? *Journal of Personality and Social Psychology*, *75*, 653–669. doi:10.1037/0022–3514.75.3.653

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, *40*, 247–255.

Bowleg, L. (2008). When Black + lesbian + woman ≠ Black lesbian woman: The methodological challenges of qualitative and quantitative intersectionality research. *Sex Roles*, *59*, 312–325. doi:10.1007/s11199-008-9400-z

Branscombe, N. R. (1998). Thinking about one's gender group's privileges or disadvantages: Consequences for well-being in women and men. *British Journal of Social Psychology*, *37*, 167–184. doi:10.1111/j.2044-8309.1998.tb01163.x

Branscombe, N. R., Ellemers, N., Spears, R., & Doosje, B. (1999). The context and content of social identity threats. In N. Ellemers, R. Spears, & B. Doosje (Eds.), *Social identity: Context, commitment, content* (pp. 35–58). Oxford, England: Wiley-Blackwell.

Branscombe, N. R., Schmitt, M. T., & Harvey, R. D. (1999). Perceiving pervasive discrimination among African Americans: Implications for group identification and well-being. *Journal of Personality and Social Psychology*, *77*, 135–149. doi:10.1037/0022-3514.77.1.135

Brown, R. P., Charnsangavej, T., Keough, K. A., Newman, M. L., & Rentfrow, P. J. (2000). Putting the "affirm" into affirmative action: Preferential selection and academic performance. *Journal of Personality and Social Psychology*, *79*, 736–747. doi:10.1037/0022-3514.79.5.736

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, *39*, 626–633. doi:10.1016/S0022-1031(03)00039-8

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, *16*, 572–578. doi:10.1111/j.0956-7976.2005.01577.x

Campbell, S. M., & Collaer, M. L. (2009). Stereotype threat and gender differences in performance on a novel visuospatial task. *Psychology of Women Quarterly*, *33*, 437–444. doi:10.1111/j.1471-6402.2009.01521.x

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology*, *97*, 1045–1060. doi:10.1037/a0016239

Clark, J. K., Eno, C. A., & Guadagno, R. E. (2011). Southern discomfort: The effects of stereotype threat on the intellectual performance of US southerners. *Self and Identity*, *10*, 248–262. doi:10.1080/15298861003771080

College Board. (2011). *SAT percentile ranks for males, females, and total group: Mathematics* [Data file]. Retrieved from http://professionals.collegeboard.com/data-reports-research/sat/data-tables

Croizet, J.-C., Després, G., Gauzins, M.-E., Huguet, P., Leyens, J.-P., & Méot, A. (2004). Stereotype threat undermines intellectual performance by triggering a disruptive mental load. *Personality and Social Psychology Bulletin*, *30*, 721–731. doi:10.1177/0146167204263961

Crosby, F. J., Iyer, A., Clayton, S., & Downing, R. A. (2003). Affirmative action: Psychological data and the policy debates. *American Psychologist*, *58*, 93–115. doi:10.1037/0003-066X.58.2.93

Crosby, F. J., Iyer, A., & Sincharoen, S. (2006). Understanding affirmative action. *Annual Review of Psychology*, *57*, 585–611. doi:10.1146/annurev.psych.57.102904.190029

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, *88*, 276–287. doi:10.1037/0022-3514.88.2.276

Davies, P. G., Spencer, S. J., Quinn, D. M., & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, *28*, 1615–1628. doi:10.1177/014616702237644

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, *11*, 315–319. doi:10.1111/1467-9280.00262

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*, 573–598. doi:10.1037//0033-295X.109.3.573

Eberhardt, J. L., & Fiske, S. T. (1994). Affirmative action in theory and practice: Issues of power, ambiguity, and gender versus race. *Basic and Applied Social Psychology*, *15*, 201–220. doi:10.1080/01973533.1994.9646078

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, *28*, 659–670. doi:10.1177/0146167202288010

Graduate Management Admission Council. (2011). *2011 profile of GMAT candidates* [Data file]. Retrieved from http://www.gmac.com/market-intelligence-and-research/research-library/gmat-test-taker-data/profile-documents/profile-of-gmat-candidates-2006-07-2010-11.aspx

Gurin, P., Dey, E. L., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review*, *72*, 330–367.

Harrison, D. A., Kravitz, D. A., Mayer, D. M., Leslie, L. M., & Lev-Arey, D. (2006). Understanding attitudes toward affirmative action programs in employment: Summary and meta-analysis of 35 years of research. *Journal of Applied Psychology*, *91*, 1013–1036. doi:10.1037/0021-9010.91.5.1013

Heilman, M. E., Battle, W. S., Keller, C. E., & Lee, R. A. (1998). Type of affirmative action policy: A determinant of reactions to sex-based preferential selection? *Journal of Applied Psychology*, *83*, 190–205. doi:10.1037/0021-9010.83.2.190

Heilman, M. E., Rivero, J. C., & Brett, J. F. (1991). Skirting the competence issue: Effects of sex-based preferential selection on task choices of women and men. *Journal of Applied Psychology*, *76*, 99–105. doi:10.1037/0021-9010.76.1.99

Heilman, M. E., Simon, M. C., & Repper, D. P. (1987). Intentionally favored, unintentionally harmed? Impact of sex-based preferential selection on self-perceptions and self-evaluations. *Journal of Applied Psychology*, *72*, 62–68. doi:10.1037/0021-9010.72.1.62

Herring, C., & Collins, S. (1995). Retreat from equal opportunity? The case of affirmative action. In M. P. Smith & J. R. Feagin (Eds.), *The bubbling cauldron: Race, ethnicity, and the urban crisis* (pp. 163–181). Minneapolis, MN: University of Minnesota Press.

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. doi:10.1111/1467-9280.00272

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology: General*, *137*, 691–705.

Kaiser, C. R., & Hagiwara, N. (2011). Gender identification moderates social identity threat effects on working memory. *Psychology of Women Quarterly*, *35*, 243–251. doi:10.1177/0361684310384102

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. doi:10.1023/A:1021003307511

Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' math performance. *British Journal of Educational Psychology*, *77*, 323–338. doi:10.1348/000709906X113662

Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, *43*, 825–832. doi:10.1016/j.jesp.2006.08.004

Koenig, A. M., Eagly, A. H., Mitchell, A. A., & Ristikari, T. (2011). Are leader stereotypes masculine? A meta-analysis of three research paradigms. *Psychological Bulletin*, *137*, 616–642. doi:10.1037/a0023557

Kravitz, D. A., & Platania, J. (1993). Attitudes and beliefs about affirmative action: Effects of target and of respondent sex and ethnicity. *Journal of Applied Psychology*, *78*, 928–938. doi:10.1037/0021-9010.78.6.928

Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology*, *80*, 942–958. doi:10.1037//0022-3514.80.6.942

Law School Admission Council. (2010). *LSAT performance with regional, gender, and racial/ethnic breakdowns: 2003-2004 through 2009-2010 testing years*. Law School Admission Council [Data file]. Retrieved from http://www.lsac.org/LsacResources/Research/TR/tech-reports.asp

Lempert, R. O., Chambers, D. L., & Adams, T. K. (2000). Michigan's minority graduates in practice: The river runs through Law School. *Law & Social Inquiry*, *25*, 395–505.

Major, B., Feinstein, J., & Crocker, J. (1994). Attributional ambiguity of affirmative action. *Basic and Applied Social Psychology*, *15*, 113–141. doi:10.1080/01973533.1994.9646075

Major, B., Gramzow, R. H., McCoy, S. K., Levin, S., Schmader, T., & Sidanius, J. (2002). Perceiving personal discrimination: The role of group status and legitimizing ideology. *Journal of Personality and Social Psychology*, *82*, 269–282. doi:10.1037/0022-3514.82.3.269

Marchand, G. C., & Taasoobshirazi, G. (2013). Stereotype threat and women's performance in physics. *International Journal of Science Education*, *35*, 3050–3061. doi:10.1080/09500693.2012.683461

Martiny, S. E., Roth, J., Jelenec, P., Steffens, M. C., & Croizet, J.-C. (2012). When a new group identity does harm on the spot: Stereotype threat in newly created groups. *European Journal of Social Psychology*, *42*, 65–71. doi:10.1002/ejsp.840

Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interactions with White and Black men. *Personality and Social Psychology Bulletin*, *28*, 939–952. doi:10.1177/014616720202800707

Miller, F. (1997). The political rhetoric of affirmative action: Infusing the debate with discussions about equity and opportunity. *American Behavioral Scientist*, *41*, 197–204. doi:10.1177/0002764297041002002

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*, 879–885. doi:10.1111/j.1467-9280.2007.01995.x

Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, *93*, 1314–1334. doi:10.1037/a0012702

Ospina, S., & Foldy, E. (2009). A critical review of race and ethnicity in the leadership literature: Surfacing context, power and the collective dimensions of leadership. *The Leadership Quarterly*, *20*, 876–896. doi:10.1016/j.leaqua.2009.09.005

Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Ditlmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream

institutions. *Journal of Personality and Social Psychology*, *94*, 615–630. doi:10.1037/0022-3514.94.4.615

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, *57*, 55–71. doi:10.1111/0022-4537.00201

Roberson, L., & Alsua, C. J. (2002). Moderating effects of goal orientation on the negative consequences of gender-based preferential selection. *Organizational Behavior and Human Decision Processes*, *87*, 103–135. doi:10.1006/obhd.2001.2960

Rosette, A. S., & Livingston, R. W. (2012). Failure is not an option for Black women: Effects of organizational performance on leaders with single versus dual-subordinate identities. *Journal of Experimental Social Psychology*, *48*, 1162–1167. doi:10.1016/j.jesp.2012.05.002

Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. *Journal of Personality and Social Psychology*, *96*, 949–966. doi:10.1037/a0014846

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, *49*, 929–954. doi:10.1037/0003-066X.49.11.929

Sanchez-Hucles, J. V., & Davis, D. D. (2010). Women and women of color in leadership: Complexity, identity, and intersectionality. *American Psychologist*, *65*, 171–181. doi:10.1037/a0017459

Scheepers, D., & Ellemers, N. (2005). When the pressure is up: The assessment of social identity threat in low and high status groups. *Journal of Experimental Social Psychology*, *41*, 192–200. doi:16/j.jesp.2004.06.002

Scheepers, D., Ellemers, N., & Sintemaartensdijk, N. (2009). Suffering from the possibility of status loss: Physiological responses to social identity threat in high status groups. *European Journal of Social Psychology*, *39*, 1075–1092. doi:10.1002/ejsp.609

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*, 194–201. doi:10.1006/jesp.2001.1500

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, *85*, 440–452. doi:10.1037/0022-3514.85.3.440

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336–356. doi:10.1037/0033-295X.115.2.336

Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, *11*, 107–130. doi:10.1177/1088868306294790

Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. *Educational Psychology Review*, *16*, 177–206. doi:10.1023/B: EDPR.0000034020.20317.89

Smith, J. L., & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles*, *47*, 179–191. doi:10.1023/A:1021051223441

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

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

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

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). San Diego, CA: Academic Press.

Tierney, W. G. (1997). The parameters of affirmative action: Equity and excellence in the academy. *Review of Educational Research*, *67*, 165–196. doi:10.3102/00346543067002165

Townsend, S. S. M., Major, B., Gangi, C. E., & Mendes, W. B. (2011). From "in the air" to "under the skin": Cortisol responses to social identity threat. *Personality and Social Psychology Bulletin*, *37*, 151–164. doi:10.1177/0146167210392384

Turner, M. E., & Pratkanis, A. R. (1993). Effects of preferential and meritorious selection on performance: An examination of intuitive and self-handicapping perspectives. *Personality and Social Psychology Bulletin*, *19*, 47–58. doi:10.1177/0146167293191006

Turner, M. E., & Pratkanis, A. R. (1994). Affirmative action as help: A review of recipient reactions to preferential selection and affirmative action. *Basic and Applied Social Psychology*, *15*, 43–69.

van Laar, C., Levin, S., & Sinclair, S. (2008). Social identity and personal identity stereotype threat: The case of affirmative action. *Basic and Applied Social Psychology*, *30*, 295–310. doi:10.1080/01973530802502200

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467. doi:10.1016/S0022-1031(03)00019-2

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, *20*, 1132–1139. doi:10.1111/j.1467-9280.2009.02417.x

Walton, G. M., Spencer, S. J., & Erman, S. (2013). Affirmative meritocracy. *Social Issues and Policy Review*, *7*, 1–35. doi:10.1111/j.1751-2409.2012.01041.x

Wilson, T. D. (2006). The power of social psychological interventions. *Science*, *313*, 1251–1252. doi:10.1126/science.1133017

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group—and self-threat. *Journal of Experimental Social Psychology*, *44*, 792–799. doi:10.1016/j.jesp.2007.07.005