Social Cognition, Vol. 33, No. 6, 2015, pp. 605–625

STEREOTYPE THREATENING CONTEXTS ENHANCE ENCODING OF NEGATIVE FEEDBACK TO ENGENDER UNDERPERFORMANCE AND ANXIETY

Chad E. Forbes, Kelly A. Duran, Jordan B. Leitner, and Adam Magerman
University of Delaware

In a recent meta-analysis, women exhibit varied performances on mathematic aptitude tests yet report more negative math perceptions compared to men. One explanation for this is that stereotype threatening (ST) contexts, that is, stressful contexts that engender negative emotions, arousal, and error vigilance, facilitate encoding of stereotype confirming information. Two studies tested this hypothesis. Study 1 found that ST women encoded error feedback better than correct feedback received on a math task compared to non-threatened women and men. Study 2 extended this effect to neutral information presented in conjunction with negative feedback. Furthermore, ST-based negative memory encoding mediated underperformance on a future difficult math test, which in turn engendered post-task anxiety. Findings indicate ST fosters negative math affect via biased memories even when women's initial performance is comparable to men.

Keywords: stereotype threat, memory encoding, memory recall, math anxiety, STEM identification

According to a recent meta-analytic study that included an international sample of over 493,000 students, women exhibited varied performances on tests of mathematic aptitude compared to men (i.e., either comparable or worse; Else-Quest, Hyde, & Linn, 2010). Despite this, women systematically reported more negative math attitudes and less math confidence compared to men. To date, explanations for why these attitudinal discrepancies between men and women exist remain unclear. The studies described here provide a mechanistic account for this paradox

The authors thank the many research assistants who helped collect data for these experiments including Katelyn Borden, Kate Feiner, and Alyssa Hoffman. We also thank Toni Schmader for invaluable comments on previous versions of the manuscript.

Address correspondence to Chad E. Forbes, Ph.D., Assistant Professor, Department of Psychological and Brain Sciences, University of Delaware, 111 Wolf Hall, Newark, DE 19716; E-mail: cforbes@psych.udel.edu.

by testing the hypothesis that women process negative performance-related feedback differently in stereotype threatening contexts compared to their male counterparts. We provide evidence that stereotype threatening contexts enhance encoding and recall of negative, stereotype confirming performance feedback compared to positive, stereotype disconfirming feedback to undermine performance on future math tests and exacerbate general domain anxiety accordingly. Such findings help inform a potential vicious cycle that prompts women to leave math testing situations with feelings that are orthogonal to the performance itself.

STEREOTYPE THREAT ENGENDERS AROUSAL AND A NEGATIVE AFFECTIVE STATE

Stereotype threat is a stressful, physiologically arousing response that stigmatized individuals, particularly those most identified with the stigmatized domain, experience when they fear that their actions may confirm a negative group stereotype (Schmader, Johns, & Forbes, 2008; Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995). For instance, stereotype threat has been shown to elicit increases in blood pressure (Blascovich, Spencer, Quinn, & Steele, 2001) and skin conductance (Osborne, 2006, 2007), as well as neurological indices of emotion regulation (Wraga, Helt, Jacobs, & Sullivan, 2007) and conflict detection (Forbes, Schmader, & Allen, 2008).

Coupled with this arousal, stereotype threat also evokes a negative affective state. For example, when stigmatized individuals are placed under stereotype threat, they sometimes report heightened levels of explicit anxiety (Spencer et al., 1999), self-doubt (Steele & Aronson, 1995), negative expectations (Stangor, Carr, & Kiang, 1998), feelings of dejection (Keller & Dauenheimer, 2003), and task-related worries (Beilock, Rydell, & McConnell, 2007; Cadinu, Maass, Rosabianca, & Kiesner, 2005). More implicit, nonverbal indicators of anxiety have also been observed in both performance and non-performance related domains, including dot probe measures typically used in clinical settings to detect abnormal anxiety responses to stimuli (Johns, Inzlicht, & Schmader, 2008; Bosson, Haymovitz, & Pinel, 2004).

Thus, it is clear that stereotype threat increases both physiological arousal and negative appraisals associated with that arousal, which in turn can undermine performance on high stakes tests like the SAT, GRE, and Advanced Placement exams (e.g., Forbes & Schmader, 2010; Good, Aronson, & Harder, 2008; Walton & Spencer, 2009). What isn't clear is why stigmatized individuals may leave a performance situation with more negative feelings and performance perceptions. One answer could lie in how stereotype threat-induced arousal and anxiety may bias what stigmatized individuals actually encode and recall about stereotype threatening situations.

STEREOTYPE THREAT MAY FACILITATE BETTER ENCODING OF NEGATIVE INFORMATION

As a negatively arousing experience, stereotype threat may recruit the same biological mechanisms that underlie the well-documented effect of enhanced encoding of emotionally evocative information. Indeed, extant literature indicates that negative, emotionally arousing information receives privileged attention compared to positive and neutral information, and as such, is better encoded (Hamann, 2001), consolidated (LaBar & Phelps, 1998), and retrieved (Ochsner, 2000) in negatively arousing, stressful contexts (e.g., Levine & Burgess, 1997; Payne, Jackson, Ryan, Hoscheidt, Jacobs, & Nadel, 2006; for a review see LaBar & Cabeza, 2006). The effects of emotional memory encoding are long lasting as well (Canli, Zhao, Brewer, Gabrieli, & Cahill, 2000; Hamann, Ely, Grafton, & Kilts, 1999).

Past research suggests that stereotype threatening contexts engender an attentional bias toward negative, but not positive, feedback very early in the information processing stream. For instance, Forbes and Leitner (2014) had women complete math problems in stereotype threatening or neutral environments while continuous electroencephalography (EEG) activity was recorded. Feedback was provided after each problem indicating whether participants solved the problem correctly or not. Results revealed that as early as 75 ms, stereotype threatened women exhibited better communication between neural regions integral for attention and working memory, and more activity in a region integral for visual word processing in response to wrong, that is, stereotype confirming, feedback but not correct feedback. Women in stereotype neutral contexts exhibited the opposite pattern. In fact, increased activity in the visual word processing area in response to the negative feedback predicted decreased performance among stereotype threatened women only, suggesting increased processing of the wrong feedback may have compromised or taxed cognitive resources otherwise needed for optimal performance. What remains unknown, however, is whether this attentional bias, in turn, prompts more efficacious encoding of the negative, stereotype confirming feedback and what role said encoding bias may play in performance and stigmatized individuals' affective experience.

Previous research supports the possibility that stereotype threat augments encoding of negative information. Evidence from social-evaluative threat studies, which have been likened to stereotype threatening contexts (Schmader et al., 2008), indicates that situations that engender increased motivation to perform well and a fear of receiving negative evaluation correspond with increased cortisol levels (Dickerson & Kemeny, 2004; see Matheson & Cole, 2004, for suggestive evidence). Importantly, stress-induced increases in cortisol prior to memory encoding predicts enhanced long-term memory for emotional episodes and decreased memory for neutral episodes (Payne, Jackson, Hoscheidt, Ryan, Jacobs, & Nadel, 2007).

Such findings might explain why women who are math, science, and engineering (MSE) majors exhibit greater physiological arousal and superior recall for details specific to a male dominated (compared to gender neutral) video and their context in social identity threatening situations (Murphy, Steele, & Gross, 2007). Whether these mechanisms affect the stigmatized in a performance situation and biases memory processes toward stereotype confirming evidence is a critical question that remains unanswered.

Thus, to the extent that stereotype threat is a stressful, negative emotional experience, it should initiate the same neural and hormonal stress response that is typically seen in other emotionally evocative situations. As such, stereotype threatening situations should ultimately facilitate encoding of negative, stereotype confirming information compared to positive information largely because this type of information would be congruent with the threatened individual's affective state.

OVERVIEW OF STUDIES AND HYPOTHESES

The present studies examined the hypothesis that stereotype threat facilitates encoding of negative, compared to positive, feedback. Studies 1 and 2 examined whether participants under stereotype threat, relative to their non-threatened counterparts, exhibited enhanced encoding of negative feedback compared to positive feedback (Study 1) and neutral information presented in conjunction with feedback (Study 2). To study these questions, men and women received positive and negative feedback on a math task that was described as either diagnostic of their math intelligence (Studies 1 and 2) or as a problem-solving task (Study 1). We varied the font of the feedback during the math task, and subsequently administered a surprise memory test to examine participants' ability to recognize previously seen feedback written in the different fonts. We hypothesized that women under stereotype threat would encode negative feedback better than positive feedback, thus exhibiting better recall for the fonts associated with wrong feedback, but that positive and negative feedback would be encoded equally in the absence of stereotype threat. Additionally, we expected that women under stereotype threat would encode negative feedback better than non-threatened women. Finally, we conducted more exploratory analyses that examined whether enhanced memory encoding of negative feedback ultimately undermined performance on future difficult math tests to promote general feelings of post-test anxiety among stereotype threatened women. Such findings could help explain why women may leave performance situations with more negative domain perceptions and decreased confidence that may be orthogonal to their actual performance, suggesting biased memory encoding could be just as deleterious for perceptions of math ability as underperforming on a math test itself.

STUDY 1

METHODS

Participants. One hundred and eleven participants (54 males) participated for course credit. To optimize our stereotype threat manipulation, we recruited participants who had knowledge of the negative female math stereotype. Specifically, participants responded with a 3 or lower to the following question during a pretesting session: "Regardless of what you think, what is the stereotype that people have about women and men's math ability" (1 = Men are better than women; 7 = Women are better than men). Grubbs' test (also known as the extreme studentized deviate method) on positive and negative feedback memory scores identified one woman and man in the stereotype threat condition and one man in the control condition as being significant outliers, that is, having Z scores greater than 3.5 (namely, standard deviations) from the grand mean ($p < .05$). This left a final sample of 108 participants for analyses. The goal for data collection across all studies was to obtain a sample of at least 25 participants per cell. This number was derived in light of previous stereotype threat studies conducted in the primary author's laboratory.

Procedure. Participants were seated at a computer and read instructions that were also read to them through headphones by either a male experimenter (stereotype threat condition) or a female experimenter (control condition). Participants were randomly assigned to either a diagnostic math test/stereotype threat (DMT) condition or problem-solving task/control (PST) condition. Following procedures previously shown to manipulate stereotype threat (e.g., Steele & Aronson, 1995), participants in the DMT condition were informed that they would be completing tasks that were diagnostic of their math intelligence. Participants in the PST condition were told that they would be completing an experimental problem-solving task that was diagnostic of the type of problem-solving strategies they prefer. To prime stereotype threat, participants completed demographic questions and indicated their gender in the DMT condition, and DMT sessions were always conducted with at least one male participant present. In contrast, PST sessions always contained either all male or all female participants. Once participants read through the instructions they completed the math feedback task (described below). For the purposes of providing a filler and manipulation check for stereotype threat effects, participants then completed a standardized, difficult math test for 5 minutes. Variations of this test have been used to demonstrate stereotype threat effects in the past (Schmader & Johns, 2003). Next, participants received a surprise memory test to assess the extent to which they encoded fonts associated with wrong or correct feedback (described below). Finally, participants reported levels of anxiety and math identification and were debriefed.

Math Feedback Task. In an attempt to create a math task that allowed for the presentation of many math problems in a timely manner, but were also novel and thus more cognitively demanding to participants, the math task consisted of problems that contained an operator term referred to as sharp (#; Paynter, Reder, & Kieffaber, 2009). In sharp problems, for example, 19 # 46 =, individuals are told to solve the problem by taking the sum of the tens place digits for the two operands, multiplying this number by the sum of the ones place digits, and multiplying this product by three (in the problem above this would equal (1 + 4) × (9 + 6) × 3, which equals 225). Participants were provided with three answer choices below each problem. The location of the correct answer was randomly presented in the three answer positions. There were two blocks of 25 problems each for a total of 50 problems. Each trial began with the presentation of a crosshair in the middle of the screen, followed by the problem. Participants were instructed to select their answer by clicking on the desired option on the screen. Following each response, participants were presented with feedback on a white background for 2 seconds that indicated whether their answer was correct (written in blue) or wrong (written in red). Critically, we varied the font of the feedback that appeared after each answer choice. That is, once participants selected an answer, they could see the words "Wrong" or "Correct," dependent on the accuracy of the participant's answer, written with an Arial font, Papyrus font, Graffiti font, etc., for example, Wrong or **Correct**. Two-hundred different fonts were collected for use in the study, as either feedback during the math feedback task or lures for the memory test. All fonts were equated for size. To ensure that a given font would never be uniquely associated with a given feedback type, all fonts had an equal likelihood of being presented as either wrong or correct feedback and all participants had an equal likelihood of seeing a given font type as wrong or correct feedback. Participants were given 16 seconds to solve each problem. Failure to respond within that time frame prompted wrong feedback. Participants took 10.69 seconds on average to answer problems. Participants did not have scratch paper.

Standard Filler Math Task. The filler math task contained 15 difficult math word problems taken from the GRE (Forbes & Schmader, 2010; Study 4). Participants were provided with scratch paper and were given 5 minutes to solve as many problems as they could. No feedback was provided to participants. A math test accuracy percentage was created by dividing the number of problems answered correctly by the number of problems attempted and multiplying the outcome by 100.

Memory Test. After the filler math task, participants received a surprise memory test containing 100 trials, 50 of which contained feedback fonts that participants may have seen during the first block of the math feedback task (25 possible wrong fonts and 25 possible correct fonts) and 50 of which were lures (25 wrong lures and 25 correct lures). In an effort to optimize the length of the memory test while also accounting for the fact that the memory test was dependent on participants' performance (e.g., any given individual could perform quite well or poorly and would see more correct or wrong feedback respectively), participants' memory for fonts was only tested from the first block of the math feedback task. Thus, the memory test likely consisted of a smaller amount of targets (feedback fonts participants actually saw) and a greater amount of lures because whatever wrong or correct font participants didn't see on a given trial invariably became a lure in the memory test. This actually results in a more conservative assessment of memory for target fonts as there were likely to be many more lures than targets. This isn't a concern with regard to memory assessment, however, as the d' memory measure accounts for unequal numbers of lures and targets by standardizing hits, misses, false alarms, and correct rejections.

Each trial of the memory test began with randomly presenting participants with the words "Wrong" or "Correct" written in one of the 100 different fonts in the middle of the computer screen. For example, using the examples provided above, if a person answered a problem incorrectly and saw wrong feedback written in this font, Wrong, then they saw Wrong during the memory test. Likewise, if they saw correct feedback presented as **Correct** during the feedback task then they saw **Correct** in the memory test. A scale was presented below each font/feedback combination and participants were asked to rate whether they thought they had seen the font/feedback during the math feedback task using a 6-point scale with labels on 1 (*I know I didn't see it*), 4 (*I think I saw it*), and 6 (*I know I saw it*). If participants had been exposed to the font during the math feedback task, responses of 4–6 were classified as a hit, and responses of 1–3 were classified as a miss. Regarding novel fonts, responses of 4–6 were classified as a false alarm, and responses of 1–3 were classified as correct rejections. To examine participants' ability to accurately discriminate seen from unseen fonts we calculated d', a measure argued to be a more sensitive assessment of memory effects that accounts for potential guessing (Wickens, 2002). To calculate d', z scores for false alarm rates are subtracted from z scores for hit rates. Thus larger d' values indicate a better ability to discriminate seen from unseen fonts. Because z scores for 0 or 1 cannot be calculated, participants without hits were given hit scores of 0.1 and participants with perfect hit rates had 0.1 subtracted from their hit score.

Post-Task Anxiety. Participants were asked to rate on a scale of 1 (*strongly disagree*) to 7 (*strongly agree*) the extent to which they felt agitated, anxious, nervous, uneasy, and worried. We calculated the mean of these responses, and higher numbers represented more anxiety ($\alpha = .87$).

RESULTS

Difficult Math Test Performance and Post-Task Anxiety. An initial 2 (Gender: Men or Women) × 2 (Condition: DMT or PST) factorial ANOVA was conducted on participants' performance on the difficult math test (i.e., the filler task). This analysis yielded no main effects or interactions, ps > .15. However, given the well-documented effects of stereotype threat on performance and engendering negative affective states (for a review see Schmader, Johns, & Forbes, 2008), planned contrasts were also conducted to compare stereotype threatened women's performance on the difficult math test to the other three conditions. As expected, women under stereotype threat ($M = 37.69$, $SD = 23.67$) performed worse on the math test compared to women ($M = 53.06$, $SD = 33.73$) and men ($M = 47.77$, $SD = 31.92$) in the PST condition, and men ($M = 48.76$, $SD = 29.60$) in the DMT condition, $t = -2.0$, $p < .05$, $d = .39$.

An initial ANOVA on participants' post-task anxiety scores also did not yield any main effects or interactions, ps > .12. However, a planned contrast analysis indicated that women under stereotype threat ($M = 3.23$, $SD = 1.29$) reported higher levels of anxiety compared to women ($M = 2.63$, $SD = 1.11$) and men ($M = 2.72$, $SD = 1.08$) in the PST condition, and men ($M = 2.61$, $SD = 1.07$) in the DMT condition, $t = 2.38$, $p < .02$, $d = .47$. These findings suggest the manipulation was at least partially successful in engendering both a stereotype threatening context and heightened levels of anxiety among stereotype threatened women.

Math Feedback Task. The math feedback task was designed to be only moderately difficult in an effort to avoid discrepant presentations of feedback type in either condition. As stereotype threat effects typically emerge on more difficult math tasks (Leitner, Jones, & Hehman, 2013; Spencer et al., 1999), we did not expect scores to differ as a function of stereotype threat. Analyses were constrained to the first block of the feedback task to provide insight into the number of target font types presented in the memory test. A 2 (Gender: Men, Women) × 2 (Task Description: DMT, PST) ANOVA conducted on the number of problems participants got wrong on the first block of the math feedback task revealed a main effect for condition, $F(1, 104) = 25.68$, $p < .001$, $\eta^2 = .20$, indicating that men and women in the PST condition ($M_{Men} = 12.32$, $SD_{Men} = 2.84$; $M_{Women} = 12.82$, $SD_{Women} = 4.38$) answered more sharp problems incorrectly than men and women in the DMT condition ($M_{Men} = 7.83$, $SD_{Men} = 4.31$; $M_{Women} = 9.53$, $SD_{Women} = 3.82$). No other effects were significant (*p*s > .15). Given that *d′* scores standardize hits, misses, false alarms, and correct rejections this was not expected to impact assessments of memory for feedback in any way.

Memory Performance. All descriptive statistics for variables associated with memory performance across the two studies are provided in Table 1. A 2 (Gender: Men, Women) × 2 (Task Description: DMT, PST) × 2 (Feedback Type: Wrong, Correct) mixed factors ANOVA with repeated measures on the latter variable was then conducted on participants' memory scores for fonts. There was no main effect for feedback type ($p = .42$) nor an interaction between feedback type and task description ($p = .32$). There was, however, an interaction between feedback type and gender, $F(1, 103) = 4.56$, $p < .04$, $\eta^2 = .04$, that was qualified by a three-way interaction between feedback type, condition, and gender, $F(1, 103) = 4.76$, $p < .04$, $\eta^2 = .04$ (Figure 1).

Simple effect analyses revealed that these interactions were largely the result of women in the DMT condition exhibiting enhanced encoding of wrong feedback fonts. Specifically, whereas women in the DMT condition showed greater encoding of wrong feedback compared to correct feedback, $F(1, 103) = 11.93$, $p < .01$, $\eta^2 = .10$, 95% CI [.21, .78], women in the PST condition exhibited inefficacious encoding of both correct and wrong feedback. Men in both conditions exhibited similar patterns to women in the PST condition, as there were no differences in the extent to which they encoded correct or wrong feedback.

Between conditions, women in the DMT condition encoded wrong feedback better than women in the PST condition, $F(1, 103) = 3.85$, $p = .05$, $\eta^2 = .04$, $\eta^2 = .04$, 95% CI [-.004, .76], and men in the DMT condition, $F(1, 103) = 12.47$, $p < .01$, $\eta^2 = .11$, 95% CI [.27, .97]. No other comparisons were significant, *p*s > .10. To determine whether performance on the math feedback task had any undue influence on memory for feedback, an identical analysis was conducted including participants' math feedback task performance (total incorrect) as a covariate. These analyses yielded nearly identical results to those described above (three-way interaction: $F[1, 103] = 4.43$, $p < .04$, $\eta^2 = .04$), indicating performance did not affect participants' memory for feedback in any way.

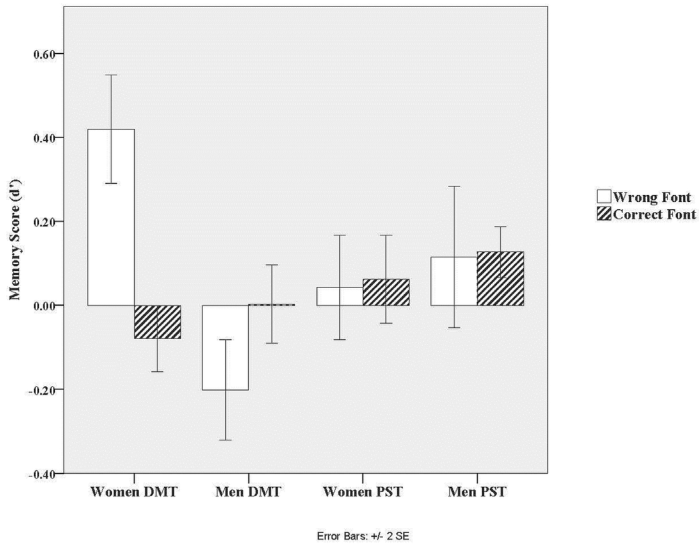

FIGURE 1. Men and women's d' memory scores as a function of feedback type and condition. (DMT = Diagnostic Math Test; PST = Problem Solving Task.)

Exploratory Analyses on Link Between Stereotype Threat, Negative Memories, Performance, and Post-Test Anxiety. To ascertain whether there was a link between stereotype threat, enhanced encoding of negative feedback on the math feedback task, future performance on difficult math tests, and general post-test anxiety, serial mediation analyses were conducted on data from Study 1. We tested for serial mediation by deriving unstandardized regression coefficients and 95% bias-corrected confidence intervals (CIs) from 10,000 bootstrap estimates (Hayes, 2013). 95% CIs are considered significant if the interval (e.g., [.5, .9]) does not contain zero (Cumming, 2008). In separate models for threat and control participants, we entered variables in a manner that most closely represented the temporal nature in which the variables were collected. Thus, gender was utilized as a predictor, d' for wrong feedback served as an initial mediator, performance was a secondary mediator, and anxiety was the outcome variable.

These analyses revealed that under threat, the indirect pathway was not significant, b = .04, 95% CI [-.0122, .2274], indicating that the relationship between gender and anxiety was not fully mediated by memory for wrong feedback and test performance. However, patterns suggested a potential connection between memory encoding and post-task feelings of anxiety among stereotype threatened women. Under threat, women compared to men showed greater memory for wrong feedback seen during the math feedback task, b = .62, 95% CI [.2566, .9753], which in turn predicted worse performance on the subsequent difficult math test, b = -9.95, 95% CI [-19.1171, -.7775]. Only gender, however, was a moderate predictor of anxiety, b = .62, 95% CI [-.0538, 1.2913]. Neither the indirect effect nor any of the direct paths were significant for participants in the control condition (indirect b = .005, 95% CI [-.0042, .0862]). There were also no relationships among these variables and memory for correct feedback (all CIs contained zero).

DISCUSSION

Findings from this study provide initial evidence that stereotype threat engenders enhanced encoding of negative feedback. When women were placed under stereotype threat they exhibited enhanced encoding of negative feedback compared to positive feedback, their male counterparts, and women not under stereotype threat. This effect was not evident in any of the other conditions, suggesting enhanced encoding of negative performance-based feedback is unique to stereotype threatening contexts. Given that women under stereotype threat tended to report higher levels of anxiety (although this pattern only emerged in the planned contrast), these findings are also consistent with the interpretation that enhanced encoding of negative feedback may have been associated with emotional/mood congruent memory encoding processes. Furthermore, mediational analyses provide at least some initial evidence that post-task feelings of anxiety may be influenced by the effects that biased memory encoding processes have on undermining performance on math tasks orthogonal to the task where the initial memories were formed.

An important additional question revolves around what stereotyped threatened individuals encode when negative feedback is presented amongst other information. Past research indicates that the experience of arousal directs attention toward the central arousing stimulus in the environment. Specifically, arousal increases encoding of information that is most relevant to the current motivational state engendered by the emotion (Levine & Pizarro, 2004; Oatley & Johnson-Laird, 1987). With respect to Study 1 findings, this suggests that wrong feedback may be encoded more efficaciously to the extent this negative stereotype confirming evidence is perceived as the centrally threatening stimulus in stereotype threatening contexts.

With that said, other research suggests that social identity threatening contexts, that is, broader identity threatening situations not tied to performance in a stigmatized domain per se (but which stereotype threatening contexts are an example of), could also facilitate attention to irrelevant information (Inzlicht & Kang, 2010) and perhaps even encoding of neutral information. For instance, Murphy and colleagues (2007) had male and female math, science, and engineering (MSE) majors watch a supposed MSE conference video that was either gender balanced or unbalanced (more men) while electrodermal and cardiovascular activity was recorded. A surprise memory test was then given to assess memory for both details specific to the video and details specific to the experimental room. Results revealed that compared to both men in either condition and women who watched the gender balanced video, women MSE majors who watched the gender unbalanced video demonstrated greater physiological arousal while watching the video and superior recall for both details specific to the video and experimental room. However, these findings do little to inform our understanding of how valenced, stereotype-confirming or disconfirming information specific to performance situations in stigmatized domains is processed, nor did the authors include a rigorous test of memory encoding (e.g., provide a memory test with adequate lures to account for participants' potential for guessing answers). Nevertheless, these findings suggest that any information presented in stereotype threatening contexts in conjunction with a centrally threatening stimulus could be encoded more efficaciously.

The goal of Study 2 was to examine the competing hypotheses that stereotype threat facilitates encoding of (a) only centrally arousing, negative feedback, or (b) both centrally arousing feedback and neutral information presented in conjunction with the negative feedback.

STUDY 2

Study 2 examined the extent to which stereotype confirming feedback is encoded with respect to stereotype-neutral information presented in conjunction with feedback in stereotype threatening contexts. Men and women completed the same math task described in Study 1 with one primary exception: Feedback was presented in different fonts and was embedded within neutral information (city and nature scenes). Consistent with Study 1, we hypothesized that compared to men, women under stereotype threat would encode negative feedback more efficaciously than positive feedback. Conversely, how negative feedback would be encoded in relation to information presented in the background of the feedback was less clear given the competing hypotheses described above. Thus, findings should provide insight into the extent to which negative feedback represents the centrally threatening stimulus or an arousing cue that facilitates encoding processes in general.

METHODS

Participants. Eighty-five participants (45 male) participated for course credit. All participants had knowledge of the negative stereotype pertaining to women's math ability (i.e., they responded with a 3 or lower on the stereotype knowledge question administered at the beginning of the semester, described above). Two women and one man were excluded from final analyses for either erroneously using scratch paper or a calculator on the math feedback task. Grubbs' test on positive and negative feedback and picture memory scores identified an additional four men as being significant outliers, that is, having Z scores less than or greater than 3.5 from the grand mean ($p < .05$). This left a final sample of 78 participants for analyses.

Procedure. Procedures were identical to Study 1 with modifications only being made to feedback stimuli on the math task and memory test. All participants were assigned to a diagnostic math test/stereotype threat (DMT) condition and instructed that they would complete tasks that were diagnostic of math intelligence. Participants were again asked to indicate their gender and sessions were always conducted with at least one male participant present. Participants then completed a math feedback task, math filler task, and memory test. Once these measures were completed participants completed anxiety questions and were debriefed.

Math Feedback Task. Participants completed the same math task described in Study 1 except that following the participants' response on each trial, wrong and correct feedback (written in different fonts) was superimposed on city and nature scenes. To avoid confounding feedback valence with font color, both wrong and correct feedback was presented in green. To ensure that a given font would never be uniquely associated with a given feedback type or city/nature scene, all fonts had an equal likelihood of being presented as either wrong or correct feedback with a given city or nature scene, and all participants had an equal likelihood of seeing a given font type as wrong or correct feedback with either a background city or nature scene.

Math Filler Task. Participants were given the same math filler task as in Study 1 and the same math test accuracy percentage was created.

Memory Test. A surprise memory test was given to participants after completion of the math filler task. The memory test contained 200 trials. Fifty of these trials contained different feedback fonts that participants may have seen during the math feedback task (25 possible wrong fonts and 25 possible correct fonts), 50 contained city or nature pictures that participants may have seen during the task (25 possible city scenes and 25 possible nature scenes), 50 contained wrong and correct font lures (25 wrong lures and 25 correct lures), and 50 contained city and nature scene lures (25 city scenes and 25 nature scenes). The memory test otherwise was similar to the test described in Study 1 in that each trial began with randomly presenting participants with different "Wrong" or "Correct" fonts on white backgrounds or different city and nature scenes and participants were asked to rate whether they thought they had seen the font or picture during the math feedback task using a 6-point scale with labels on 1 (*I know I didn't see it*), 4 (*I think I saw it*), and 6 (*I know I saw it*). Participants' responses were converted to binary responses of hits, misses, false alarms, or correct rejections and these responses were transformed in to d' scores. Again participants with no hits were given a hit score of 0.1 and participants with a perfect hit rate had 0.1 subtracted from their hit score.

Post-Task Anxiety. Participants were given the same questionnaire as in Study 1 and mean composites were created such that higher numbers represented more anxiety ($\alpha = .87$).

RESULTS

Difficult Math Test Performance and Post-Task Anxiety. An initial independent samples t-test was conducted on participants' accuracy scores on the difficult math test filler task. Similar to Study 1, these analyses revealed that men ($M = 54.95$, $SD = 31.72$) performed better on the difficult math test than did women ($M = 33.37$, $SD = 26.52$), $t(75) = 3.23$, $p < .01$, $d = .75$.[1] A comparable analysis conducted on participants' post-task anxiety composites indicated that women ($M = 3.10$, $SD = 1.18$) experienced more anxiety than men ($M = 2.53$, $SD = 1.06$), $t(69) = -2.17$, $p < .04$, $d = .52$.[2] Together, these findings suggest that our manipulation was successful at engendering a stereotype threatening situation among women.

Math Feedback Task. An independent samples t-test conducted on the number of sharp problems answered incorrectly on the first block of the math feedback task

1. Six participants did not complete this portion of the questionnaire.
2. Seven participants did not complete this portion of the questionnaire.

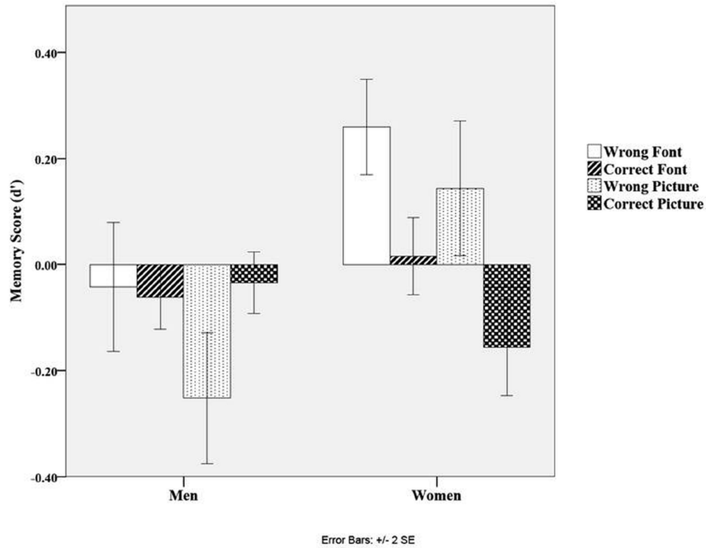

FIGURE 2. Men and women's d' memory scores as a function of feedback type and information presented in the background of feedback type. (Wrong/Correct font = recall for feedback; Wrong/Correct picture = recall for information presented with feedback type.)

indicated there were no differences between women (M = 9.24, SD = 4.04) and men (M = 7.65, SD = 4.25), $t(76) = -1.69$, $p = .10$; gender did not affect the amount of wrong and correct feedback participants received.

Memory Performance. To test our primary hypothesis, a 2(Gender: Men, Women) × 2(Feedback type: Wrong, Correct) × 2(Stimulus Type: Font, Picture) mixed factors ANOVA with repeated measures on the latter two variables was conducted on participants' memory scores for fonts and pictures. There were no effects for feedback type ($p = .20$), stimulus type ($p = .11$), stimulus type by gender ($p = .72$), feedback type by stimulus type ($p = .53$), or the three-way interaction ($p = .31$). There was, however, a main effect for gender $F(1, 76) = 6.57$, $p < .02$, $\eta^2 = .08$, that was qualified by a feedback type by gender interaction, $F(1, 76) = 7.67$, $p < .01$, $\eta^2 = .09$ (Figure 2; Table 1). Simple effects analyses indicated that women encoded both wrong fonts better than correct fonts, $F(1, 76) = 3.79$, $p = .055$, $\eta^2 = .05$, 95% CI [-.006, .49], and pictures associated with wrong feedback compared to pictures associated with correct feedback, $F(1, 76) = 3.83$, $p = .05$, $\eta^2 = .05$, 95% CI [-.005, .61]. These patterns were not evident among men (ps > .14). Women also encoded wrong fonts, $F(1, 76) = 3.90$, $p = .05$, $\eta^2 = .05$, 95% CI [-.002, .61], and pictures associated with wrong feedback, $F(1, 76) = 4.98$, $p < .03$, $\eta^2 = .06$, 95% CI [.04, .75], better than men. Gender did not affect the extent to which correct feedback or pictures associated with correct feedback were encoded (ps > .25). There were also no differences in either men or women in the extent to which they encoded wrong or correct fonts compared to pictures (ps > .09).

Exploratory Analyses on Link Between Stereotype Threat, Negative Memories, Performance, and Post-Test Anxiety. Serial mediational analyses comparable to Study 1 were conducted again to ascertain whether there was a link between stereotype threat, enhanced encoding of negative feedback on the math feedback task, future performance on difficult math tests, and general post-test anxiety. Unstandard-

TABLE 1. Mean Memory Scores (Standard Deviations in Parentheses), Total Hits, False Alarms and Standardized Hits and False Alarms for Women and Men Across Studies

		Study 1			
		DMT Women	DMT Men	PST Women	PST Men
Wrong Font	D′	.42 (.74)	-.20 (.66)	.04 (.58)	.11 (.79)
	Raw Hits	5.00 (3.29)	4.43 (2.25)	3.36 (1.87)	3.45 (3.25)
	Raw FAs	27.09 (10.30)	24.73 (11.47)	30.00 (9.72)	26.77 (12.35)
	Standard Hits	.22 (1.05)	.04 (.72)	-.31 (.60)	-.28 (1.03)
	Standard FAs	-.03 (.86)	-.22 (.96)	.21 (.81)	-.05 (1.03)
Correct Font	D′	-.08 (.45)	.00 (.51)	.06 (.49)	.13 (.28)
	Raw Hits	6.44 (3.0)	6.20 (3.44)	7.95 (3.98)	7.68 (3.21)
	Raw FAs	32.62 (12.21)	27.70 (10.33)	34.86 (9.91)	31.27 (9.74)
	Standard Hits	-.20 (.78)	-.26 (.91)	.20 (1.05)	.13 (.85)
	Standard FAs	.08 (1.05)	-.34 (.88)	.27 (.85)	-.03 (.83)
		Study 2			
Wrong Font	D′	.26 (.55)	-.04 (.77)		
	Raw Hits	3.38 (2.36)	2.92 (3.06)		
	Raw FAs	9.45 (6.13)	11.30 (5.88)		
	Standard Hits	.09 (.86)	-.08 (1.12)		
	Standard FAs	-.16 (1.01)	.15 (.97)		
Correct Font	D′	.02 (.45)	-.06 (.38)		
	Raw Hits	5.93 (3.68)	7.70 (4.07)		
	Raw FAs	14.21 (7.51)	17.40 (7.58)		
	Standard Hits	-.23 (.93)	.22 (1.03)		
	Standard FAs	-.21 (.98)	.20 (.99)		
Wrong Picture	D′	.14 (.78)	-.25 (.78)		
	Raw Hits	3.51 (3.00)	2.64 (2.47)		
	Raw FAs	17.37 (12.91)	23.47 (12.46)		
	Standard Hits	.16 (1.08)	-.15 (.90)		
	Standard FAs	-.24 (1.00)	.23 (.96)		
Correct Picture	D′	-.16 (.56)	-.03 (.37)		
	Raw Hits	4.19 (3.11)	6.88 (3.88)		
	Raw FAs	17.37 (12.91)	23.47 (12.46)		
	Standard Hits	-.37 (.83)	.35 (1.03)		
	Standard FAs	-.24 (1.00)	.23 (.96)		

Note. DMT = Diagnostic math test, PST = Problem solving task, FAs= False alarms, Standard = Standardized. False alarms are much higher because participants were exposed to markedly more lures compared to previously seen targets due to the nature of the study designs.

ized regression coefficients and 95% bias-corrected confidence intervals (CIs) from 10,000 bootstrap estimates (Hayes, 2013) were derived for men and women. Again, gender was utilized as a predictor, d' for wrong feedback served as an initial mediator, performance was a secondary mediator, and anxiety was the outcome variable. These analyses revealed that the indirect pathway was significant, $b = .03$, 95% CI [.0003, .1584], indicating that the relationship between gender and anxiety was mediated by memory for wrong feedback and test performance. Specifically, under threat, women compared to men showed greater memory for wrong feedback seen during the math feedback task, which in turn predicted worse performance on the subsequent difficult math test, which in turn predicted greater anxiety. The indirect pathways including memory for correct fonts, memory for pictures as-

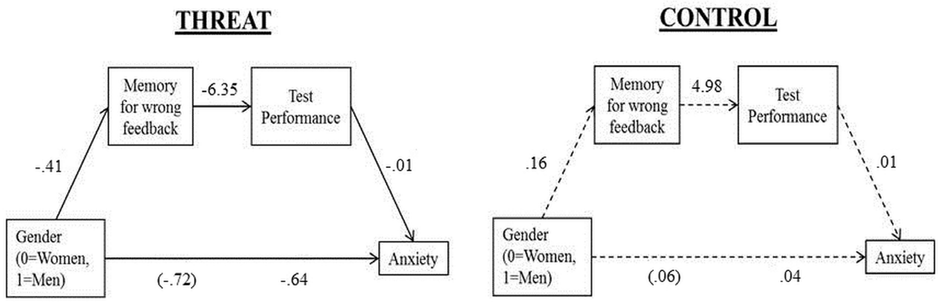

FIGURE 3. Serial mediation pathways showing that, under threat, women exhibited more efficacious encoding of wrong feedback than men, which predicted worse test performance on the future difficult math test, and increased anxiety. Solid paths are significant (95% CIs do not contain zero), whereas dashed paths are not significant (95% CIs contain zero). Values are unstandardized regression coefficients. Values in parentheses represent coefficients before mediators are entered in the model.

sociated with wrong feedback, and memory for pictures associated with correct feedback were all nonsignificant (i.e., the CIs contained 0).

To determine whether these patterns held in the aggregate across studies, serial mediation analyses were conducted on data from both Studies 1 and 2. In separate models for threat and control participants, gender was utilized as a predictor, *d′* for wrong feedback served as an initial mediator, performance was a secondary mediator, and anxiety was the outcome variable.

These analyses revealed that the indirect pathway indicating that better memory for stereotype confirming evidence was associated with underperformance on the difficult math test and greater anxiety held for women only, b = -.02, 95% CI [-.0936, -.0001] (Figure 3). There were also no relationships among these variables and memory for correct feedback.

Finally, because stereotype threatening situations also are known to elicit anxiety in general that has been linked to underperformance, we tested a model that utilized gender as a predictor, *d′* for wrong feedback served as an initial mediator, anxiety was a secondary mediator, and performance was the outcome variable. The indirect pathway for this analysis was not significant b = -.09, 95% CI [-.9483, .3431], suggesting that under stereotype threat women report more general post-test anxiety to the extent they encode negative performance related feedback, that is, stereotype-confirming evidence, and subsequently underperform on future difficult math tests.

DISCUSSION

Results from Study 2 demonstrated that stereotype threat augments encoding of negative feedback as well as information presented in conjunction with negative feedback. Consistent with Study 1, women under stereotype threat showed enhanced memory for negative feedback compared to positive feedback and men. Extending on these findings, women under stereotype threat demonstrated enhanced encoding of stereotype-neutral information (i.e., city and nature scenes) associated with wrong feedback compared to stereotype-neutral information associated with correct feedback. Conversely, males did not show this asymmetry in memory encoding. Also consistent with Study 1, stereotype threatened women reported more anxiety compared to men, suggesting that enhanced encoding of negative information may have been associated with emotional/mood congruent memory encoding processes. This anxiety was also in part augmented by the relationship between memory for stereotype-confirming evidence and subsequent underperformance on a future math test. Stereotype threatened women reported greater anxiety to the extent they exhibited enhanced recall of negative feedback and underperformed on a difficult follow-up math test. These results are consistent with findings from Murphy et al. (2007), which indicated that in viscerally arousing situations both information specific to an identity threat (stimuli presented in a gender unbalanced STEM video) and stimuli in the immediate surroundings are encoded more efficaciously by women compared to men and identity neutral contexts. We extend upon these findings by systematically demonstrating that stereotype-consistent performance feedback and stereotype-neutral information presented in conjunction with said feedback is more efficaciously encoded during performance situations in stigmatized domains.

GENERAL DISCUSSION

Results from two studies provide converging evidence that stereotype threatening contexts facilitate encoding of negative performance-related feedback compared to positive performance-related feedback. Study 1 provided evidence that stereotype threat facilitates encoding of negative feedback; women under threat exhibited greater memory sensitivity for fonts associated with wrong feedback compared to correct feedback, men, and women taking a supposed problem-solving task. Study 2 demonstrated that both stereotype-confirming, negative feedback and stereotype-neutral information presented in conjunction with negative feedback is encoded more efficaciously; women under stereotype threat again exhibited enhanced memory sensitivity for fonts associated with wrong feedback but also stereotype-neutral information presented in conjunction with wrong feedback. Importantly, when examining studies independently (Study 2 specifically) and combining data from the two studies, enhanced encoding of negative feedback on an initial math test predicted underperformance on a subsequent difficult math test, which in turn engendered greater post-test anxiety among stereotype threatened women only. These findings highlight both the deleterious consequences of biased negative feedback encoding on performance, and may provide insight in to why women may systematically report more negative affect in the math domain compared to men (Else-Quest et al., 2010).

Findings from both studies are consistent with recent work suggesting that stereotype threat biases attention toward negative feedback at a fundamental level and at the expense of performance on math tests similar to the math feedback task used in these studies (Forbes & Leitner, 2014). We extend upon these findings by providing evidence that this bias toward negative feedback engenders more efficacious encoding of negative feedback. Yet, enhanced encoding of either feedback type did not relate to performance on the math feedback task itself. This provides provocative evidence that stereotype threat can bias perceptions of one's performance that is orthogonal to their actual performance. That is, stigmatized individuals may perform comparable to their non-stigmatized counterparts, but if they are encoding negative, stereotype-consistent information at disproportionate levels, they may still leave a stereotype threatening situation with the perception that they performed poorly. This notion is supported by recent work indicating that despite stereotype threatened minorities learning implicit relationships on a supposed intelligence test at comparable rates compared to whites, they still believed they underperformed to the extent they exhibited decreased connectivity between brain regions integral for self-oriented processing and coping at rest (Forbes et al., 2015).

Relevant to this point, findings from Study 2 indicate that stigmatized individuals are likely to encode both stereotype-consistent and stereotype-neutral information presented in conjunction with the stereotype-consistent stimulus in stereotype threatening contexts. Such findings are consistent with evidence suggesting women attend to irrelevant information in stereotype threatening contexts (Inzlicht & Kang, 2010) and possibly encode irrelevant information presented in social identity threatening situations (Murphy et al., 2007). Our results extend upon these findings by suggesting that specific negative performance relevant feedback, that is, negative feedback written in a specific font, received while stigmatized individuals performed (a) was more efficaciously encoded compared to positive feedback and was not a product of guessing, (b) was not evident in stereotype-neutral contexts, (c) was only recalled to the extent the information was consistent with the negative stereotype, and (d) undermined performance on a future difficult math test to engender general post-test anxiety.

Furthermore, in Study 2 the centrally arousing stimulus was provided to individuals in the form of stereotype confirming feedback. But what form does this centrally arousing stimulus take in other stereotype threatening contexts? In real world settings, while individuals are likely to get feedback in classroom settings, they do not typically receive feedback while they complete GRE or math AP exams; thus it is possible that negative thoughts associated with the exam and perceived performance become the centrally arousing stimulus. Past research implicates a prominent role for negative thoughts, doubt, and anxiety-oriented thoughts experienced while individuals perform under stereotype threat (e.g., Cadinu et al., 2004; Schmader et al., 2008). Thus, in the absence of direct feedback, it's quite possible negative thoughts associated with perceived performance, task difficulty, or even manifestations or interpretations of the stress-oriented visceral arousal engendered by threatening contexts (e.g., "why can't I stop sweating?" or "why is it so hot in here?") are better encoded as well. Ultimately, the centrally arousing stimulus is identified as that stimulus that prevents the actualization of a primed goal state (Levine & Burgess, 1997; Levine & Pizarro, 2004; Oatley & Johnson-Laird, 1987) and therefore can take myriad forms, including overt feedback, or a thought or feeling that is perceived by the stigmatized person to interfere with his or her performance and ultimate success. With that said, future research would be necessary to determine how neutral information is encoded when it is presented in competition with the centrally arousing stimulus. That is, in this study neutral information was presented in conjunction with negative feedback which could have confounded the neutral information with the threatening stimulus (i.e., the neutral information was perceived to be a component of the threatening stimulus). We would hypothesize that when a centrally threatening stimulus is presented in competition with the neutral information that only the threatening stimulus would be encoded more efficaciously.

Both in Study 2 and when the data from the two studies were combined, the deleterious consequences of enhanced encoding of negative feedback on performance and anxiety emerged. While it is always difficult to make definitive conclusions regarding directionality and causation in mediation analyses, our serial mediation analyses suggested that stereotype threatened women underperformed on a difficult math test to the extent they encoded negative feedback more efficaciously on a previous math task (in which there were no performance differences, only biased encoding responses). Underperformance on the difficult math test, in turn, fostered greater post-test anxiety. How and why this happened is a more difficult question to answer, but one possibility is that exposure to a second math test prompted the spontaneous recall of negative feedback that in turn interfered with cognitive resources necessary to perform optimally. One could also argue, however, that underperformance was due to negative memory recall decreasing confidence or engendering more negative expectations for performance which then prompted stereotype threatened individuals to perform in line with these new negative expectations. Another possibility is that biased, exacerbated attention and encoding of negative information is cognitively taxing in and of itself, which would deplete cognitive resources at disproportionate levels among stereotype threatened individuals, and compromise performance on difficult downstream tasks accordingly. Finally, it is also possible that anxiety had a more general effect on performance (i.e., anxiety engendered underperformance), as previous research has suggested (e.g., Eysenck & Calvo, 1992; Eysenck & Derakshan, 2011). However, this path was not significant when tested, nor is it consistent with the temporal nature in which memory, performance, and anxiety measures were collected across studies rendering this conceptualization of the model outside the scope of the present studies. Thus future research would be necessary to determine which of these possibilities is the more adequate explanation for why negative feedback encoding decreased performance under stereotype threat but it's possible that all routes play some role in undermining performance overall.

Regardless of the precise mechanism, these findings are particularly provocative because of their implications for stigmatized individuals' performance and domain perceptions. Recall that memory measures were obtained from the moderately difficult math feedback task where performance differences between men and women were not evident. This suggests then that even if women perform comparably to men on a math test they might not only be more likely to encode negative feedback relative to their performance, but this enhanced encoding may ultimately undermine their performance on future difficult math tests. This underperformance, in turn, culminates in fostering general feelings of anxiety toward the domain that the stigmatized individual leaves the testing situation with. This process is bound to have ramifications for individuals' identification with a stigmatized domain over time. Indeed, given that stereotype threat, or social identity threat more broadly, can be triggered in various, oftentimes innocuous ways (e.g., via solo status or even posters on the wall), one manifestation of this process could be more negative perceptions and reduced confidence in the stigmatized domain not unlike those found among girls in Else-Quest et al.'s (2010) cross-cultural study.

These studies also raised some interesting questions. For instance, it is unclear why men in the DMT condition in both studies exhibited poor encoding of either type of feedback. It could be that the feedback is not deemed important or motivationally relevant in situations where one is expected to excel, and thus men in the DMT condition simply did not attend to the feedback. Such an interpretation would be consistent with the fact that while men performed poorly on the memory test they outperformed women in the DMT condition and performed comparably to everyone else on the difficult math test. The disconnect between performance on the memory test compared to the math test was found among all participants, however, and speaks to the possibility that regardless of performance, men may not leave diagnostic math testing situations with the same negative memories as women.

The studies presented here lay an important foundation for the deleterious effects of stereotype threat on memory encoding. Future research should directly examine whether enhanced encoding of negative feedback that women demonstrated under stereotype threat is a product of the same stress-oriented responses driven by the HPA-glucocorticoid system. To the extent stereotype threat engenders feelings of stress in stigmatized individuals, it is possible that cortisol-dependent long-term potentiation of associations between negative information and stereotype threatening contexts is the precise mechanism that underlies these findings. Past research has provided indirect evidence for the presence of cortisol in stereotype threatening situations (Matheson & Cole, 2004), but more research is needed to provide causal evidence for the role of cortisol, as well as the amygdala and hippocampus, in facilitating encoding of negative information received in stereotype threatening situations.

In sum, findings from two studies provide converging evidence that stereotype threat enhances encoding of negative feedback to undermine performance and engender anxiety. This suggests that the stereotype threat experience may endure in the mind of the stigmatized after the test and not necessarily reflect the reality of their performance. Rather, these situations lay the foundation for encoding negative aspects about an event that may persist through time. These findings not only provide more evidence that stereotype threat is a highly pervasive problem that can undermine performance, they highlight a new threat suggesting that stigmatized individuals' failures may be unforgettable.

REFERENCES

Beilock, S. L., Rydell, R. J., & McConnell, A. R., (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spill-over. *Journal of Experimental Psychology: General, 136*, 256-276.

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

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.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 16*, 572-578.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D. E., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience, 20*, RC99.

Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, *3*(4), 286-300.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*, 355-391.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103.

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, *6*(6), 409-434.

Eysenck, M. W., & Derakshan, N. (2011). New perspectives in attentional control theory. *Personality and Individual Differences*, *50*(7), 955-960.

Forbes, C. E., & Leitner, J. B. (2014). Stereotype threat engenders neural attentional bias towards negative feedback to undermine performance. *Biological Psychology, 102*, 98-107. doi: 10.1016/j.biopsycho.2014.07.007

Forbes, C. E., Leitner, J. B., Jordan, K., Magerman, A., Schmader, T., & Allen, J. J. B. (2015). Spontaneous default mode network phase locking moderates performance perceptions under stereotype threat. *Social Cognitive and Affective Neuroscience, 10*(7), 994-1002.

Forbes, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of Personality and Social Psychology, 99*, 740-754.

Forbes, C. E., Schmader, T., & Allen, J. J. (2008). The role of devaluing and discounting in performance monitoring: A neurophysiological study of minorities under threat. *Social Cognitive and Affective Neuroscience, 3*, 253-261.

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*, 17-28.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Science, 5*, 394-400.

Hamann, S. B., Ely, T. D., Grafton, S. T., & Kilts, C. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience, 2*, 289-293.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford.

Inzlicht, M., & Kang, S. K. (2010). Stereotype threat spillover: how coping with threats to social identity affects aggression, eating, decision making, and attention. *Journal of Personality and Social Psychology*, *99*(3), 467-481.

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

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.

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience, 7*, 54-64.

LaBar, K. S., & Phelps, E. A. (1998). Arousal-mediated memory consolidation: Role of the medial temporal lobe in humans. *Psychological Science, 9*, 527-540.

Leitner, J. B., Jones, J. M., & Hehman, E. (2013). Succeeding in the face of stereotype threat: The adaptive role of engagement regulation. *Personality and Social Psychology Bulletin, 39*, **17-27.**

Levine, L. J., & Burgess, S. L. (1997). Beyond general arousal: Effects of specific emotions on memory. *Social Cognition, 15*, 157-181.

Levine, L. J., & Pizarro, D. A. (2004). Emotion and memory research: A grumpy overview. *Social Cognition, 22*, 530-554.

Matheson, K., & Cole, B. M. (2004). Coping with a threatened group identity: Psychosocial and neuroendocrine responses. *Journal of Experimental Social Psychology, 40*, 777-786.

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

Oatley, K., & Johnson-Laird, P. N. (1987). Toward a cognitive theory of emotions. *Cognition and Emotion, 1*, 29-50.

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology, 129*, 242-261.

Osborne, J. W. (2006). Gender, stereotype and threat anxiety: Psychophysiological and cognitive evidence. *Journal of Research in Educational Psychology, 8*, 109-138.

Osborne, J. W. (2007). Linking stereotype and threat anxiety. *Educational Psychology, 27*, 135-154.

Payne, J. D., Jackson, E. D., Hoscheidt, S., Ryan, L., Jacobs, W. J., & Nadel, L. (2007). Stress administered prior to encoding impairs neutral but enhances emotional long-term episodic memories. *Learning and Memory, 14*, 861-868.

Payne, J. D., Jackson, E. D., Ryan, L., Hoscheidt, S., Jacobs, W. J., & Nadel, L. (2006). The impact of stress on neutral and emotional aspects of episodic memory. *Memory, 14*, 1-16.

Paynter, C. A., Reder, L. M., & Kieffaber, P. D. (2009). Knowing we know before we know: ERP correlates of initial feeling-of-knowing. *Neuropsychologia*, *47*(3), 796-803.

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

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review, 115*, 336-356.

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.

Stangor, C., Carr, C., & Kiang, L. (1998). Activating stereotypes undermines task performance expectations. *Journal of Personality and Social Psychology, 75*, 1191-1197.

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

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of women and ethnic minority students. *Psychological Science, 20*, 1132-1139.

Wickens, T. D. (2002). *Elementary signal detection theory.* New York: Oxford University Press.

Wraga, M., Helt, M., Jacobs, E., & Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognition and Affective Neuroscience, 2*, 12-19.