Journal of Personality and Social Psychology
2007, Vol. 93, No. 4, 544–564

0022-3514/07/$12.00 DOI: 10.1037/0022-3514.93.4.544

Mere Effort and Stereotype Threat Performance Effects

Jeremy P. Jamieson and Stephen G. Harkins
Northeastern University

Although the fact that stereotype threat impacts performance is well established, the underlying process(es) is(are) not clear. Recently, T. Schmader and M. Johns (2003) argued for a working memory interference account, which proposes that performance suffers because cognitive resources are expended on processing information associated with negative stereotypes. The antisaccade task provides a vehicle to test this account because optimal performance requires working memory resources to inhibit the tendency to look at an irrelevant, peripheral cue (the prepotent response) and to generate volitional saccades to the target. If stereotype threat occupies working memory resources, then the ability to inhibit the prepotent response and to launch volitional saccades will be impaired, and performance will suffer. In contrast, S. Harkins's (2006) mere effort account argues that stereotype threat participants are motivated to perform well, which potentiates the prepotent response, but also leads to efforts to counter this tendency if participants recognize that the response is incorrect, know the correct response, and have the opportunity to make it. Results from 4 experiments support the mere effort but not the working memory interference account.

Keywords: stereotype threat, mere effort, antisaccade, working memory

Stereotype threat "refers to the phenomenon whereby individuals perform more poorly on a task when a relevant stereotype or stigmatized social identity is made salient in the performance situation" (Schmader & Johns, 2003, p. 440). A wide range of stereotypes have been tested, from women's supposed lack of ability in math and science domains (Ben-Zeev, Fein, & Inzlicht, 2005; Brown & Josephs, 1999; Brown & Pinel, 2003; Davies, Spencer, Quinn, & Gerhardstein, 2002; Johns, Schmader, & Martens, 2005; Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Pronin, Steele, & Ross, 2004; Schmader & Johns, 2003; Sekaquaptewa & Thompson, 2003; Spencer, Steele, & Quinn, 1999) to African American's underperformance on standardized tests (Aronson, Fried, & Good, 2002; Blascovich, Spencer, Quinn, & Steele, 2001; Steele & Aronson, 1995) to White males' athletic inferiority (Stone, 2002; Stone, Lynch, Sjomeling, & Darley, 1999; Stone, Perry, & Darley, 1997). In each of these cases, concern about confirming the relevant stereotype have been shown to negatively impact the performance of the stigmatized individuals.

However, an important question remains: How does stereotype threat produce these performance effects? A number of explanations have been proposed (*anxiety*: Bosson, Haymovitz, & Pinel, 2004; Spencer et al., 1999; Steele & Aronson, 1995; *expectancy*: Cadinu, Maass, Frigerio, Impagliazzo, & Latinotti, 2003; *arousal*: Ben-Zeev et al., 2005; Blascovich et al., 2001; O'Brien & Crandall, 2003; *working memory interference*: Schmader & Johns, 2003; *cognitive load*: Croizet et al., 2004; *withdrawal of effort*: Stone, 2002; Stone et al., 1997; *reactance*: Kray, Thompson, & Galinsky, 2001), but researchers have yet to reach a consensus as to what process(es) actually mediate(s) the performance effects.

One explanation, working memory interference, has garnered significant attention in the literature as of late. This explanation conceptualizes stereotype threat as "a stressor in that a negative social stereotype that is primed in a performance

Jeremy P. Jamieson and Stephen G. Harkins, Department of Psychology, Northeastern University.

Special thanks to Neal Pearlmutter for all his help and advice in the conduct of this research and for use of the eye-tracker (supported by National Institutes of Health Grant R01-DC05237). Thanks also go to Sean Allen for writing the computer programs used in this work, Richard Gramzow for his helpful comments during the writing of this paper, and Rhea Eskew for his assistance with data analyses.

Correspondence concerning this article should be addressed to Jeremy P. Jamieson, Department of Psychology, 125 Nightingale Hall, Northeastern University, 360 Huntington Avenue, Boston, MA 02115. E-mail: jamieson.jp@gmail.com

situation poses a threat to one's social identity (Schmader, 2002)" (Schmader & Johns, 2003, p. 442). Cognitive resources that could be devoted to task performance are instead expended on processing the information resulting from the activation of the negative stereotype. It is this reduction in working memory capacity that produces the performance debilitation reported in the stereotype threat literature.

The work of Engle and his colleagues (e.g., Engle, 2001; Engle, Tuholski, Laughlin, & Conway, 1999) forms the conceptual basis for Schmader and John's (2003) research. These researchers view working memory as a subset of highly activated long-term memory units, an array of processes that produce and maintain activation of those units, and an executive attention component. The executive attention component of working memory is responsible for maintaining task goals, processing incoming information, and blocking external and internal interference. Engle and his colleagues argued that the domain-free executive attention ability "is important for predicting performance on higher order cognitive tasks (e.g., Engle et al., 1999; Kane et al., 2004)" (Unsworth, Schrock, & Engle, 2004, p. 1303) and is crucial in the performance of inhibition tasks, like the antisaccade task, which require maintenance of the goal in the face of potent environmental distractors.

On the antisaccade task (e.g., Roberts, Hager, & Heron, 1994), participants are asked to fixate a cross that appears in the center of the visual display and to respond to a target presented randomly on one side of the display or the other (see Figure 1). However, before the target appears, a cue (a white square) is presented on the opposite side of the display. Participants are explicitly instructed not to look at this cue, but rather to look to the opposite side of the display where the target will appear. However, there is a reflexivelike tendency to look at the cue that must be inhibited to optimize performance.

To perform the antisaccade task well, one must marshal working memory resources to inhibit the tendency to look at the peripheral cue as well as to move the eyes to the target. As Unsworth et al. (2004) wrote:

> In situations in which the central executive and automatic attentional capture are in opposition, the central executive will direct the focus of attention only when the intent to do so is actively maintained. A momentary lapse in intention will result in automatic attentional capture. Within the antisaccade task, the central executive is needed not only to block automatic attentional capture, but also to effectively direct focus of attention to the correct location. (p. 1318)

Four articles (Kane, Bleckley, Conway, & Engle, 2001; Roberts et al., 1994; Stuyven, Van der Goten, Vandierendonck, Claeys, & Crevits, 2000; Unsworth et al., 2004) report a total of eight experiments using the antisaccade task that test the hypothesis that deficits in working memory deleteriously af-

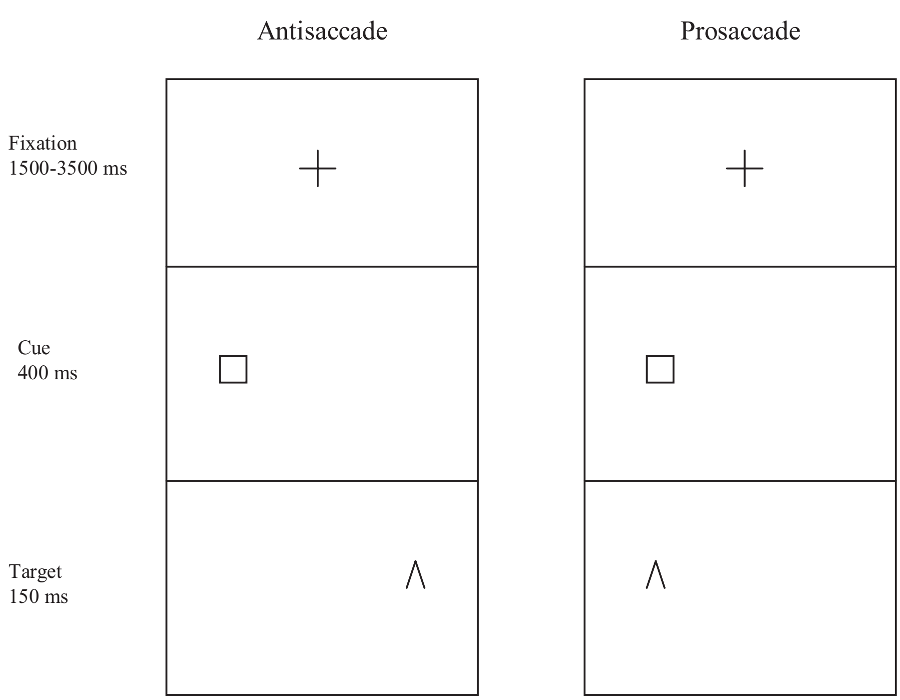

Figure 1. Sequence of events for the antisaccade and prosaccade tasks. Each frame represents what was displayed on the monitor for the period of time shown to the left of the figure. The target appeared in one of three orientations: pointing up (shown), to the right, or to the left.

fect performance on the antisaccadetask.[1] Four of these experiments (Kane et al., 2001, Experiment 2; Unsworth et al., 2004, Experiments 1, 2, and 3) operationalized working memory capacity as an individual-difference variable, operation span, and are, thus, correlational. The other four (Roberts et al., 1994, Experiments 1 and 2; Stuyven et al., 2000, Experiments 1 and 2) incorporated experimental manipulations of working memory capacity (concurrent tasks).

Seven of these eight experiments found that participants who had less working memory capacity (either as an individual-difference variable or through a concurrent task manipulation) produced more reflexive saccades (looking in the wrong direction toward the cue) than participants who had more working memory capacity (weighted mean $r = .43$, combined $Z = 6.84$, $p < .00001$).[2] Eight of the eight experiments found that less working memory capacity led to longer latencies to launch correct saccades (saccades launched toward the target following successful inhibition of the reflexive saccade; see Figure 2) (weighted mean $r = .48$, combined $Z = 7.55$, $p < .00001$). The three experiments that measured corrective saccades (saccades launched toward the target following reflexive saccades; see Figure 2) also found reliable effects (weighted mean $r = .42$, combined $Z = 3.66$, $p = .0001$).

Thus, the antisaccade task is ideally suited for testing Schmader and Johns's (2003) working memory interference account of stereotype threat effects. First, reductions in working memory capacity lead directly to specific predictions about performance on the antisaccade task: If the experience of stereotype threat decreases participants' working memory capacities during task performance, then participants subject to threat should perform more poorly on the antisaccade task than control participants because these participants should be less able to inhibit the tendency to look toward the cue and slower to launch saccades toward the target. Second, by tracking eye movements, it is possible to isolate the various components of performance on this task (e.g., proportion of reflexive responding, saccade launch latencies), which allows the assessment of their relative contribution to the terminal measures, identification of target orientation, and reaction time. Most tasks do not permit this molecular level of analysis (e.g., quantitative Graduate Record Examination [GRE] items). Third, our pilot research shows that participants can easily be convinced that men are better than women at this test of visuospatial skill, as demonstrated by the effects produced on manipulation checks typical of work in this area.

The aim of the present research was to use the antisaccade task to compare the working memory interference account of stereotype threat effects with the mere effort account, an explanation suggested by Harkins's (2006) effort to isolate the process(es) that mediate(s) the effect of evaluation on complex task performance. The potential for evaluation, like stereotype threat, arouses participants' concern about their ability to perform well on the task. In fact, many of the processes proposed to mediate stereotype threat performance effects (e.g., processing interference, withdrawal of effort, and arousal) have also been proposed to explain performance differences in the evaluation-performance domain, and like the stereotype threat literature, researchers have not come to a consensus as to what the mediating mechanism(s) is (are).

Harkins (2006) suggested that the mediating process could be identified through the molecular analysis of performance on a specific task. To this end, Harkins (2006) examined the effect of

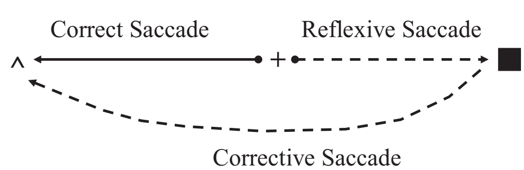

Figure 2. Response maps for different types of saccadic eye movements on antisaccade trials. The solid line represents trials on which participants make a correct saccade toward the target. The broken line represents trials on which participants first make a reflexive saccade and then generate a corrective saccade back toward the target.

[1] Mitchell, Macrae, and Gilchrist (2002) also examined the effect of working memory deficits on antisaccade performance. However, their methodology differed substantially from ours and from that of the other researchers in the area. Mitchell et al. presented their fixation stimulus for a fixed interval (2,000 ms), whereas other research used a randomly determined interval (present research: 1,500 ms–3,500 ms; Kane et al., 2001: 600 ms–2,200 ms; Roberts et al., 1994: 1,500 ms–3,500 ms; Stuyven et al., 2000: 1,500 ms–3,000 ms; Unsworth et al., 2004: 600 ms–2,200 ms). Mitchell et al. (2002) used a cue eccentricity (distance from the fixation point to the cue) of only 4.8°, whereas other research used larger cue eccentricities (present research: 11°; Kane et al., 2001: 11.5°; Roberts et al., 1994: 10.5°; Stuyven et al., 2000: 6.7°; Unsworth et al., 2004: 11.5°). Finally, Mitchell et al. (2002) used a cue that subtended 5.2° of visual angle, whereas other research used cues that were substantially smaller (present research: 0.5°; Kane et al., 2001: 0.6°; Stuyven et al., 2000: 0.4°; Unsworth et al., 2004: 0.6°).

Each of these differences has the effect of making Mitchell et al.'s (2002) task easier than the versions used in the other research. A fixed fixation interval makes the task easier by eliminating the need for a "when" decision so that participants are only left with a "where" decision ("Do I need to launch a saccade to the right or left?"). Smaller cue eccentricities produce faster correct saccades than larger cue eccentricities (except for eccentricities below 2°) (Smyrnis et al., 2002). Finally, Roberts et al. (1994) manipulated cue size (0.4° vs. 2.0° vs. 3.4°) and found that larger cue sizes were associated with better performance than smaller cue sizes. They argued that "larger cues are noticed more easily and quickly, allowing deliberate processes of working memory more opportunity to program an antisaccade"(p. 383). Once again, Mitchell et al. (2002) used a cue that subtended 5.2° of visual angle.

Because Mitchell et al.'s (2002) version of the antisaccade task is much easier than the tasks used in the other research, this may account for the fact that in their first experiment, cognitive load did not affect correct saccade latency. In Experiment 2, Mitchell et al. (2002) increased task difficulty by increasing the number of possible cue locations from two to four (left, right, up, down). Kveraga, Boucher, and Hughes (2002) have shown that correct saccade latencies and keypress responses follow Hick's law: Response time increases as the number of stimulus response alternatives increase. Mitchell et al. (2002) also increased task difficulty by changing the cue "from a single white stimulus to a collection of 60 different color photographs of various objects" (p. 102). With these changes in effect, Mitchell et al. did find that cognitive load significantly increased the latencies of correct saccades. However, given the large number of differences in methodology between this research and that of the other research, we did not include it in our meta-analyses.

[2] In their first experiment, Stuyven et al. (2000) included a control condition and two cognitive load conditions. For both dependent measures, proportion of reflexive saccades and saccade launch latency, we computed the effect size for the contrast between the control condition and each of the load conditions and then averaged the two effect sizes for entry in the meta-analysis.

evaluation on the performance of the remote associates task (RAT), which requires participants to look at sets of three words (e.g., *lapse*, *elephant*, *vivid*) and to generate a fourth word that is related to each word in the given triad (in this case, *memory*; Harkins, 2006). Harkins (2001) has shown that the potential for evaluation produces the typical pattern of performance on this task: Participants who anticipate evaluation by the experimenter solve more simple trials than no-evaluation participants, whereas participants who anticipate experimenter evaluations solve fewer difficult triads than no-evaluation participants.

Harkins's (2006) analysis showed that these findings result from the fact that the potential for evaluation motivates participants to want to do well, which potentiates whatever response is prepotent on the given task. On the RAT, the prepotent response is to generate words that are closely related to one of the three triad members. On simple triads, because the correct answer tends to be a close associate of one of the triad members, participants subject to evaluation perform better than controls. However, on complex triads, the correct answer is remotely associated with each triad member, and the accumulation of the weak activation contributed by each triad member is required for the correct answer to be brought to mind. In this case, the potentiation of the prepotent response, generating close associates, inhibits the activation of the correct answer, resulting in poorer performance by participants subject to evaluation.

Zajonc's (1965) drive theory account of social facilitation effects also accords a central role to prepotent or dominant responses. Drive theory contends that the presence of others produces arousal, which increases drive. Increased drive enhances the probability of the emission of dominant responses, which are likely to be correct on simple tasks but incorrect on difficult ones. In fact, Cottrell (1972) argued that this drive was the result of the participants' apprehension about the fact that they would be evaluated.

Thus, both mere effort and Cottrell's (1972) evaluation apprehension account of social facilitation effects predict that the potential for evaluation will potentiate dominant or prepotent responses. However, in the case of mere effort, this potentiation results from the motivation to perform well, which should also lead to an effort to correct the incorrect response if the participant recognizes that his or her response is incorrect, knows the correct response, and has the opportunity to make it. In contrast, Cottrell's (1968, 1972) modification of Zajonc's drive theory suggests only that the positive or negative anticipations produced by the presence of others nonselectively energize individual performance (i.e., potentiate the dominant response). Of course, on a task like the RAT, we are unable to distinguish between mere effort and evaluation apprehension accounts because even if the participants know that the response is incorrect, they do not know how to correct it.

However, an inhibition task, like the Stroop, does allow us to see the effect of the motivation to correct. The Stroop color-word task (Stroop, 1935) requires participants to name the ink color of a color word. For example, they may see the word *red* printed in blue, and the correct response is blue. The prepotent tendency in this task is to read the color word, and the mere effort account would predict that the potential for evaluation would potentiate this incorrect, prepotent response. However, on inhibition tasks, like the Stroop, the correct answer is quite obvious. So, although the initial tendency for participants subject to evaluation to read the color will be stronger than that of their no-evaluation counterparts, it will be quite clear to these participants that this response is incorrect. Given enough time to counter the effect of response potentiation, the mere effort account predicts that the heightened motivation of the evaluation participants to do well will lead them to produce the correct answer more quickly than the no-evaluation participants, whereas drive theory or evaluation apprehension predicts only response potentiation. Consistent with the mere effort prediction, McFall, Jamieson, and Harkins (2007) found that if required to produce a response in a brief time (1 s or 750 ms), then participants subject to evaluation made more errors than no-evaluation participants, but when given up to 2 s to respond, participants subject to evaluation responded more quickly than no-evaluation participants with no difference in accuracy.[3]

The mere effort account of stereotype threat performance effects argues that stereotype threat should operate like the potential for evaluation in that threat will motivate participants to want to perform well on the task. Thus, stereotype threat should produce the same basic pattern of findings on the antisaccade task that is produced by the potential for evaluation on inhibition tasks like the Stroop. When not given sufficient time to correct for the prepotent tendency (i.e., brief display time), the more motivated stereotype threat participants should be less accurate than controls in their ability to correctly identify target orientation.[4] However, when the display time is increased enough to allow for correction, stereotype threat participants should be able to respond to the target more quickly than controls as a result of increased motivation to perform well.

More specifically, like the working memory interference account, the mere effort account predicts that the participants under stereotype threat will look in the wrong direction, toward the cue, more often than participants in the control group, but for a different reason. The working memory interference account argues that because the experience of stereotype threat diminishes working memory capacity, the participants are less able to inhibit looking in the wrong direction ("foot off the brake"). The mere effort account argues that the motivation to perform well potentiates the prepo-

[3] When studying the effect of motivation on inhibition tasks like the Stroop, it is common to find differences reflected in accuracy when response time is limited (e.g., Hochman, 1967; Pallak, Pittman, Heller, & Munson, 1975) but reflected in speed when the time provided for a response is essentially unlimited (e.g., Huguet, Galvaing, Monteil, & Dumas, 1999; MacKinnon, Geiselman, & Woodward, 1985; O'Malley & Poplawsky, 1971). The correction of a prepotent tendency requires time and cognitive resources. When given limited time to respond, more motivated participants do not have enough time to correct for the tendency to make the prepotent response, but when given more time to respond they do. Because all participants can respond accurately when response time is increased, motivation produces faster responses. We would expect to find a similar pattern of findings on the antisaccade task.

[4] As on the Stroop, to succeed on the antisaccade task, participants must inhibit their prepotent response, in this case, the reflexive tendency to look at the peripherally flashed stimulus. However, the "brief" intervals on the antisaccade task are much briefer than the time periods used in the Stroop. On this task, the "long" response interval gave the participants up to 2 s to produce their responses, whereas the "brief" intervals provided them from 750 ms to 1,000 ms. In contrast, on the antisaccade task, a typical exposure time is only 150 ms (e.g., Roberts et al., 1994), and then the target is immediately masked. The participant does have time following target exposure in which to make a response (e.g., 1,500 ms), but if the target was not seen, the participant has no basis for making a response.

tent response, looking at the cue, which makes stereotype threat participants more likely than control participants to look in the wrong direction ("foot on the gas").

However, the mere effort account predicts that stereotype threat will motivate participants to launch correct and corrective saccades faster, not slower, than control participants. Sereno (1992) noted that correct and corrective saccades represent "an extreme example of a voluntary saccade" (p. 92). They are not produced in response to a stimulus but are instead endogenously generated and launched to a predicted target region (e.g., Crevits & Vandierendonck, 2005; Godijn & Kramer, 2006; Massen, 2004). Thus, if one is motivated to generate correct and corrective saccades as quickly as possible, then more effort should produce faster saccade launch latencies. Consistent with this notion, research has shown that motivation/reward facilitates the generation of volitional eye movements in animal models (e.g., Kawagoe, Takikawa, & Hikosaka, 1998; Nakamura & Hikosaka, 2006).

Finally, after the participants' eyes arrive at the target area, the participant must determine the target orientation and press the appropriate response key. There is no reason to believe that working memory interference would impact this decision-making aspect of trial performance (see Engle & Kane, 2003). Motivation to perform well, however, should make stereotype threat participants try to respond as quickly as possible. Thus, when the participants see the target, the mere effort account would predict that participants subject to stereotype threat would respond more quickly than participants in the control condition.

To test the full range of predictions for the mere effort account, the target display time must be long enough for the motivation to correct to have an opportunity to play a role. If the display period is too short, then participants subject to stereotype threat will not be able to recover quickly enough to see the target. Therefore, we began by parametrically varying the target display time and collected data only on the terminal measures, accuracy and reaction time.

Experiment 1: Antisaccade Task at 150 ms

Because the working memory interference account predicts that the experience of threat will reduce working memory capacity, participants subject to threat should perform more poorly than controls regardless of target display time. Specifically, working memory resources are required to inhibit the prepotent response to look toward the cue and to generate saccades to the target. If working memory is impaired, then participants should look toward the cue more often and take longer to launch saccades to the target than control participants. The mere effort account also predicts that, at some display time, participants in the stereotype threat condition will perform more poorly than participants in the control condition because the participants will not have sufficient time to correct for the prepotent response, looking at the cue. However, at some longer display time, there will be no difference, and at a longer display time, a reversal, as the threat participants do have time to make the correction.

Pilot testing showed that no actual gender differences exist on the antisaccade task when participants were given no experimental manipulations. As noted previously, this work also showed that participants were convinced that the antisaccade task is a measure of visuospatial capacity, which, in turn, is indicative of mathematical ability, a domain in which males are stereotypically superior. As a first step, we set the display time of the target at 150 ms because this exposure time had been used in previous antisaccade research (e.g., Roberts et al., 1994).

Method

Participants

Eighty Northeastern University undergraduate students (40 men and 40 women) participated in this experiment in exchange for partial fulfillment of a course requirement. All participants reported normal vision or corrected-to-normal vision.

Tasks and Apparatus

Participants completed two eye movement tasks, the antisaccade and prosaccade tasks. Each participant was seated in front of a computer monitor in a small room. A Macintosh G5 computer controlled the stimulus presentation and recorded the keypress timing and the accuracy of the responses. Both tasks were presented on a 17-in. (43.18-cm) monitor. Participants' heads were stabilized throughout the experiment by a chin rest positioned 54 cm from the monitor.

In the antisaccade task (see Figure 1), each trial began with the presentation of a fixation cross, subtending 1° of visual angle, in the center of the screen for a randomly determined interval ranging from 1,500 to 3,500 ms. The cue, a white square that subtended 0.5° of visual angle, was then presented 11° to either the left or the right of the fixation cross for 400 ms. When the cue was extinguished, the target, an arrow also subtending 0.5° of visual angle, then appeared on the opposite side of the screen from the cue, 11° from the center fixation cross. The target was presented in one of three orientations: pointing up, to the left, or to the right. The target was displayed for 150 ms, after which a mask, another white square subtending 0.5° of visual angle, appeared in its place. This mask remained until the participant responded with a keypress. If no response was made, then it was removed after 1,500 ms, and the next trial began after a 1,750-ms intertrial interval.

Participants were instructed to look at the fixation cross in the center of the screen until the cue was presented, at which point they were to look away from the cue and indicate the orientation of the target located on the opposite side of the screen as quickly and accurately as possible by pressing the corresponding arrow key on a keyboard (left, up, or right). Cue side (left or right) and arrow direction were randomized across trials.

As shown in Figure 1, the prosaccade task was identical to the antisaccade task except that the target (the arrow) was presented on the same side of the screen as the cue (the white square). Participants were instructed to look toward the cue and identify the orientation of the target that appeared in its place. The prepotent tendency to look toward the peripherally flashed cue is correct on prosaccade trials, whereas on the antisaccade task, this prepotent response is incorrect. Thus, prosaccade trials are structurally similar to antisaccade trials but do not require the inhibition and/or correction of prepotent responses.

Participants completed six practice trials prior to the beginning of each task and then completed 90 antisaccade or prosaccade trials. Task order was counterbalanced across participants. As is common in cognitive load research using the antisaccade task (e.g., Roberts et al., 1994; Stuyven et al., 2000), participants did not receive feedback after each trial.

Procedure

Participants were brought into the lab one at a time and gave written consent. After consent was obtained, verbal and written (on the computer screen) instructions were presented to each participant. When the task had been thoroughly explained and participants had completed the practice trials, the experimenter then implemented the stereotype threat manipulation. In the stereotype threat condition, the experimenter explained,

> The task you are about to complete is a test of visuospatial capacity. This measure is closely linked to math ability. As you may know, there has been some controversy about whether there are gender differences in math and spatial ability. Previous research has demonstrated that gender differences exist on some tasks, but not on others. In our lab, we examine performance on both kinds of tasks. The task on which you are about to participate has been shown to produce gender differences.

The control condition instructions were identical to the stereotype threat instructions except the last sentence read, "The task on which you are about to participate has not been shown to produce gender differences." A similar manipulation has been shown to produce stereotype threat effects in previous research (e.g., Brown & Pinel, 2003; Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Spencer et al., 1999). No specific mention was made as to whether men outperformed women or vice versa, only that gender differences did or did not exist on the task. Participants were expected to infer that women would perform more poorly than men on the basis of the societal stereotype that men are superior to women in mathematical and spatial ability.

Each participant responded to a questionnaire after completion of each saccade task (one after the prosaccade task and one after the antisaccade task). Two questions allowed us to evaluate the effectiveness of the stereotype threat manipulation: To what extent are there gender differences in performance on this task? (1 = *no gender differences* and 11 = *gender differences*) and Who do you believe performs better on this task? (1 = *males perform better*, 6 = *males and females perform the same*, and 11 = *females perform better*). Participants were also asked to rate the extent to which they felt that they could evaluate their performance and the extent to which their performance could be evaluated by the experimenter. Finally, they were asked to rate how interesting the task was, how anxious they felt about their performance, how well they thought they performed, and how much effort they put into the task, all on 11-point scales.

Results

Manipulation Check for Stereotype Threat

The manipulation checks were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (gender: male vs. female) × 2 (task order: antisaccade first vs. prosaccade first) × 2 (task: antisaccade vs. prosaccade) analyses of variance (ANOVAs). Condition, gender, and task order were analyzed as between-subjects effects, and task was analyzed as a within-subjects effect. Participants in the stereotype threat condition reported that gender differences existed to a greater extent ($M = 6.85$, $SD = 3.05$) than participants in the control condition ($M = 1.48$, $SD = 1.38$), $F(1, 72) = 100.50$, $p < .001$, $d = 2.34$. Threatened participants reported that men performed these tasks better than women to a greater extent ($M = 3.78$, $SD = 2.01$) than participants in the control condition ($M = 5.85$, $SD = 1.03$), $F(1, 72) = 33.90$, $p < .001$, $d = 1.39$. These results indicate that the stereotype threat manipulation used in the present experiment was successful. Participants in the threat condition were aware of the negative group stereotype, and women were expected to perform more poorly than men.

Performance

The performance data (accuracy and reaction time measured from the onset of the target) were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (gender: male vs. female) × 2 (task order: antisaccade first vs. prosaccade first) × 2 (task: antisaccade vs. prosaccade) ANOVAs. Condition, gender, and task order were analyzed as between-subjects effects, and task was analyzed as a within-subjects effect.

Accuracy. Replicating past work (e.g., Roberts et al., 1994; Stuyven et al., 2000; Unsworth et al., 2004), participants correctly reported the orientation of a greater percentage of targets in the prosaccade task ($M = 98.70\%$, $SD = 2.30\%$) than in the antisaccade task ($M = 90.10\%$, $SD = 8.20\%$), $F(1, 72) = 110.31$, $p < .001$, $d = 2.49$. This finding is expected because on the prosaccade task, unlike the antisaccade task, good performance does not require the inhibition of the prepotent response tendency.

Each of the other reliable effects in the overall analysis, gender, $F(1, 72) = 4.39$, $p < .05$, $d = 0.50$; Condition × Gender, $F(1, 72) = 7.71$, $p < .01$, $d = 0.65$; and Task × Gender, $F(1, 72) = 9.00$, $p < .01$, $d = 0.70$, must be interpreted in the context of the Condition × Task × Gender interaction, $F(1, 72) = 6.00$, $p < .02$, $d = 0.58$ (see Figure 3). A series of contrasts was used to decompose this three-way interaction (Kirk, 1995).

As can been seen in Figure 3, on the antisaccade task, females in the stereotype threat condition performed more poorly ($M = 84.30\%$, $SD = 10.50\%$) than participants in any other condition (M female/no stereotype threat = 91.50%, SD female/no stereotype threat = 6.80%), $F(1, 72) = 17.28$, $p < .001$, $d = 0.98$; (M male/stereotype threat = 93.60%, SD male/stereotype threat = 4.30%), $F(1, 72) = 28.83$, $p < .001$, $d = 1.28$; (M male/no stereotype threat = 91.20%, SD male/no stereotype threat = 7.40%), $F(1, 72) = 15.87$, $p < .001$, $d = 0.95$, which did not differ from each other ($ps > .20$). Thus, only the women subject to

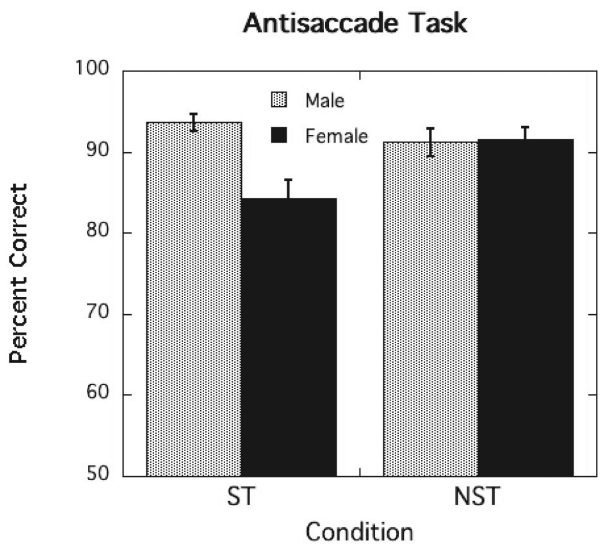

Figure 3. Accuracy for correctly identifying the target on antisaccade trials at a display time of 150 ms as a function of stereotype threat (ST) condition and gender in Experiment 1. NST = no stereotype threat.

stereotype threat experienced performance decrements in their ability to accurately report target orientation.

There were no reliable differences among the conditions on the prosaccade task ($ps > .20$). The largest difference between the pairs of these means was only 1.1%.

Reaction time. In this experiment and all that follow, only reaction times for correct responses were submitted to analysis. Replicating past work (e.g., Roberts et al., 1994; Stuyven et al., 2000), we found a main effect for task. Participants responded to targets more quickly in the prosaccade task ($M = 468.79$ ms, $SD = 63.28$ ms) than in the antisaccade task ($M = 509.06$ ms, $SD = 77.11$ ms), $F(1, 72) = 57.19$, $p < .001$, $d = 1.81$. There was also a marginal gender main effect, $F(1, 72) = 3.74$, $p < .06$, $d = 0.45$, with men tending to respond more quickly ($M = 475.45$ ms, $SD = 69.06$ ms) than women ($M = 502.40$ ms, $SD = 67.91$ ms) across tasks.

A Condition × Task Order interaction, $F(1, 72) = 5.54$, $p < .05$, $d = 0.56$, resulted from the fact that in the stereotype threat condition, it did not matter whether the antisaccade task or the prosaccade task came first. However, in the control condition, participants completing the prosaccade task first responded more quickly across both tasks than participants who responded to the antisaccade task first. There was also a significant Condition × Task interaction, $F(1, 72) = 4.85$, $p < .05$. This interaction resulted from the fact that participants in both the stereotype threat and control conditions responded to the target reliably more quickly on prosaccade trials than on antisaccade trials, but the difference was greater in the stereotype threat condition.

However, as a result of differences in accuracy, the means of participants in the stereotype threat and control groups are based on different numbers of trials. Thus, any analysis comparing these groups on a measure of reaction time is suspect. In any event, these interactions were not predicted, nor do they alter the interpretation of results in the present experiment.

Ancillary Measures

Analysis of the difficulty measure revealed a Gender × Task interaction, $F(1, 72) = 4.52$, $p < .05$, $d = 0.49$. A Tukey's Honestly Significant Difference test (Kirk, 1995) showed that, on the prosaccade task, women ($M = 1.9$, $SD = 1.33$) and men ($M = 1.75$, $SD = 0.97$) did not differ in their ratings of task difficulty ($p > .20$), but female participants in the stereotype threat condition rated the antisaccade task as more difficult ($M = 6.15$, $SD = 2.80$) than male participants ($M = 4.05$, $SD = 2.59$; $p < .05$). The gender, $F(1, 72) = 6.02$, $p < .05$, $d = 0.58$, and task, $F(1, 72) = 50.99$, $p < .001$, $d = 1.67$, main effects must be interpreted in the context of this interaction.

Participants did not differ in their ratings of the extent to which the experimenter knew how well they performed or in their ratings of the extent to which they could evaluate their own performances ($ps > .20$). Analyses of the self-reports of anxiety experienced during task performance, task interest, how well they thought they performed, and effort were also all nonsignificant ($ps > .20$).

Discussion

Consistent with the success of the stereotype threat manipulation, participants in the stereotype threat condition reported that there were gender differences in the performance of the task and that male performance was superior to that of female performance. Although male performance was unaffected by this manipulation, women in the stereotype threat condition were less accurate in their identification of target orientation than women in the control condition and men in either condition.

This finding is consistent with the working memory interference account of stereotype threat effects (Schmader & Johns, 2003), which would argue that cognitive resources that could be devoted to inhibiting the tendency to look at the cue are instead expended on processing the information resulting from the activation of the negative stereotype. It is this reduction of working memory capacity that accounts for the performance debilitation. However, this finding is not necessarily inconsistent with the mere effort account. If the 150-ms presentation time did not provide sufficient time for the stereotype threat participants to correct for the potentiation of the prepotent response, mere effort would also predict debilitated performance. Thus, the results of Experiment 1 are not conclusive.

Experiment 2: Antisaccade Task at 250 ms

In Experiment 2, the target display time was increased from 150 ms to 250 ms. At this exposure time, the predictions of the working memory interference and mere effort accounts diverge. Mere effort predicts that participants subject to threat will report target orientation either as quickly as or more quickly than control participants because the additional time should provide them with the opportunity to correct for the potentiated prepotent response and still see the target, and they are motivated to do so. If the experience of stereotype threat impairs working memory capacity, then stigmatized individuals should perform more poorly than the control group on the antisaccade task regardless of display time.

Method

Participants

Because male performance did not differ as a function of the stereotype threat manipulation in Experiment 1, only women were used in the subsequent experiments. Thirty-six Northeastern University undergraduate students participated in this experiment in partial fulfillment of a course requirement. All participants reported normal or corrected-to-normal vision.

Materials and Apparatus

The saccade tasks were identical to those described in Experiment 1 except that the target display time was changed from 150 ms to 250 ms.

Procedure

The procedure was identical to that in Experiment 1, including all manipulations. Questionnaire items were identical to those used in Experiment 1. However, as task order did not affect participants' questionnaire responses in Experiment 1, only one questionnaire was given to each participant upon completion of the first block of trials.

Results

Manipulation Check for Stereotype Threat

Manipulation check and questionnaire items were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task: antisaccade vs. prosaccade) ANOVAs, with condition and task as between-subjects effects. Participants in the stereotype threat condition reported that gender differences existed on the saccade tasks to a greater extent ($M = 7.22$, $SD = 2.78$) than participants in the control condition ($M = 1.83$, $SD = 1.38$), $F(1, 32) = 53.08$, $p < .001$, $d = 2.58$. The threatened participants also reported that men performed these tasks better than women to a greater extent ($M = 3.22$, $SD = 2.46$) than the participants in the control condition ($M = 6.39$, $SD = 0.92$), $F(1, 32) = 28.01$, $p < .001$, $d = 1.85$.

Performance

The performance data were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task order: antisaccade first vs. prosaccade first) × 2 (task: antisaccade vs. prosaccade) ANOVAs, with condition and task order as between-subjects factors and task as a within-subjects factor. The dependent measures were accuracy in identifying the target orientation and reaction time, as measured from the onset of the target.

Accuracy. As in Experiment 1, participants correctly reported the orientation of the target more accurately in the prosaccade task ($M = 98.67\%$, $SD = 2.02\%$) than in the antisaccade task ($M = 96.20\%$, $SD = 3.20\%$), $F(1, 32) = 36.41$, $p < .001$, $d = 2.14$. No other effects were reliable. In the present experiment, the target in the antisaccade task was more easily identified (stereotype threat: $M = 96.20\%$, $SD = 3.20\%$) than in Experiment 1 (stereotype threat: $M = 84.30\%$, $SD = 10.50\%$). This increase in accuracy likely results from the increase in display time.

Reaction time. As in Experiment 1, participants responded to target orientation more quickly in the prosaccade condition ($M = 462.73$ ms, $SD = 72.45$ ms) than in the antisaccade condition ($M = 524.63$ ms, $SD = 82.16$ ms), $F(1, 32) = 35.88$, $p < .001$, $d = 2.14$.

More important, there was a main effect for stereotype threat, $F(1, 32) = 19.52$, $p < .001$, $d = 1.58$. Participants subject to stereotype threat responded more quickly than control participants. Separate contrasts (Kirk, 1995) show that this pattern characterized performance on both antisaccade (stereotype threat: $M = 480.62$ ms, $SD = 59.85$ ms; no stereotype threat: $M = 568.64$ ms, $SD = 78.82$ ms), $F(1, 32) = 9.06$, $p < .01$, $d = 1.06$, and prosaccade trials (stereotype threat: $M = 420.62$ ms, $SD = 32.17$ ms; no stereotype threat: $M = 504.83$ ms, $SD = 77.57$ ms), $F(1, 32) = 8.30$, $p < .01$, $d = 1.01$. No other effects were reliable.

Ancillary Measures

Participants rated the antisaccade task ($M = 3.56$, $SD = 2.23$) as more difficult than the prosaccade task ($M = 1.83$, $SD = 1.54$), $F(1, 32) = 6.89$, $p = .013$, $d = 0.93$. There were no reliable differences in the participants' ratings of self-reported anxiety, effort, how well they performed, or the potential for experimenter or self-evaluation ($ps > .20$).

Discussion

Once again, the manipulation checks suggest that the stereotype threat manipulation was successful. Participants exposed to the stereotype threat manipulation believed that the saccade tasks were diagnostic of their math ability and that these tasks were gender biased.

In Experiment 1, on the antisaccade task, participants under stereotype threat were less accurate in their identifications of target orientation than participants in the control condition. However, in Experiment 2, in which the display time of the target was increased from 150 ms to 250 ms, participants in the stereotype threat condition were as accurate in their identifications as participants in the control group, and reported target orientation reliably more quickly. This finding is inconsistent with an explanation that relies solely on working memory interference, which would predict that participants subject to threat should perform more poorly than controls on the antisaccade task at all target display times. However, the finding is consistent with the mere effort account, which argues that the 100-ms increase in display time provided the more motivated stereotype threat participants with enough time to correct for the potentiation of the prepotent response, accounting for the fact that they identified the orientation of the target more quickly than the control group.

Participants in the stereotype threat condition in Experiment 2 also reported target orientation on prosaccade trials more quickly than control participants. In Experiment 1, although stereotype threat participants did respond slightly more quickly than controls, the difference was not significant. The pattern of results of Experiment 1 could have resulted from a shift in participants' decision process as to what constituted an optimal approach to performance. At 150 ms, participants subject to threat may have been more motivated than control participants, but the briefer display time may have led them to place a premium on responding accurately instead of quickly. At 250 ms, because the target was easier to see, participants may have shifted their focus to responding quickly. This finding is also consistent with the mere effort account in that it suggests that participants under stereotype threat are more motivated than control participants, which is reflected in faster reaction times on the prosaccade task.

Although the performance data from Experiments 1 and 2 are consistent with the mere effort account, this support comes from the accuracy and reaction time measures that are collected at the end of each trial. The mere effort and working memory interference accounts can be examined in detail by dividing the individual antisaccade trial into three phases. At the outset of the trial, the cue is presented. At this point, the participant is either successful in inhibiting an eye movement to the cue or is unsuccessful. The working memory interference and the mere effort accounts each predict that the participants under stereotype threat will look in the wrong direction, toward the cue, more often than participants in the control group, but for different reasons. The working memory interference account argues that because the experience of stereotype threat diminishes working memory capacity ("foot off the brake"), the participants are less able to inhibit looking in the wrong direction, whereas the mere effort account argues that the motivation to perform well potentiates the prepotent response ("foot on the gas"), looking at the cue, which makes stereotype threat participants more likely than control participants to look in the wrong direction.

In the second phase, the participant launches one of two types of saccades to the target. Once again, correct saccades are those that are launched toward the target following successful inhibition of a reflexive saccade. Corrective saccades are those that are launched toward the target following reflexive saccades (see Figure 2). Working memory resources are required not only to inhibit reflexive saccades but also to generate correct and corrective saccades. Thus, to the extent that females subject to stereotype threat suffer from reduced working memory capacity, they should not only produce more reflexive saccades, but, as noted previously, also launch correct and corrective saccades more slowly than female controls. However, motivation to perform well should have the opposite effect on saccade launch latencies. Participants subject to stereotype threat should produce faster correct and corrective saccades than control participants.

Finally, in the third phase, the participants see the target and make the keypress. Once again, there is no reason to believe that a working memory deficit would impact this decision-making aspect of trial performance, whereas the mere effort account predicts that participants subject to stereotype threat will respond more quickly than participants in the control condition. These predictions were tested in Experiment 3.

Experiment 3: Antisaccade Task and Eye Tracking

In Experiment 3, the type and latency of participants' eye movements were recorded during performance of the antisaccade task. Our analysis focused on the three types of eye movements produced in the course of task performance (see Figure 2): reflexive saccades; corrective saccades; and correct saccades. We also isolated the effect of the motivation to make the keypress. The isolated reaction time differs from the usual reaction time in that the latter represents the time taken to respond measured from when the target appears (400 ms after cue presentation; e.g., Kane et al., 2001; Roberts et al., 1994). However, in some cases, the eyes will have arrived in the target area after the target has appeared, whereas in other cases, the eyes will have arrived before the target appears. In the latter cases, no adjustment is necessary. Reaction time measured from the 400-ms mark makes sense, but in the former, the reaction time includes time when the participants could not have responded. Therefore, to isolate the keypress component, in those cases in which the saccade reached the target after its presentation, we subtracted the amount of time by which it came after from the reaction time.

We argue that an examination of the contribution of each of these components of performance will aid us in understanding the process(es) that produce(s) the results for the terminal performance measures, accuracy of target orientation, and overall reaction time. In Experiment 3, we once again used a 250-ms display time, which allowed us to see whether we could replicate the performance findings from Experiment 2.

Method

Participants

Thirty-six Northeastern University students participated in this experiment in exchange for class credit. All participants reported normal vision or corrected-to-normal vision, but none wore eyeglasses, which interfered with eye tracking.

Tasks and Apparatus

The tasks were identical to those described in Experiment 2 except that participants completed the saccade tasks in blocks of 74 trials rather than 90. Each participant was seated in front of a computer screen in a cubicle. A Macintosh computer controlled the stimulus presentation and recorded the keypress, timing the accuracy of the responses and eye movement measures. Head position was stabilized with a chin rest throughout the course of each task. Eye movement data were collected using a Dr. Bouis infrared oculometer (Dr. Bouis Devices, Karlsruhe, Germany) interfaced with the computer. The oculometer measured eye position by projecting an infrared light into the eye at an intensity limited to 3×10^{-4} W/cm^2 and calculating the angular disparity between pupil reflectance and maximum corneal reflectance. The resolution was only limited by the fact that the infrared light illuminating the eye was pulsed at 4 kHz. Thus, the oculometer allowed eye position to be tracked with a resolution of 0.1°, which is ideal for measuring small eye movements such as saccades (Bach, Bouis, & Fischer, 1983). To ensure that the oculometer remained calibrated for luminance and spatial accuracy throughout the experiment, an onscreen calibration test was presented every 20 trials.

Procedure

Manipulation check and questionnaire items were identical to those used in Experiment 2. The eye tracker, however, required changes in the procedure. Because of the calibration test, the experimenter was required to remain in the same room as the participant during performance of the antisaccade and prosaccade tasks. However, he was seated so that the screen remained out of his sight. Participants were aware that the experimenter could not see the computer screen during task performance and were instructed to inform the experimenter when the calibration test screen appeared. When the calibration screen appeared, the experimenter then repositioned himself to conduct the calibration test. Upon completion of the first saccade task, participants filled out a questionnaire in an adjacent room.

Data Preparation

Filters were used prior to data analysis to ensure that eye movements recorded by the eye tracker represented responses to the stimuli and were not random movements. Prior to beginning each trial, participants were required to fixate on a center fixation cross. If in a period of 200 ms preceding the onset of the cue a participant's eye position did not vary by more than 2.82° (50 pixels), then that trial was considered as having a valid baseline. If gaze strayed more than 2.82° from the center of the central position during this 200-ms pretrial window, then that trial was considered as having a bad baseline and was excluded from the analysis. A total of 4.1% of the total number of trials across the prosaccade and antisaccade tasks were excluded due to bad baselines.

Trials on which participants initiated saccades in 80 ms or less were considered anticipatory (e.g., Crevits & Vandierendonck, 2005; Ford, Goltz, Brown, & Everling, 2005) and were excluded from the analyses. Additionally, saccades beginning at 1,000 ms or more were excluded from the data analyses because these eye movements could not have been initiated in response to either the cue or the target because both stimuli had been previously extinguished. Using these criteria resulted in the exclusion of another 8.4% of the trials. Thus, a total of 12.5% of trials were excluded from the analyses as a result of poor baselines and threshold and limit violations. The percentage of excluded trials did not differ by condition ($ps > .20$). In addition, previous antisaccade research using eye tracking measures has excluded approximately the same percentage of trials (e.g., Kane et al., 2001; Unsworth et al., 2004).

Eye movements were classified as saccades if participants shifted their gaze position by more than 4.25°; however, movements less than 4.25° were uncommon as participants exhibited a tendency to generate consistent saccadic movements ($M = 10.5°$, $SD = 3.03°$) to either the target or the cue, which were each located 11° from the center of the computer screen. Participants' average saccade velocity for an 11° eye movement ($M = 221°/s$) fell below peak human saccade velocity for 11° eye movements and was within the normal range for eye movements of this magnitude (e.g., Montagnini & Chelazzi, 2005).

Results

Manipulation Checks for Stereotype Threat

Responses to the stereotype threat manipulation checks and questionnaire items were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task: antisaccade vs. prosaccade) ANOVAs, with condition and task as between-subjects factors. Once again, the stereotype threat manipulation was successful: Participants in the stereotype threat condition indicated that gender differences on this task existed to a greater extent ($M = 6.39$, $SD = 3.55$) than participants in the no-stereotype threat condition ($M = 2.67$, $SD = 2.50$), $F(1, 32) = 12.96$, $p = .001$, $d = 1.28$. Stereotype threat participants also reported that men performed better than women on this task to a greater extent ($M = 4.00$, $SD = 2.20$) than participants in the control condition ($M = 6.11$, $SD = 1.37$), $F(1, 32) = 11.96$, $p = .002$, $d = 1.22$.

Performance

The performance data were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task: antisaccade vs. prosaccade) × 2 (task order: antisaccade first vs. prosaccade first) ANOVAs, with condition and task order as between-subjects factors and task as a within-subjects factor. The dependent measures for these analyses were accuracy and reaction times measured from target onset.

Accuracy. Participants were more accurate on prosaccade trials ($M = 99.90\%$, $SD = 3.00\%$) than on antisaccade trials ($M = 97.30\%$, $SD = 2.50\%$), $F(1, 32) = 40.76$, $p < .0001$, $d = 2.27$. Replicating the pattern of results from Experiment 2 in which a 250-ms display time was also used, participants did not differ in their ability to identify target orientation as a function of stereotype threat ($p > .20$).

Reaction time. As in Experiment 2, participants responded to target orientation more quickly on prosaccade trials ($M = 439.57$ ms, $SD = 82.71$ ms) than on antisaccade trials ($M = 477.30$ ms, $SD = 79.53$ ms), $F(1, 32) = 11.03$, $p < .002$, $d = 1.19$. As previously noted, the antisaccade task requires the inhibition of a prepotent response, whereas the prosaccade task does not. Thus, slower reaction times are expected on the antisaccade task.

Again, as in Experiment 2, which also used a target display time of 250 ms, participants subject to stereotype threat identified the target orientation more quickly ($M = 407.29$ ms, $SD = 63.98$ ms) than control participants ($M = 504.19$ ms, $SD = 70.30$ ms), $F(1, 32) = 30.74$, $p < .001$, $d = 1.96$. As in Experiment 2, separate contrasts (Kirk, 1995) showed that this pattern characterized performance on both antisaccade (stereotype threat: $M = 431.09$ ms, $SD = 70.88$ ms; no stereotype threat: $M = 518.65$ ms, $SD = 63.34$ ms), $F(1, 32) = 29.53$, $p < .001$, $d = 1.91$, and prosaccade trials (stereotype threat: $M = 383.50$ ms, $SD = 47.06$ ms; no stereotype threat: $M = 489.74$ ms, $SD = 75.54$ ms), $F(1, 32) = 43.47$, $p < .001$, $d = 2.34$.

Eye Movement Measures

Analyses were conducted on the three types of saccades produced on the antisaccade task—reflexive saccades, corrective saccades, and correct saccades (see Figure 2)—and on the eye movement data for the prosaccade task. Adjusted reaction time data (reaction times adjusted for time of arrival at the target area) for antisaccades and prosaccades were also analyzed.

Reflexive saccades. The percentage and latency of reflexive saccades were analyzed in 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task order: antisaccade first vs. prosaccade first) ANOVAs. This analysis included all the trials that met the inclusion criteria, whether the trial ended with a correct response or not. Both the working memory interference and mere effort accounts predict that participants in the stereotype threat condition will launch more reflexive saccades than participants in the control condition. These predictions hold whether the trial culminates in a correct response or not.

Consistent with the predictions of these accounts, participants in the stereotype threat condition launched reflexive saccades on a greater percentage of the trials (M = 59.90%, SD = 25.10%) than control participants (M = 26.90%, SD = 17.00%), $F(1, 32) = 20.54$, $d = 1.58$. There was also a tendency for stereotype threat participants to launch these saccades more quickly (M = 131.56 ms, SD = 18.28 ms) than participants in the control condition (M = 156.76 ms, SD = 64.26 ms), $F(1, 32) = 2.85$, $p = .12$, $d = 0.61$.

Corrective saccades. The latencies of corrective saccades were analyzed in a 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task order: antisaccade first vs. prosaccade first) ANOVA. This analysis included the trials that met the inclusion criteria and were correctly answered because we are attempting to account for differences in reaction time on trials for which the response was correct. In fact, however, as noted above, the participants correctly identified the orientation of the target on 97.3% of the trials, and stereotype threat did not affect their ability to do so ($p > .20$). As a result, including the few incorrect trials makes no difference in the pattern of the results. This analysis revealed that participants subject to stereotype threat launched corrective saccades more quickly (M = 370.49 ms, SD = 53.18 ms) than control participants (M = 418.44 ms, SD = 67.43 ms), $F(1, 32) = 5.48$, $p < .03$, $d = 0.82$ (see Figure 4). This finding is consistent with the mere effort account, but not an account relying solely on working memory interference.

Correct saccades. The latencies for correct saccades were analyzed in a 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task order: antisaccade first vs. prosaccade first) ANOVA. Once again, we excluded the few trials on which the participants answered incorrectly. Participants subject to stereotype threat launched correct saccades more quickly (M = 296.55 ms, SD = 66.11 ms) than control participants (M = 376.99 ms, SD = 75.86 ms), $F(1, 32) = 10.86$, $p < .003$, $d = 1.15$ (see Figure 4). This finding is consistent with the mere effort account and not with a working memory interference account that relies solely on a reduction of working memory capacity during task performance.

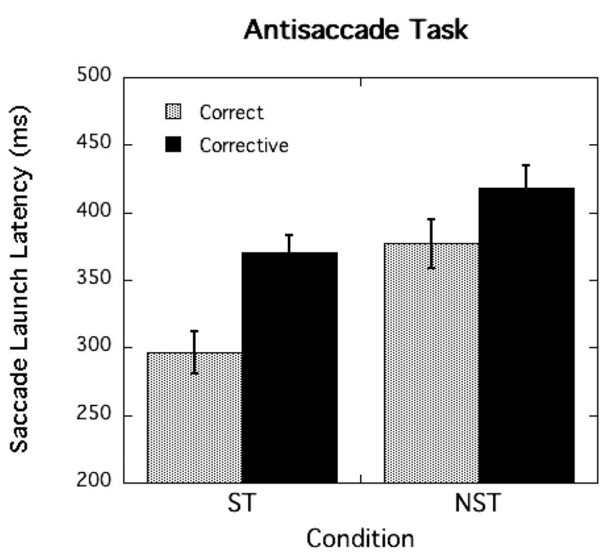

Figure 4. Saccade launch latencies for correct and corrective saccades as a function of stereotype threat (ST) condition in Experiment 3. NST = no stereotype threat.

Prosaccades. On the prosaccade task, the cue and target appear on the same side. Prosaccades are eye movements launched in the direction of the cue/target on this task. The latencies of these saccades were analyzed in a 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task order: antisaccade first vs. prosaccade first) ANOVA. Once again, we used only trials that met the inclusion criteria and were answered correctly. However, participants answered correctly on 99.9% of the prosaccade trials. In fact, out of all the prosaccade trials, only two responses were in error. This analysis revealed a tendency for participants in the stereotype threat condition to launch prosaccades more quickly (M = 127.34 ms, SD = 14.65) than controls (M = 137.18 ms, SD = 25.45), $F(1, 32) = 3.69$, $p = .06$, $d = 0.68$.

Adjusted reaction times. As described previously, to test the hypothesis that the motivation to press the key could contribute to performance on the antisaccade task, we had to take into account the time at which the participant's eyes arrived at the target area. If the eyes arrived before the target even appeared (400 ms from the beginning of the trial), then no adjustment was necessary. The participant was looking at the target area when the target appeared (at the 400-ms mark) and reaction time, as measured from the 400-ms mark until the keypress was appropriate. However, if, for example, the eyes did not reach the target until the 450-ms mark, then reaction time measured from the 400-ms mark would include 50 ms in which the participant could not have responded. Therefore, to isolate the keypress component, in those cases in which the participant's eyes arrived at the target area prior to the target's appearance, we left the reaction time unchanged (i.e., measured from the 400-ms mark in the trial). In those cases in which the saccade reached the target after its presentation, we subtracted the amount of time by which it came after from the reaction time. This procedure was followed for each trial for each person, and the resultant adjusted reaction time scores were averaged for each person.

The adjusted reaction times were analyzed in a 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task order: antisaccade first vs. prosaccade first) × 2 (type of saccade: corrective vs. correct) ANOVA. The adjusted reaction times for participants subject to stereotype threat were significantly faster (M = 381.33 ms, SD = 58.02 ms) than those for participants in the control group (M = 453.67 ms, SD = 84.81 ms), $F(1, 31) = 10.06$, $p < .01$, $d = 1.12$. This finding is consistent with the mere effort account and would not be predicted by working memory interference alone.

This analysis also revealed a main effect for type of response. Adjusted reaction times for corrective saccades were longer (M = 430.12 ms, SD = 85.20 ms) than for correct saccades (M = 411.08 ms, SD = 78.25 ms), $F(1, 31) = 6.29$, $p < .02$, $d = 0.90$. There are at least two features of these saccades that could produce this effect. First, corrective saccades cover twice as much distance as correct saccades in that the former are launched from the location of the incorrect cue, whereas the latter are launched from the middle of the screen. It is possible that a longer recovery period is required after a longer saccade. Second, on trials on which participants make correct saccades, they tend to arrive at the target area before the target appears, whereas on trials on which they make corrective saccades, they tend to arrive after the target has appeared. It is possible that arriving before the target appears provides some time for response preparation. Certainly, both launching the eye movement and deciding to press the key require central executive processing capacity. Completing the eye movement may free resources that then are more quickly available to ready the keypress.

To evaluate these possibilities, we looked at adjusted reaction times for correct and corrective saccades as a function of time of arrival (before vs. after target presentation). When one holds constant time of arrival, there is virtually no difference in adjusted reaction times between correct and corrective saccades. When the participants' eyes arrived at the target site early, the adjusted reaction time for corrective saccades was 411.11 ms, whereas for correct saccades, it was 410.07 ms. When the eyes arrived at the target site after the target had appeared, the adjusted reaction times were longer but virtually identical for corrective ($M = 438.57$ ms) and correct saccades ($M = 436.25$ ms). These findings suggest that correct saccades have faster adjusted reaction times than corrective saccades because the eyes are more likely to arrive at the target site prior to target presentation in the former case, and this early arrival may confer some advantage in response preparation.

Although there was only a 10-ms difference in the latency to launch a prosaccade, participants in the stereotype threat condition pressed the response key an average of 100 ms faster than participants in the control group. This response advantage cannot be attributed to eye movement because the combination of the latency to launch and travel time brought the stereotype threat and control participants to the target area at the same time (stereotype threat: $M = 188.05$ ms, $SD = 19.68$ ms; no stereotype threat: $M = 194.02$ ms, $SD = 27.41$ ms, $F < 1$), over 200 ms before the target even appeared. These findings suggest that participants saw the target at the same time, at the 400-ms mark. Thus, it was the motivation to press the key to make the response that produced the reaction time difference between the conditions on the prosaccade trials. This finding is also consistent with the notion that participants in the stereotype threat condition are motivated to perform well.

Ancillary Measures

Replicating the findings from Experiments 1 and 2, participants rated the antisaccade task as more difficult ($M = 2.89$, $SD = 1.78$) than the prosaccade task ($M = 1.78$, $SD = 1.17$), $F(1, 32) = 4.64$, $p = .04$, $d = 0.77$. There were no reliable differences in the participants' ratings of self-reported anxiety, effort, how well they performed, or the potential for experimenter or self-evaluation ($ps > .20$).

Discussion

As in Experiments 1 and 2, the stereotype threat manipulation was successful. Women exposed to the stereotype threat manipulation indicated that the saccade tasks were gender biased and expected men to outperform women. Also, replicating the finding from Experiment 2, women subject to stereotype threat responded to target orientation as accurately but more quickly than controls.

The eye tracking data allowed us to examine the processes that produced this outcome. Consistent with the predictions of both the working memory interference and mere effort accounts, on the antisaccade task, participants subject to the threat manipulation launched more reflexive saccades than controls (60% vs. 27%, see Table 1). On these trials, the threat group also launched corrective saccades more quickly than controls (373.40 ms vs. 440.00 ms, see Figure 4). In addition, on trials on which participants were successful in inhibiting the tendency to launch a reflexive saccade, participants subject to stereotype threat launched correct saccades faster than participants in the control group (296.55 ms vs. 376.99 ms, see Figure 4). These findings are consistent with the mere effort account but are inconsistent with the working memory interference account. Finally, finding faster adjusted reaction times for stereotype threat than control participants is not predicted by the working memory interference account but is consistent with the mere effort account.

These eye tracking data also make clear exactly why participants subject to stereotype threat outperform control participants, despite the fact that the former participants look the wrong way more than twice as often as the control participants. On 27% of the trials, both participants subject to threat and controls look in the incorrect direction, toward the cue (see Table 1). On this subset of trials, stereotype threat participants generate corrective saccades more quickly than control participants. On 40% of the trials, both threat participants and controls are able to inhibit the prepotent response and do not look toward the cue (see Table 1). On this subset of trials, the threat participants also launch correct saccades more quickly than controls. Thus, on 67% of the trials, stereotype threat participants launch saccades toward the target more quickly than control participants.

On the remaining 33% of the trials, participants subject to threat launch reflexive saccades followed by corrective saccades, whereas control participants are able to inhibit this response and launch correct saccades. On this subset of trials, threat participants launch corrective saccades in approximately the same amount of time ($M = 373.40$ ms) that it takes control participants to launch correct saccades ($M = 376.99$ ms). However, participants in the stereotype threat condition have to move their eyes twice as far

Table 1
Percentage of Correct and Corrective Antisaccade Trials by Condition

	Correct		Corrective	
Condition	*M*	*SD*	*M*	*SD*
Stereotype threat	40.10	22.07	59.90	25.10
No stereotype threat	73.10	16.80	26.90	17.00

Note. Correct saccades refer to eye movements directed to the target location, whereas corrective saccades refer to eye movements launched to the target location that were preceded by reflexive saccades made to the cue.

(22°) as control participants (11°) to see the target because threat participants start their corrective saccades at the cue location, not at the center of the screen. As a result, the stereotype threat participants arrive at the target area on average 26.32 ms after the control participants. However, the reaction times of stereotype threat participants on trials when they launch corrective saccades is still faster (M = 453.85 ms, SD = 90.25 ms) than the reaction times of participants in the control condition on trials when they launch correct saccades (M = 503.52 ms, SD = 61.79 ms), $F(1, 31) = 10.86$, $p < .01$, $d = 1.19$. The analysis for reaction times adjusted for arrival time shows that this 50-ms advantage for the stereotype threat participants is a result of the fact that stereotype threat participants are more motivated to press the key than the control participants, which more than makes up for their late arrival at the target area.

Thus, on 67% of the trials, the stereotype threat participants arrive at the target area before the control participants, and, as shown by the adjusted reaction time analysis, they also respond to the target more quickly. On the remaining 33% of the trials, threat participants arrive at the target area after the control participants, but their motivation to press the key makes up for their late arrival.

The eye tracking data also provide more insight into the findings of Experiment 1. Specifically, in Experiment 1, in which a display time of 150 ms was used, women in the control condition identified the orientation of the target more accurately than women in the stereotype threat condition (no stereotype threat: M = 91.50%; stereotype threat: M = 84.30%, see Figure 2). Experiment 3 suggests that this effect is produced by the trials on which the stereotype threat participants arrive at the target area after the participants in the control condition. Of course, our analysis suggests that if the stereotype threat participants could see the target, then their greater motivation would lead them to press the appropriate key faster than the control participants, but they cannot respond to what they cannot see. That is, if they arrive too late to see the signal, then their faster keypress can make no difference.

To test this hypothesis, we used the eye movement results from Experiment 3 (250-ms display time) to estimate the percentage of the signals these participants would have missed if they had only 150 ms to see the target. To do so, for each individual in Experiment 3, we computed the time at which their eyes arrived at the target site on a trial-by-trial basis. However, after the eyes move to the target site, some amount of time is required to recognize the signal, which must also be taken into account.

Processing time, or low-level sensory encoding, can depend on many variables, such as the size, complexity, and contrast of the target, in addition to environmental factors, such as whether participants are light or dark adapted and how well practiced they are (see Breitmeyer, 1984). To obtain an estimate of processing time in our particular setting, we conducted an experiment using the same room and general setup as was used in Experiment 1. However, because we were interested in the amount of time it takes to recognize the orientation of the arrow, the arrow always appeared in the middle of the screen, replacing the fixation cross after a randomly determined interval of from 1,500 ms to 3,500 ms.

The participants were given the same instructions as those used in all of our experiments (i.e., do your best to report the orientation of the arrow as quickly and as accurately as you can). Only women took part, but we said nothing about gender differences in performance on this task. Our assumption is that although there are most certainly individual differences in processing time, motivation should not affect visual processing time. Consistent with this notion, previous research using a visual recognition task suggests that "the visual system processes new stimuli at a speed and with a number of stages that cannot be compressed" (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001, p. 171). Thus, low-level sensory encoding is a feed-forward process that is unaffected by motivation as long as the individual is attending to the stimulus. In our studies, the relatively high level of performance by both stereotype threat and no-stereotype threat participants suggests that they were attending to the tasks throughout all experiments. Of course, our own data show that stereotype threat affects the speed of the keypress; thus, our assumption is limited to the recognition process that precedes the decision to press the key.

On the basis of pretesting, we used three display values for the arrow: 30 ms, 40 ms, and 50 ms, after which the arrow was masked by a white box. Nine participants were each presented three blocks of 48 trials, one block at each of the display values. The order of presentation of the blocks was counterbalanced using a Latin square. The accuracy data were analyzed in a one-way ANOVA, with display time as a within-subjects factor. Participants' accuracy did not drop significantly as the display time was reduced from 50 ms (M = 97.90%, SD = 1.00%) to 40 ms (M = 94.40%, SD = 5.50%; $p > .20$), but there was a highly significant drop from 40 ms to 30 ms (M = 37.00%, SD = 16.70%; $p < .001$) (Tukey's Honestly Significant Difference Test; Kirk, 1995), $F(1, 16) = 96.50$, $p < .0001$, $d = 5.06$. In fact, performance at 37% accuracy does not differ significantly from chance (33%; $p > .20$).

Thus, participants easily identify the orientation of the arrow when the target is displayed for 40 ms, but they are unable to do better than chance when the target is presented only for 30 ms. This finding suggests that it takes participants somewhere between 30 ms and 40 ms, on average, to be able to process the target orientation reliably in our paradigm. As a conservative estimate of processing time, we added a constant of 30 ms to the arrival time for each trial for each participant. Using this value most likely slightly underestimates the processing time, but we can have a high degree of certainty that it takes at least this long for participants to be able to respond to the target used in the present research.

To estimate the percentage of trials participants would have missed had the target been displayed for only 150 ms, as in Experiment 1, we added the 400-ms cue display time to the 150-ms display time for the target. This 550 ms represents the amount of time from the beginning of the trial (when the fixation cross is extinguished) to the point at which the target is masked. According to the present analysis, if, on a given trial, the participant's arrival time plus processing time (30 ms) totaled more than 550 ms, then the participant should miss the signal. For each participant in Experiment 3, the percentage of total trials on which they exceeded this cutoff (> 550 ms) was computed, and these data were analyzed in a 2 (condition: stereotype threat vs. no stereotype threat) × 2 (task order: antisaccade first vs. prosaccade first) × 2 (type of saccade: corrective vs. correct) ANOVA. This analysis revealed a Saccade Type × Condition interaction, $F(1, 32) = 9.91$, $p = .004$, $d = 1.12$. A series of contrasts (Kirk, 1995) was used to decompose this interaction. This analysis showed that participants subject to stereotype threat exceeded the cutoff on 12.91% of trials on which they made corrective saccades (SD = 10.12%), significantly more than the percentage observed in any other condition (stereotype threat/correct saccade: M = 4.86%, SD = 6.00%), $F(1, 32) = 20.37$, $p < .001$, $d = 1.58$; (no stereotype threat/correct saccade: M = 5.63%, SD = 4.80%), $F(1, 32) = 16.66$, $p < .001$, $d = 1.46$; (no stereotype threat/corrective saccade: M = 5.72%, SD = 5.63%), $F(1, 32) = 15.40$, $p < .001$, $d = 1.39$, which did not differ among themselves (ps > .20). The main effect for saccade type, $F(1, 32) = 10.48$, $p = .003$, $d = 1.15$, must be interpreted in the context of this interaction.

These findings suggest that if the target in Experiment 3 had been displayed for 150 ms instead of 250 ms, then participants in the stereotype threat condition would have made 7.28% more errors than participants in the control conditions, and these errors would have occurred on the trials on which they made reflexive saccades followed by corrective saccades. Of course, this analysis requires various assumptions (e.g., estimate of processing time), and as a result, it should be treated with caution. Nonetheless, it is interesting to note that this analysis suggests that there would be an accuracy difference of 7% between participants in the stereotype threat condition and controls at 150 ms as a result of the late arrival of the stereotype threat participants at the target site, and this is the difference between these conditions that we obtained in Experiment 1, in which we used a 150-ms display time.

Taken together, the findings of Experiments 1–3 are consistent with the mere effort account and are inconsistent with an account that relies on working memory interference alone. One could argue for a dual-process model in which stereotype threat not only interferes with participants' working memory capacity, making it more difficult for them to inhibit the tendency to look at the cue, but also increases the participants' motivation to correct this tendency, and so, when given enough time (e.g., 250 ms), they are able to respond more quickly than controls. However, the findings of Experiment 3 suggest that this possibility is not likely.

Working memory interference ("foot off the brake"), mere effort ("foot on the gas"), or both could be responsible for the finding that participants subject to stereotype threat produce more reflexive saccades than control participants. However, in the next phase of trial performance, working memory interference would produce slower saccade launch latencies for stereotype threat participants than for control participants (Kane et al., 2001; Roberts et al., 1994; Stuyven et al., 2000; Unsworth et al., 2004), not the faster saccade launch latencies that were found in Experiment 3.

Thus, mere effort alone is sufficient to account for the pattern of findings in these experiments (more reflexive saccades [potentiated prepotent response], faster launch times for correct and corrective saccades, and faster adjusted reaction times). However, it should be noted that working memory is as essential to the mere effort account as it is to the working memory interference account. It is just that the mere effort account requires an intact, not a compromised, central executive. According to the mere effort account, threat participants make more reflexive saccades because the prepotent response has been potentiated ("foot on the gas"), not because they lack the working memory capacity necessary to inhibit this response ("foot off the brake"). In fact, if the working memory system were not intact, then the threat participants would not be able to launch correct and corrective saccades toward the target faster than control participants.

This analysis suggests that if participants were faced with reduced working memory capacity in addition to stereotype threat, then they would launch even more reflexive saccades ("foot off the brake" plus "foot on the gas"), their saccade launch latencies would be increased, and their overall performance advantage over control participants would be diminished, eliminated, or reversed. This hypothesis was tested in Experiment 4 by adding a concurrent task to the threat paradigm.

Experiment 4: Antisaccade and the *n*-Back Task

As previously noted, research has shown that requiring participants to perform a high-load concurrent task degrades performance on the antisaccade task because working memory capacity that is needed for the inhibition of reflexive responding and the generation of correct and corrective saccades is used instead to perform the concurrent task (Kane et al., 2001; Roberts et al., 1994; Stuyven et al., 2000; Unsworth et al., 2004). In the present experiment, we crossed a manipulation of cognitive load with the stereotype threat manipulation and measured accuracy and terminal reaction times, not eye movements or adjusted reaction times.

We predicted that under high cognitive load, stereotype threat participants should no longer have the working memory capacity necessary to produce their performance advantage. The potentiated tendency to look at the cue ("foot on the gas") produced by stereotype threat in combination with the reduced ability to inhibit looking at the cue ("foot off the brake") produced by the high-load concurrent task should increase the likelihood that participants in the stereotype threat condition will produce reflexive saccades. In addition, reduced working memory capacity should negatively impact the advantage in saccade launch latency enjoyed by threat participants. Under low-cognitive load, however, the pattern of performance should replicate the pattern observed in Experiments 2 and 3: Participants in the stereotype threat condition should have faster reaction times than participants in the control group.

To test these predictions, participants in Experiment 4 completed the antisaccade task, while also performing the *n*-back task (e.g., Jonides et al., 1997). On the *n*-back task, participants are presented with a series of letters, and on each trial, they must decide whether the present letter matches the letter that preceded it by *n* places in the series. For example, a 1-back task requires participants to judge (either true or false) whether the letter presented to them is the same as the one that appeared in the previous trial.

We used the 2-back version of this task in the present experiment. Participants were presented a letter displayed in the center of the screen (location identical to the fixation cross in the previous experiments), followed by a saccade trial. After the saccade trial, another letter was then presented, followed by another saccade trial. This pattern of letter-then-saccade trial was repeated throughout the block of trials. The participant's task was to indicate whether the current letter matched the letter that appeared two places back in the sequence and to perform well on the saccade trials. As a control condition, we used the 0-back task, which required participants to evaluate whether the letter that was presented between the antisaccade trials was one of two predetermined target letters. Thus, we held constant the requirement that participants remember two letters, but in the control condition, the two letters remained the same throughout the task.

Method

Participants

Seventy-two Northeastern University female undergraduates participated in partial fulfillment of a course requirement. All participants reported normal or corrected-to-normal vision.

Tasks and Apparatus

Participants completed two eye movement tasks, the antisaccade and prosaccade tasks. Each participant performed both saccade tasks in one of two *n*-back conditions, the 0-back or 2-back. Each participant was seated in front of a computer screen in a cubicle. A Macintosh G5 computer controlled the stimulus presentation and recorded the keypress timing and accuracy. All tasks were presented on a 17-in. (43.18-cm) monitor. Participants' heads were stabilized by a chin rest throughout the course of each task.

Eye movement tasks. The antisaccade and prosaccade tasks administered to participants were identical to those described in Experiment 2, with the exception that participants completed 72 rather than 90 saccade trials per set.

n-back tasks. Participants were instructed to fixate the fixation cross at the beginning of each trial. After the cross was extinguished, participants were presented with a letter that appeared in the center of the computer screen for 750 ms. After presentation of the letter, a saccade trial began with the appearance of a fixation cross 1,250 ms after the letter regardless of whether a response was made. After the participants responded to the target (the arrow), and an intertrial interval of 1,750 ms, the fixation cross reappeared on the screen. This fixation cross was again followed by the presentation of a letter. This pattern of letter presentation repeated itself throughout the duration of the saccade tasks. On the 2-back task, participants were instructed to monitor the letters and indicate whether each letter matched the item two back in the series by pressing keys on the keyboard designated *yes* and *no*. The 0-back task required participants to monitor the series of letters and indicate (yes or no) whether the letter matched one of two target letters that were specified by the experimenter before the task began.

Each 2-back series consisted of a sequence of 72 letters, 23 of which should have produced a yes response (31.9%) and 49 a no response (68.1%), with the first two trials always producing no answers because no letter appears two places back. Consistent with previous research (e.g., Jonides et al., 1997), all vowels and the consonants *L*, *W*, and *Y* were omitted from the stimulus sets. This yielded a total of 18 stimulus letters, with each letter appearing four times per 2-back stimulus set. We used two 2-back stimulus sets and counterbalanced the sets such that half the participants received Set A first and the other half received Set B first. To ensure that participants were not responding on the basis of the familiarity of the letters, the 2-back sequence included some 1-back matches (the present letter matched the previous letter) and 3-back matches (the present letter matched the letter that was three letters back).

There were also two 0-back stimulus sets, each of which used two different target letters. Because the 0-back stimulus set required that the target letters be presented 23 times, the 16 other letters were dispersed throughout the remaining 49 trials such that 15 of those letters appeared three times and one appeared four times. Set had no effect on the results in either the 2-back or 0-back conditions and is not discussed further.

In the 2-back condition, participants could not begin responding until the third item had been presented, because prior to this letter, no letter had appeared two places before. Thus, 70 trials were used in the analyses (both *n*-back and saccade performance) in which 2-back performance was examined, whereas analyses using the 0-back task used all 72 trials because participants were able to make a response on every trial.

Procedure

Participants completed three sets of six practice trials to familiarize themselves with the *n*-back and the saccade tasks before beginning the first block of trials. The first set of practice trials instructed participants to focus on completing the saccade part of the trial, the second set of practice trials instructed participants to focus on completing the *n*-back part of the trial, and the third practice set required participants to complete both parts of the trial. All *n*-back practice trial sets consisted of two hits and four misses. This ratio (33.3%) approximates the percentage of hits and misses used in both blocks of experimental trials (31.9%). After participants completed the practice trials, the stereotype threat manipulation was implemented. All manipulations and questionnaires were identical to those described in Experiment 2, with the exception that an additional question asked participants to assess the difficulty of the *n*-back task.

Results

Manipulation Check for Stereotype Threat

The stereotype threat manipulation checks and other questionnaire items were analyzed in 2 (stereotype threat condition: stereotype threat vs. no stereotype threat) × 2 (saccade task: antisaccade vs. prosaccade) × 2 (*n*-back task: 0-back vs. 2-back) ANOVAs, with condition, task, and *n*-back task as between-subject factors. Participants subject to stereotype threat indicated that gender differences existed to a greater extent ($M = 6.49$, $SD = 2.31$) than did controls ($M = 2.97$, $SD = 2.39$), $F(1, 66) = 88.90$, $p < .001$, $d = 2.34$. Stereotype threat participants also indicated that men performed better than women to a greater extent ($M = 4.31$, $SD = 1.86$) than did controls ($M = 6.14$, $SD = 1.08$), $F(1, 66) = 59.27$, $p < .001$, $d = 1.91$. Thus, once again, the stereotype threat manipulation was successful.

n-Back Task

The percentage of correct responses on the *n*-back task was analyzed in a 2 (stereotype threat condition: stereotype threat vs. no stereotype threat) × 2 (saccade task: antisaccade vs. prosaccade) × 2 (task order: antisaccade first vs. prosaccade first) × 2 (*n*-back task: 0-back vs. 2-back) ANOVA, with condition, task order, and *n*-back task as between-subjects factors and saccade task as a within-subjects factor. Participants identified the target letter with higher accuracy in the 0-back condition ($M = 97.07\%$, $SD = 2.60\%$) than in the 2-back condition ($M = 87.78\%$, $SD = 10.20\%$), $F(1, 64) = 64.25$, $p < .001$, $d = 2.02$. This effect was expected because the 2-back condition requires working memory resources to keep updating the last two letters, whereas the 0-back condition only requires participants to match the current letter with one of two preassigned target letters. Participants did not differ in their ability to identify the target letter as a function of the stereotype threat manipulation. No other effects were reliable.

Saccade Tasks

Accuracy and reaction time measured from target onset were analyzed in 2 (stereotype threat condition: stereotype threat vs. no stereotype threat) × 2 (saccade task: antisaccade vs. prosaccade) × 2 (task order: antisaccade first vs. prosaccade first) × 2 (*n*-back task: 0-back vs. 2-back) ANOVAs, with condition, task order, and *n*-back task as between-subjects factors and task as a within-subjects factor.

Accuracy. Participants responded to target orientation with higher accuracy on prosaccade trials (M = 98.15%, SD = 2.40%) than on antisaccade trials (M = 95.45%, SD = 4.10%), $F(1, 64) = 28.66$, $p < .001$, $d = 1.35$. This effect is consistent with findings from Experiments 1 through 3. In the low-load 0-back condition, participants were marginally more accurate in identifying target orientation (M = 97.28%, SD = 3.10%) than participants in the high-load 2-back condition (M = 96.33%, SD = 2.00%), $F(1, 64) = 2.82$, $p = .09$, $d = 0.43$. No other effects were reliable. As in Experiments 2 and 3, stereotype threat did not affect the participants' ability to correctly identify target orientation.

Reaction time. Consistent with our previous experiments, participants identified target orientation significantly more quickly on prosaccade trials (M = 464.16 ms, SD = 58.64 ms) than on antisaccade trials (M = 554.40 ms, SD = 75.98 ms), $F(1, 64) = 201.45$, $p < .001$, $d = 3.53$.

This analysis also revealed a significant *N*-Back Task × Condition × Saccade Task interaction, $F(1, 64) = 10.62$, $p < .01$, $d = 0.82$. A series of contrasts was used to decompose this three-way interaction and to test our predictions (Kirk, 1995). The 0-back condition represents a minimal memory load and should not have taxed the participants' working memory. Consistent with this notion, replicating Experiments 2 and 3, on the antisaccade task, participants in the stereotype threat condition identified target orientation more quickly (M = 514.53 ms, SD = 53.18 ms) than control participants (M = 570.59 ms, SD = 70.59 ms), $F(1, 64) = 13.67$, $p < .01$, $d = 0.93$. However, in the 2-back condition, the performance advantage enjoyed by participants in the stereotype threat condition was not only eliminated but also reversed. Participants in the control condition responded more quickly on antisaccade trials (M = 534.67 ms, SD = 85.67 ms) than participants under stereotype threat (M = 597.83 ms, SD = 67.76 ms), $F(1, 64) = 12.15$, $p < .01$, $d = 0.87$. This finding is consistent with the prediction that introducing a taxing concurrent task will undermine the performance of stereotype threat participants.

Replicating Experiments 2 and 3, on the prosaccade task, stereotype threat participants in the 0-back condition responded more quickly (M = 450.51 ms, SD = 42.74 ms) than control participants (M = 474.30 ms, SD = 65.40 ms), although this contrast was only marginally reliable, $F(1, 64) = 3.26$, $p = .09$, $d = 0.45$. In the 2-back condition, there was no difference between these conditions (stereotype threat: M = 475.47 ms, SD = 67.09 ms; no stereotype threat: M = 459.64 ms, SD = 57.42 ms; $p = .19$).

Ancillary Measures

As in Experiments 1 through 3, participants rated the antisaccade task as being more difficult than the prosaccade task, $F(1, 66) = 16.11$, $p < .001$, $d = 0.99$. As expected, the 2-back condition yielded higher task difficulty ratings (M = 7.06, SD = 2.28) than did the 0-back condition (M = 4.24, SD = 2.57), $F(1, 66) = 59.38$, $p < .001$, $d = 1.91$. Participants in the 2-back condition reported being more concerned with the experimenter's evaluation (M = 3.98, SD = 2.70) than did participants in the 0-back condition (M = 2.93, SD = 2.09), $F(1, 66) = 6.31$, $p = .013$, $d = 0.63$. Also, participants in the 2-back condition reported lower levels of effort (M = 7.90, SD = 2.15) than did participants in the 0-back condition (M = 8.62, SD = 2.21), $F(1, 66) = 4.39$, $p = .038$, $d = 0.52$. Finally, participants in the 2-back condition did not expect to perform as well (M = 5.66, SD = 2.33) as participants in the 0-back condition (M = 7.46, SD = 1.80), $F(1, 66) = 33.07$, $p < .001$, $d = 1.42$. No other effects were reliable ($ps > .20$).

Discussion

Replicating the findings of Experiments 2 and 3, in the 0-back condition, stereotype threat participants responded more quickly than control participants, as would be expected when working memory resources are available. However, when working memory resources were consumed by the 2-back task, threat participants performed more poorly than control participants. Thus, adding a high-load concurrent task that taxed working memory debilitated the performance of participants subject to threat on the antisaccade task, whereas stereotype threat alone facilitated performance when there was no (see Experiments 2 and 3) or a minimal cognitive load (0-back, Experiment 4).

These findings suggest that the potentiation of the prepotent response puts the threat participants close to their performance limit at the 250-ms display time even with their working memory intact (i.e., no or minimal cognitive load). In fact, in follow-up research, we have found that control participants still outperform threat participants on the antisaccade task at a display time of 200 ms ($p < .01$), suggesting that threat participants need more than 200 ms of display time to reach parity with the control participants, and then more time than that to surpass them, perhaps as much as the full 250 ms.

It is also interesting to note that the high-load concurrent task manipulation did not impair the performance of control participants. In fact, participants in the control group responded more quickly in the 2-back (M = 534.67 ms, SD = 85.67 ms) than in the 0-back condition (M = 570.59 ms, SD = 70.59 ms), $F(1, 64) = 7.45$, $p < .01$, $d = 0.68$, and were equally accurate (M 2-back = 94.90%, SD = 4.60%; M 0-back = 96.00%, SD = 3.70%; $p > .20$). Of course, in Experiment 4, the display interval was 250 ms, whereas in the previous research that has demonstrated the negative effect of high-load concurrent task on antisaccade performance (e.g., Roberts et al., 1994), a briefer display interval was used (150 ms). It appears that participants in the control group were not sufficiently taxed at 250 ms for their performance to even be affected by the addition of the high-load concurrent task. To the contrary, the concurrent task may have led 2-back participants to become more engaged in the task, resulting in faster reaction times on the antisaccade task than for the 0-back participants. Consistent with this interpretation when we reran the control condition with a display time of 150 ms, we replicated Robert et al.'s (1994) findings of debilitated performance with a high-load concurrent task. In any event, these findings do not impact our interpretation of the effect of manipulating cognitive load on the performance of participants subject to stereotype threat.

General Discussion

Although stereotype threat has been studied extensively, researchers have yet to reach a consensus as to what mechanism(s) mediate(s) its effect on performance. Recent research has suggested that the experience of threat reduces working memory capacity and that it is this reduction in processing capacity that leads to poor task performance (e.g., Schmader & Johns, 2003). More specifically, Schmader and Johns (2003) argued that it is the executive attention component (central executive) of working memory (Engle, 2001, 2002) that is impaired by the threat manipulation.

The central executive is also essential for effective performance on inhibition tasks like the antisaccade task. First, its resources are required to inhibit the prepotent tendency to look at the cue. Then, when inhibition is successful, working memory resources are needed to generate correct saccades. When the effort to inhibit is unsuccessful, these resources are required to launch a corrective saccade (e.g., Kane et al., 2001; Roberts et al., 1994; Stuyven et al., 2000; Unsworth et al., 2004). Thus, if, as Schmader and Johns (2003) argued, stereotype threat interferes with working memory, then participants under threat should produce more reflexive saccades than control participants because they have less ability to inhibit their tendency to look at the cue. They also should launch correct and corrective saccades more slowly than control participants because the capacity to launch these eye movements also requires the central executive.

Consistent with the notion that stereotype threat impairs working memory capacity, Experiment 1 demonstrated that when the display time was 150 ms, participants subject to threat performed more poorly than controls. However, when display time was increased to 250 ms in Experiment 2, stereotype threat participants performed better than controls. This finding is inconsistent with an explanation that relies on working memory interference alone, which suggests that participants subject to stereotype threat should perform more poorly than control participants regardless of display time.

However, the findings of Experiments 1 and 2 are consistent with the account provided by mere effort (Harkins, 2006). The mere effort account argues that stereotype threat motivates participants to strive to perform well, which potentiates prepotent responding. Thus, the mere effort account also predicts that participants subject to stereotype threat should produce more prepotent responses (reflexive saccades) than control participants, but because the prepotent response is potentiated, not because they have less ability to inhibit the response. However, the stereotype threat participants are motivated to do well, and so, if given the opportunity, they will correct these responses. Because the 150-ms display time provided in Experiment 1 does not provide sufficient time for correction, threat participants perform more poorly than controls, but at 250 ms (see Experiment 2), there is sufficient time for correction, and stereotype threat participants outperform control participants.

In Experiment 3, we examined the participants' eye movements during task performance to identify the processes that resulted in the terminal performances outcomes of Experiments 1 and 2. Consistent with the mere effort account, participants subject to stereotype threat launched more reflexive saccades, but they also generated corrective and correct saccades more quickly than participants in the control condition. Also consistent with the mere effort account was the finding that the adjusted reaction times for participants subject to threat were faster than those for control participants.

The results of Experiments 1–3 are consistent with the operation of the single process suggested by the mere effort account. In Experiment 4, we asked participants to complete the antisaccade task while also performing a 2-back concurrent task. When stereotype threat participants completed the antisaccade task under low-working memory load (0-back), they performed better than control participants, just as they did in Experiments 2 and 3, but under high-working memory load (2-back), they performed more poorly. These findings show that working memory resources are required for threat participants to outperform the control participants in Experiments 2 and 3.

In summary, in this series of experiments, we used a manipulation of stereotype threat that has been used successfully in previous research (e.g., Brown & Pinel, 2003; Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Spencer et al., 1999). The manipulation checks suggest that the manipulation was also successful in the present research. We used a task, the antisaccade task, for which the working memory interference and mere effort accounts make specific predictions. On the basis of previous research on the effects of reduced working memory capacity on performance on the antisaccade task (Kane et al., 2001; Roberts et al., 1994; Stuyven et al., 2000; Unsworth et al., 2004), the working memory interference account predicted that threat should degrade performance on this task regardless of target display time. This prediction was not supported, whereas the findings are consistent with the mere effort account.

Previously, we noted that Harkins's (2006) mere effort account of the evaluation–performance relationship and Cottrell's (1968, 1972) evaluation apprehension account of social facilitation each predict the potentiation of prepotent or dominant responses. This focus on the energization of dominant responses is also central to the arousal/drive explanations proposed by O'Brien and Crandall (2003) and Ben-Zeev et al. (2005) to account for the effect of stereotype threat on performance. The mere effort account differs in a crucial way from these explanations in that they argue only that the prepotent or dominant response is potentiated. For example, Cottrell (1968) wrote: "Only when positive or negative anticipations are produced by the presence of others will it nonselectively energize individual performance" (p. 107). The mere effort account goes beyond "nonselective" energization by arguing that the motivation to perform well also leads to an effort to correct the incorrect response if the participant recognizes that his or her response is incorrect, knows the correct response, and has the opportunity to make it. As shown in detail by the eye movement and adjusted reaction time data of Experiment 3, the findings support the mere effort account over explanations that rely on response potentiation alone.

Research conducted subsequent to Schmader and Johns's (2003) work that has invoked the concept of working memory interference is also consistent with the mere effort account. For example, Croizet et al. (2004) found that participants subject to stereotype threat scored lower on the Raven Advanced Progressive Matrices Test (Raven, 1962) than control participants. They also collected data on heart rate variability indices to assess mental workload and found evidence consistent with the notion that stereotype threat debilitated intellectual performance by producing a disruptive mental load. However, as Croizet et al. (2004) noted, their heart rate findings are consistent with either a disruptive mental load or stronger involvement in the focus task. Thus, Croizet et al.'s findings are certainly compatible with the mere effort account, which would argue that the heart rate findings reflect greater motivation on the part of participants under threat to perform well on a task on which the prepotent response is incorrect.

More recently, Beilock, Jellison, Rydell, McConnell, and Carr (2006) have argued that "stereotype threat may simultaneously affect working memory availability and direct attention in ways that are counterproductive. However, these two effects may be differentially relevant to performance depending on the attentional demands of the task being performed" (p. 1062). Performance on tasks that rely heavily on proceduralized routines but not on working memory (e.g., well-learned sensorimotor skills) may suffer from stereotype threat, "because of the stereotype threat-induced shift of attention to step-by-step control and not because the overall capacity of working memory has been reduced" (Beilock et al., 2006, p. 1062).

In their first experiment, Beilock et al. (2006) found that expert golfers subject to threat performed more poorly than controls and argued that this finding resulted from the fact that they directed attention to processes that normally operated automatically. In a second experiment, Beilock et al. (2006) manipulated whether or not the highly skilled golfers were given a secondary task and found that threat participants performed significantly better in the dual-task format than in the single-task condition. Beilock et al. argued that having to perform the concurrent task took up working memory resources that were being used for performance monitoring, thereby allowing the well-learned putting behavior to be enacted without interference.

We would argue that these findings are compatible with the mere effort account. Experts' golf strokes become proceduralized (prepotent) through "countless hours of practice that they [experts] put into honing their skill" (Beilock et al., 2006, p. 1062). Mere effort predicts that stereotype threat should motivate expert golfers to do well, which potentiates the prepotent, proceduralized stroke. However, in their effort to perform well, the expert golfers attend to the components of their putting stroke, which disrupts the performance of this well-learned behavior. Performing a concurrent task eliminates the potential for this explicit attention. Under these conditions, potentiation should now facilitate performance because the ability to explicitly monitor task components is eliminated by the concurrent task. Consistent with this notion, when performing the dual task, the experts performed better when they were under stereotype threat than when they were not (Beilock et al., 2006, Experiments 2 and 3).

Although the present findings are consistent with the mere effort account, there are a number of factors that limit the range of conditions under which we would expect these effects to be produced. For example, the mere effort account does not make predictions for self-reported effort, only for task performance. It is not at all unusual to find a lack of correspondence between measures of self-reported effort and actual measures of performance. In a meta-analysis of social loafing research, Karau and Williams (1993) found that the social loafing effect was robust, but the average effect size for self-reported effort in these experiments was not significantly different from zero. Karau and Williams (1993) reported that "only slightly more than half of the reported differences were in the direction of reduced collective effort" (p. 699). Thus, it is not surprising that a number of experiments in the stereotype threat domain have found that threat affected performance but not self-reported effort (e.g., Brown & Pinel, 2003; Keller, 2002; Steele & Aronson, 1995). In any event, the focus of the mere effort account is on task performance not on self-reported effort.

In addition, the mere effort account argues that women subject to stereotype threat are motivated to perform well when they find themselves in a given performance setting. We do not argue that these women seek out opportunities to demonstrate their proficiency. In fact, there is reason to believe that given the chance, they will avoid having to show how well they can perform. For example, a large number of experiments show that participants are more motivated to perform well when they are subject to experimenter evaluation than when they are not subject to this evaluation (e.g., Harkins, 2000; Harkins & Lowe, 2000; Harkins & Szymanski, 1988; Harkins, White, & Utman, 2000; Szymanski & Harkins, 1987; White, Kjelgaard & Harkins, 1995). Nonetheless, Szymanski and Harkins (1993) have also shown that when given the choice, participants avoid subjecting themselves to the scrutiny of this source of evaluation.

Thus, our research shows how threat participants perform when they find themselves in a situation that they did not seek out and that does not provide a plausible explanation as to why they may not perform well. Under these circumstances, they are motivated to perform well. As a result, the mere effort account is not inconsistent with research that suggests that instead of seeking out the opportunity to demonstrate how well they can perform, participants under stereotype threat are quite willing to take advantage of explanations that will allow them to deflect responsibility for their performance (e.g., Ben-Zeev et al., 2005; Steele & Aronson, 1995; Stone, 2002).

The next step in the present line of research will be to test the mere effort account on the GRE quantitative problems often used in the stereotype threat literature. Key to this effort will be the identification of the prepotent response on these problems. For example, one might argue that if stereotype threat motivates women, then they should spend more time working on GRE quantitative problems and/or attempt more problems than women not subject to threat. However, there is no evidence to suggest that spending more time per problem or attempting more problems represents the prepotent response. In fact, no consistent effects have been reported on these measures. For example, although Steele and Aronson (1995, Experiment 2) found that threat participants spent more time per problem than controls, Spencer et al. (1999, Experiment 1) found no differences on this measure. Steele and Aronson (1995, Experiment 4) also found that threat participants completed fewer problems than control participants, whereas Keller (2002) found no difference on this measure.

Previous research does show that when faced with quantitative problems, the participants' prepotent response is to attempt to solve the equations presented to them, rather than simplifying terms or using logic, estimation, or intuition to find the correct answer. So, for example, Gallagher and De Lisi (1994) found that 63% of their participants used the solving approach as opposed to one of the other methods; Gallagher et al. (2000, Experiment 2) found that 56% did so; whereas Quinn and Spencer (2001; Experiment 2) found that 60% relied on the solving approach. We must now determine what impact potentiation of this prepotent response has on the performance of GRE problems like the ones used by Schmader and Johns (2003).

The mere effort account differs from other accounts of the effects of stereotype threat on performance in that it suggests that participants are not falling victim to a process that negatively affects their cognitive capacities (e.g., processing interference) or leads them to withdraw effort. Instead, it suggests that participants subject to stereotype threat may actually be trying very hard during task performance to disprove the negative stereotypes directed at their group. It is just that these efforts may be misdirected. Although the present series of experiments suggests that motivation is an important factor in explaining the effects of stereotype threat on task performance, additional research is required to determine whether motivation also contributes to the effect of stereotype threat on tasks traditionally used in the stereotype threat literature (e.g., quantitative GRE problems). If this research supports the mere effort model, then the task becomes determining how to assist stigmatized individuals in channeling their efforts more effectively during task performance.

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.

Bach, M., Bouis, D., & Fischer, B. (1983). An accurate and linear infrared oculometer. *Journal of Neuroscience Methods, 9,* 9–14.

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? *Personality and Social Psychology Bulletin, 32,* 1059–1071.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41,* 174–181.

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

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.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.

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

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.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. *European Journal of Social Psychology, 33,* 267–285.

Cottrell, N. B. (1968). Performance in the presence of other human beings: Mere presence, audience, and affiliation effects. In E. C. Simmell, R. A. Hoppe, & G. A. Milton (Eds.), *Social facilitation and imitation behavior* (pp. 91–110). Boston: Allyn & Bacon.

Cottrell, N. B. (1972). Social facilitation. In N. B. McClintock (Ed.), *Experimental social psychology* (pp. 185–236). New York: Holt, Rinehart, & Winston.

Crevits, L., & Vandierendonck, A. (2005). Gap effect in reflexive and intentional prosaccades. *Neuropsychobiology, 51,* 39–44.

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.

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.

Engle, R. W. (2001). What is working memory capacity? In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Suprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 297–314). Washington, DC: American Psychological Association.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19–23.

Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *The Psychology of Learning and Motivation, 44,* 145–199.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128,* 309–331.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13,* 171–180.

Ford, K. A., Goltz, H. C., Brown, M. R. G., & Everling, S. (2005). Neural processes associated with antisaccade task performance investigated with event-related fMRI. *Journal of Neurophysiology, 94,* 429–440.

Gallagher, A. M., & De Lisi, R. (1994). Gender differences in Scholastic Aptitude Test–Mathematics problem solving among high-ability students. *Journal of Educational Psychology, 86,* 204–211.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology, 75,* 165–190.

Godijn, R., & Kramer, F. (2006). Prosaccades and antisaccades to onsets and color singletons: Evidence that erroneous prosaccades are not reflexive. *Experimental Brain Research, 172,* 439–448.

Harkins, S. (2000). The potency of the potential for experimenter and self-evaluation in motivating task performance. *Basic and Applied Social Psychology, 22,* 277–289.

Harkins, S. (2001). The three-variable model: From Occam's razor to the black box. In S. Harkins (Ed.), *Multiple perspectives on the effects of* *evaluation on performance: Toward an integration* (pp. 99–131). Norwell, MA: Kluwer Academic Press.

Harkins, S. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91,* 436–455.

Harkins, S., & Lowe, M. (2000). The effects of self-set goals on task performance. *Journal of Applied Social Psychology, 30,* 1–40.

Harkins, S., & Szymanski, K. (1988). Social loafing and self-evaluation with an objective standard. *Journal of Experimental Social Psychology, 24,* 354–365.

Harkins, S., White, P., & Utman, C. (2000). The role of internal and external sources of evaluation in motivating task performance. *Personality and Social Psychology Bulletin, 26,* 100–117.

Hochman, S. (1967). The effects of stress on Stroop color-word performance. *Psychonomic Science, 9,* 475–476.

Huguet, P., Galvaing, M., Monteil, J., & Dumas, F. (1999). Social presence effects in the Stroop task: Further evidence for an attentional view of social facilitation. *Journal of Personality and Social Psychology, 77,* 1011–1025.

Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle: Teaching stereotype threat as a means of improving women's math performance. *Psychological Science, 16,* 175–179.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience, 9,* 462–475.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General, 130,* 169–183.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217.

Karau, S. J., & Williams, S. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65,* 681–706.

Kawagoe, R., Takikawa, Y., & Hikosaka, O. (1998). Expectation of reward modulates cognitive signals in the basal ganglia. *Nature Neuroscience, 1,* 411–416.

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

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

Kirk, R. (1995). *Experimental design*. Pacific Grove, CA: Brooks/Cole.

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.

Kveraga, K., Boucher, L., & Hughes, H. C. (2002). Saccades operate in violation of Hick's law. *Experimental Brain Research, 146,* 307–314.

MacKinnon, D., Geiselman, E., & Woodward, J. (1985). The effects of effort on Stroop interference. *Acta Psychologica, 58,* 225–235.

Massen, C. (2004). Parallel programming of exogenous and endogenous components in the antisaccade task. *The Quarterly Journal of Experimental Psychology: Section A, 57,* 475–498.

McFall, S., Jamieson, J., & Harkins, S. (2007). *Testing the generalizability of the mere effort account of the evaluation-performance relationship.* Manuscript submitted for publication.

Mitchell, J. P., Macrae, C. N., & Gilchrist, I. D. (2002). Working memory and the suppression of reflexive saccades. *Journal of Cognitive Neuroscience, 14,* 95–103.

Montagnini, A., & Chelazzi, L. (2005). The urgency to look: Prompt saccades to the benefit of perception. *Vision Research, 45,* 3391–3401.

Nakamura, K., & Hikosaka, O. (2006). Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *Journal of Neuroscience, 26,* 5360–5369.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin, 29,* 782–789.

O'Malley, J., & Poplawsky, A. (1971). Noise-induced arousal and breadth of attention. *Perceptual and Motor Skills, 33,* 887–890.

Pallak, M., Pittman, T., Heller, J., & Munson, P. (1975). The effect of arousal on Stroop color-word task performance. *Bulletin of the Psychonomic Society, 6,* 248–250.

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

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

Raven, J. C. (1962). *Advanced progressive matrices: Sets I and II*. London: H. K. Lewis.

Roberts, R. J., Hagar, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123,* 374–393.

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

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

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: The effects on women's performance. *Journal of Experimental Social Psychology, 39,* 68–74.

Sereno, A. B. (1992). Programming saccades: The role of attention. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 89–107). New York: Springer.

Smyrnis, N., Evdokimides, I., Stefanis, N. C., Constantinidis, T. S., Avramopoulos, D., Theleritis, C., et al. (2002). The antisaccade task in a sample of 2,006 young males. *Experimental Brain Research, 147,* 53–63.

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

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

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. *Journal of Personality and Social Psychology, 77,* 1213–1227.

Stone, J., Perry, Z. W., & Darley, J. M. (1997). "White men can't jump": Evidence for the perceptual confirmation of racial stereotypes following a basketball game. *Basic and Applied Social Psychology, 19,* 291–306.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662.

Stuyven, E., Van der Goten, K., Vandierendonck, A., Claeys, K., & Crevits, L. (2000). The effect of cognitive load on saccadic eye movements. *Acta Psychologica, 104,* 69–85.

Szymanski, K., & Harkins, S. (1987). Social loafing and self-evaluation with a social standard. *Journal of Personality and Social Psychology, 53,* 891–897.

Szymanski, K., & Harkins, S. (1993). The effect of experimenter evaluation on self-evaluation within the social loafing paradigm. *Journal of Experimental Social Psychology, 29,* 268–286.

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1302–1321.

White, P., Kjelgaard, M., & Harkins, S. (1995). Testing the contribution of self-evaluation to goal-setting effects. *Journal of Personality and Social Psychology, 69,* 69–79.

Zajonc, R. (1965, July 16). Social facilitation, *Science, 149,* 269–274.

Received July 21, 2006
Revision received May 21, 2007
Accepted May 22, 2007 ■