

Emotion blocks the path to learning under stereotype threat

Jennifer A. Mangels,^{1,2} Catherine Good,¹ Ronald C. Whiteman,² Brian Maniscalco,³ and Carol S. Dweck⁴

¹Department of Psychology, Baruch College, Baruch College of City University of New York, New York, NY 10010, ²Department of Psychology, Program in Cognitive Neuroscience, The City College of the City University of New York, New York, NY 10031, ³Department of Psychology, Columbia University, New York, NY 10027, and ⁴Department of Psychology, Stanford University, Stanford, CA 94305, USA

Gender-based stereotypes undermine females' performance on challenging math tests, but how do they influence their ability to learn from the errors they make? Females under stereotype threat or non-threat were presented with accuracy feedback after each problem on a GRE-like math test, followed by an optional interactive tutorial that provided step-wise problem-solving instruction. Event-related potentials tracked the initial detection of the negative feedback following errors [feedback related negativity (FRN), P3a], as well as any subsequent sustained attention/arousal to that information [late positive potential (LPP)]. Learning was defined as success in applying tutorial information to correction of initial test errors on a surprise retest 24-h later. Under non-threat conditions, emotional responses to negative feedback did not curtail exploration of the tutor, and the amount of tutor exploration predicted learning success. In the stereotype threat condition, however, greater initial salience of the failure (FRN) predicted less exploration of the tutor, and sustained attention to the negative feedback (LPP) predicted poor learning from what was explored. Thus, under stereotype threat, emotional responses to negative feedback predicted both disengagement from learning and interference with learning attempts. We discuss the importance of emotion regulation in successful rebound from failure for stigmatized groups in stereotype-salient environments.

Keywords: LPP; FRN; ERP; math; emotion regulation; memory

Gender gaps in math achievement widen as women progress through the academic and career pipeline. Although various reasons have been considered for this gap (Ceci *et al.*, 2009), there is increasing evidence that stereotype threat (ST) is an important societal factor that may be preventing women from reaching their full potential (Spencer *et al.*, 1999). ST occurs when a member of a group fears they will confirm a stereotype of poor performance in a given domain, which often leads to performance decrements for individuals who identify strongly with both the stereotyped group and domain (Steele, 1997). Although much research has been aimed at understanding the mechanisms underlying its effects on performance outcomes, there has been little published work addressing its effects on learning (Rydell *et al.*, 2010a,b), and none that has directly addressed the underlying cognitive and emotional mechanisms for any observed learning decrements.

Under what circumstances would ST affect learning the most? When material is new or difficult, the probability of a

negative outcome is increased. If ST is 'in the air,' stigmatized individuals may interpret this negative feedback as confirmation of the stereotype of low ability, rather than as a natural part of the learning process. These negative appraisals of ability could then activate negative thoughts and feelings about the self that interfere with effective engagement with learning opportunities, the very opportunities that could foster mastery of the material. Over time, ineffective learning could translate into real achievement gaps that could perpetuate negative stereotypes. Yet, even in challenging, stereotype-activating situations, some females are able learn effectively (Good *et al.*, 2008). What are the factors that mitigate the damaging effects on learning?

The present study focused on the emotional and cognitive processes that differentiate those females who rebound from math errors under ST from those who succumb. To this aim, we created a math-based version of a paradigm we have previously used to study error correction in general knowledge domains (Butterfield and Mangels, 2003; Mangels *et al.*, 2006). In this task, responses to challenging math problems were followed by both accuracy feedback and an opportunity to engage with an on-line math tutor after each problem. Later, participants were given a surprise retest to evaluate the extent to which they were able to learn from the tutor and rebound from their initial errors. Our primary questions concerned whether ST influences females' emotional responses to negative feedback, and whether these emotional responses predicted: (i) use of the tutor

Received 8 July 2010; Revised 11 October 2010; Accepted 29 November 2010

Advance Access publication 19 January 2011

The authors thank Matthew Greene for assistance with programming, Joanie Dean for assistance with creating math stimuli, Justin Lamb for assistance with testing subjects and Belén Guerra-Carrillo for assistance with preparation of the article. The authors also acknowledge the contributions of the many undergraduate research assistants who dedicated their time to various aspects of this project. This work was funded by the Institute for Educational Sciences (IES) Cognition and Student Learning (CASL) Grant (to J.A.M., C.G., and C.S.D.).

Correspondence should be addressed to Jennifer A. Mangels, Department of Psychology, Baruch College, 55 Lexington Ave., NY, NY 10010. E-mail: jenimangels@gmail.com

(i.e. engagement with learning) and (ii) the ability to correct the errors on a retest (i.e. success in learning).

Much recent research indicates that emotional processing interferes with cognitive processing under ST (Wraga *et al.*, 2007; Krendl *et al.*, 2008), either by directly or indirectly sapping the working memory resources needed for successful task performance (for review see Schmader *et al.*, 2008). These emotion-related costs appear to stem both from increased vigilance for threat-relevant information, such as signs of failure (Forbes *et al.*, 2008), as well as counterproductive attempts to regulate the negative emotions elicited by this information, such as through rumination (Beilock *et al.*, 2007) or emotion suppression (Johns *et al.*, 2008). Although both vigilance and maladaptive emotion regulation may disrupt learning, they may do so through somewhat different cognitive mechanisms. Studies of anxious individuals sometimes find that increased vigilance for threat-relevant information is followed by active avoidance of that information when it is detected, as means of regulating the anxiety it provokes (Mogg *et al.*, 2004; Buckner *et al.*, 2010). Indeed, studies have shown that some individuals under ST will buffer threats to self-esteem by dissociating themselves from threat-relevant information (Cohen *et al.*, 1999). Extending this to the present study, greater vigilance and initial sensitivity to negative feedback may undermine learning by motivating individuals to regulate their arousal by distancing themselves from the tutorial learning opportunity. In contrast, attempts to regulate one's emotional response by focusing on the response itself, through either rumination or suppression, will likely divert cognitive resources from processing task-relevant information (Richards and Gross, 2000, 2006). If this were the case, individuals may appear motivated to investigate the tutor, but their ability to deeply process the information it provided would be limited by the demands of concurrently monitoring the persistent thoughts and emotions the negative feedback elicited.

Given the different roles that initial threat-sensitivity and sustained emotional processing may play in ST effects on learning, we employed the high temporal resolution of event-related potentials (ERPs) to capture how emotional responses to negative feedback unfold under ST without invoking the demand characteristics of self-report (Spencer *et al.*, 1999). We focused our analysis on three ERP waveforms: the medial frontal feedback-related negativity (FRN), the anterior P3 (P3a) and the posteriorly-maximal late positive potential (LPP). Drawing upon previous research, we relate these waveforms, respectively, to the relatively automatic appraisal of feedback valence (Simons, 2010), the subsequent orienting to the feedback as a function of its motivational salience (Friedman *et al.*, 2001; Mangels *et al.*, 2006), and the sustained attentional processes that are sensitive to the subjective level of emotional arousal the feedback elicits (Hajcak *et al.*, 2010). We expect that, together, the FRN and P3a will primarily measure the effects

of ST on individuals' initial sensitivity to feedback valence, whereas the LPP should track the dynamics of arousal up-regulation, as might be produced by rumination or suppression (Hajcak *et al.*, 2010). Although the LPP has traditionally been assessed with complex picture stimuli (e.g. faces, scenes), it is relatively insensitive to stimulus complexity or repetition (Codispoti *et al.*, 2006), and thus, should serve as an informative index of students' emotional response even to simple feedback stimuli.

In order to differentiate the two proposed mechanisms by which the emotional response to feedback may influence learning success under ST (i.e. disengagement from learning *vs* interference with learning from distracting thoughts), our learning opportunity took the form of an interactive math tutor where post-error problem-solving information would only be revealed by the participant's active engagement. Specifically, strategies for solving a given problem type appeared only when students made mouse clicks within the tutor framework. In addition, because we did not inform students that they would need to learn this information for an upcoming retest, they would not have expected any immediate negative consequences for disengaging from the tutor before all steps had been fully explored. As such, the proportion of tutor steps participants chose to examine served as a straightforward measure of their 'quantity' of knowledge seeking following errors, and thus, would be sensitive to disengagement from active learning. On the other hand, if attempts at engagement were rendered ineffective because of interference from distracting thoughts, we might find that even extensive tutor exploration would result in little error correction. In other words, this type of interference would not necessarily reduce students' motivation to explore the tutor, but it would reduce the 'quality' of any knowledge seeking attempts.

By examining both the quantity and quality of tutor processing we may gain insight into the means by which emotional processing, as indexed by our ERP measures, may influence learning under ST. An understanding of these emotion-cognition relationships could provide a foundation for interventions that would promote successful learning and rebound from failure for stereotyped individuals even when stereotypes are highly salient and unavoidable.

MATERIALS AND METHODS

Subjects

We tested 71 English-speaking, right-handed, neurologically healthy, consenting undergraduate females at Columbia University. Both Caucasians ($N=61$) and non-Caucasians ($N=10$) were tested, although we did not include Asians given the potential for conflict between gender- and race-based stereotypes (Shih *et al.*, 1999). After excluding one student who never used the tutor and three subjects with EEG recording difficulties, there were 32 and 36 students in the ST and non-threat (NT) conditions, respectively.

Design and procedure

The task took place over three days. On the first day, participants completed pre-measures including gender and math identification, and their perception of environmental stereotype threat (PEST; Good *et al.*, manuscript under review), a measure of the extent to which participants believe that people in their environment hold negative stereotypes about women in math. The second testing session began with some additional questionnaires assessing mood and confidence in math ability, after which participants were prepared for EEG recording (see details below). Once EEG preparation was complete, participants were first instructed about the procedure for the math tasks (see details below), and then received either the ST or NT framing (see details below). Participants completed 48 math problems while EEG was recorded. Approximately 24 h later, participants returned for the third session, during which they were presented with a surprise retest of math problems that were isomorphic to those from the first test (i.e. involving identical concepts but using different numbers). Some students reported suspicions that they might be given more math questions during this final session, but none reported

expecting to be retested on the original problem set or having used the tutorial for that purpose. At the conclusion of the retest, participants answered some manipulation check questions, and reported their math SAT score (Table 1).

Math task and procedure

We created 48 multiple-choice GRE-like math questions and 48 isomorphic 'yokes'. Each question tapped a different math concept in order to strengthen inferences about the relationship between tutor use on a given problem and likelihood of solving a yoked retest question. For each question, a tutorial was created that provided students with a step-wise solution for that problem. Solutions were either 3-steps (32 problems) or 4-steps (16 problems) depending on problem complexity.

At first test, the math problems were randomly presented in 4 blocks of 12 questions. Each trial began with the presentation of a math problem (see Figure 1A for illustration of a trial sequence). Participants then had 1-min to click on their multiple-choice answer. After the answer was recorded, they rated their confidence in their answer on a Likert scale

Table 1 Subject characteristics

	Stereotype threat		Non-threat		Significance testing		
	Better learners	Poorer learners	Better learners	Poorer learners	Threat (T)	Learn (L)	TxL
Demographic							
N	16	16	18	18			
Age (years)	20.2 (0.46)	20.3 (0.46)	20.7 (0.43)	19.9 (0.43)			
Education (years) ^a	14.3 (0.32)	14.3 (0.32)	14.4 (0.31)	14.2 (0.31)			
Gender ID ^b	6.4 (0.29)	6.3 (0.29)	5.8 (0.27)	6.4 (0.27)			
Math characteristics							
Math SAT	689.4 (16.7)	701.3 (16.7)	674.4 (16.7)	705.6 (16.7)			
Math confidence ^c	3.2 (1.0)	2.8 (1.0)	4.4 (0.98)	1.9 (0.98)			
Math ID ^d	7.3 (0.68)	8.5 (0.68)	9.0 (0.64)	7.6 (0.64)			*
PEST ^e	4.1 (0.43)	4.7 (0.43)	4.8 (0.40)	4.5 (0.40)			
State-trait anxiety index (STAI)^f							
Pre-instruction	1.8 (0.12)	1.7 (0.11)	1.6 (0.11)	1.6 (0.11)			
Post-test	2.0 (0.11)	1.8 (0.11)	1.6 (0.10)	1.6 (0.10)	**		
Post-test perceptions							
First-test difficulty ^g	4.8 (0.27)	4.8 (0.27)	1.7 (0.25)	1.2 (0.25)	*		
Expectation ^h	5.1 (0.21)	5.1 (0.21)	4.3 (0.39)	4.6 (0.39)	**		
MC: math IQ ⁱ	4.8 (0.27)	4.8 (0.27)	1.7 (0.25)	1.2 (0.25)	**		
MC: gender Differences ^j	5.2 (0.27)	5.0 (0.27)	1.7 (0.25)	1.3 (0.25)	**		
MC: problem solving ^k	4.9 (0.26)	4.2 (0.26)	5.7 (0.24)	5.9 (0.24)	**		

Note: Mean \pm s.e.m., * $P \leq 0.05$, ** $P \leq 0.01$. ^aYears of education. Year in college was added to an assumed 12 years of pre-college education. ^bGender identity. Average rating on four questions (e.g. Being a woman is an important reflection of who I am), scaled from 1 (strongly disagree) to 8 (strongly agree); two questions were reverse coded. ^cConfidence in math ability. Total sum of ratings across three questions (e.g. When I get new math material, I'm usually sure I will be able to learn it), each scaled from -3 (low confidence) to +3 (high confidence). Overall values could thus range from -9 (low confidence) to +9 (high confidence). ^dIdentification with the math domain. Average rating on five questions (e.g. Is your math ability important to you?), scaled from 1 (not at all) to 15 (extremely). ^ePerception of environmental stereotype threat. Average rating on six questions (e.g. In the context of math abilities, other people believe that females can do just as well as males in mathematics.), scaled from 1 (strongly agree) to 8 (strongly disagree). ^fState anxiety portion of the state-trait anxiety index (STAI). Average rating on 20 questions (e.g. I feel nervous), scaled from 1 (not at all) to 4 (very much so). ^gIn general, how difficult did you find the first set of math problems? Rating scale: 1 (not at all) to 15 (extremely). ^hWhat did you think were the expectations of the experimenters regarding the performance of male students relative to female students? Rating scale: 1 (females would perform better) to 7 (males would perform better). ⁱThreat manipulation check. 'The purpose of these math problems was to measure my math intelligence.' Rating scale for this and all manipulation check (MC) questions (i-k): 1 (strongly disagree) to 6 (strongly agree). ^jThreat manipulation check. 'The purpose of these math problems was to measure gender differences.' ^kNon-threat manipulation check. 'The purpose of these math problems was to investigate psychological factors involved in problem solving.'

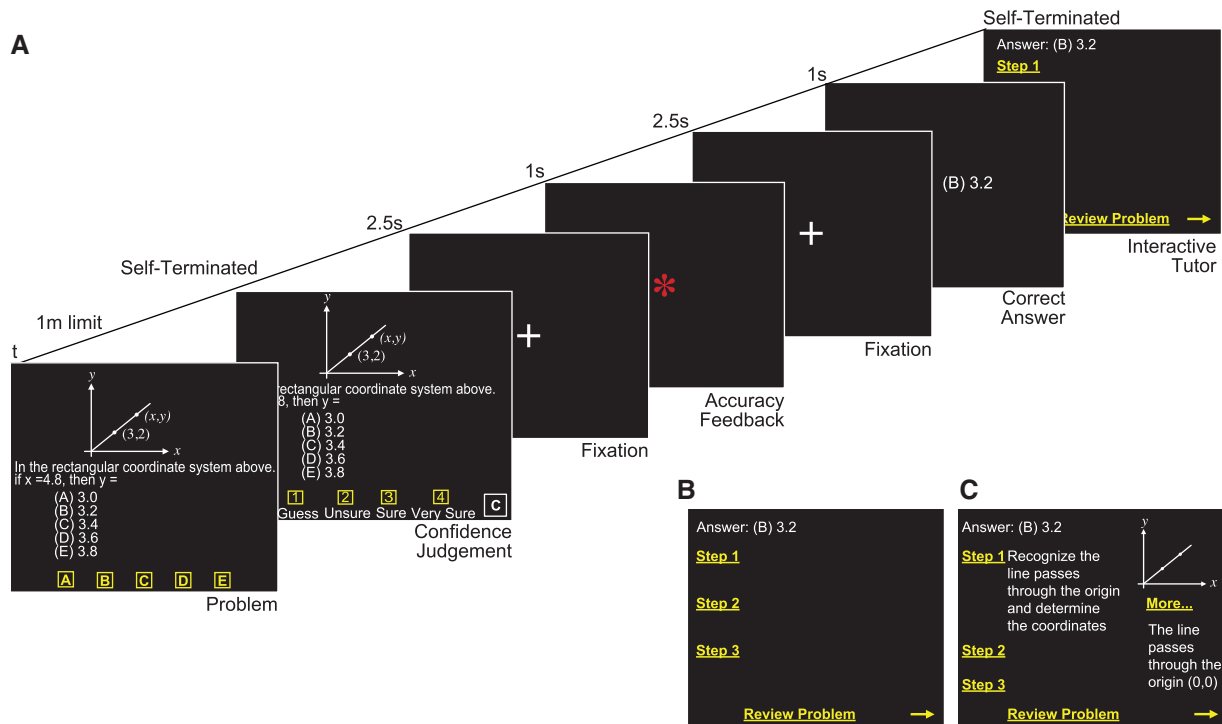


Fig. 1 Trial structure. (A) Schematic of a first-test trial when the answer was incorrect. Subjects selected a multiple-choice answer within a 1m time limit, then were given unlimited time to judge their confidence in that answer using a 4-point scale (1 = guess, 2 = unsure, 3 = sure, 4 = very sure). After confidence was entered, the feedback sequence automatically commenced. During accuracy feedback, the red asterisk associated with negative feedback was accompanied by a low tone. If the subject's answer had been correct, all feedback components would have been identical except that accuracy feedback would have been a green asterisk and high tone. (B) Example of initial screen of the tutor for problem shown in (A). This initial screen showed the correct answer, placeholders for the steps, a link to the original question ('review problem'), and a yellow arrow in the lower right corner used to advance to the next trial. Problems with a graphical element would show this element in the upper right corner. (C) Example of tutor for problems shown in (A) when Step 1 and its associated 'More' had been revealed. Clicking on a step revealed a key procedure in general conceptual terms, and activated the link to an associated 'More', which if clicked, revealed concrete procedural details of that step.

ranging from 1 (not at all sure correct) to 4 (very sure correct). A crosshair fixation appeared for 2.5s, followed by accuracy feedback presented for 1s. A green asterisk and a high tone or a red asterisk and a low tone indicated correct or incorrect responses, respectively. A second crosshair fixation then appeared for 2.5s, followed by the correct answer for 1s. Finally, the interactive math tutor appeared for a duration that was self-terminated by the student. Figure 1B shows an example of the tutor before any clicks were made. Figure 1C provides an example of what was shown for that item if Step 1 and More 1 were clicked.

To make the first test suitably challenging, mean difficulty within a block was set at ~65%, using difficulty ratings obtained through a prior normative study conducted with male and female Columbia University undergraduates ($N=434$) under NT, untimed conditions. For the present study, we only included questions that did not show gender differences in the normative sample.

Retest and first-test procedures were identical, with the exception that no tutors were offered during the retest and EEG was recorded only during first test. For each yoked pair of math problems, the order in which the items were presented across first test and retest was randomized across subjects.

ST manipulation

The first test began with visually displayed ST or NT instructions concurrently spoken by a recorded male voice. This manipulation was similar to that used in previous studies (Spencer *et al.*, 1999). Specifically, the ST instructions stressed that math intelligence/ability would be assessed ('our aim is to understand what makes some people better at math than others'), and that participants' performance would be compared to others ('we will be comparing your score to other students for the purpose of studying gender differences in math'). Students also explicitly indicated their gender. In contrast, the NT condition explicitly de-emphasized ability comparisons ('the purpose of this study is not to see how smart you are'), and stressed an interest in problem-solving ('rather, we want to examine psychological processes associated with effortful problem solving'). The potential for gender differences was also explicitly minimized ('analysis of thousands of students' results has shown that males and females perform equally well on this problem set'), and students did not indicate their gender at the start of the test (see Supplementary Data for verbatim instructions). A brief version of the relevant instruction was repeated at the start of each block, as well as at the beginning of the retest.

As shown in Table 1, the subjective experience of the participants indicated that the ST manipulation appeared to have functioned as intended. Students in the ST condition reporting greater overall levels of post-retest anxiety, greater difficulty with the first-test, and greater likelihood of perceiving the experimenter to be biased (i.e. expecting that males would perform better than females). The effectiveness of the manipulation was further verified by the pattern of students' ratings on the manipulation check questions.

EEG recording and measurement

Continuous EEG was recorded during the first test from 64 sintered Ag/AgCl electrodes with an A/D conversion rate of 500 Hz and band-pass of DC-100 Hz. Impedance was kept below 11 k Ω . EEG was initially referenced to Cz and converted to an average reference off-line. We compensated for blinks and other eye movement artifacts with PCA-derived ocular components. Off-line, EEG was cut into 1100 ms epochs starting 100 ms prior to the feedback onset. Following baseline correction, we rejected epochs containing excessive noise (± 120 mV), applied 0.15 Hz high-pass and 35 Hz low-pass, zero-phase filters, and then averaged to create the ERPs.

To increase signal-to-noise ratios, we excluded participants with less than nine trials (Olvet and Hajcak 2009), resulting in elimination from the EEG analysis of two NT subjects who erred on less than nine first-test questions. These subjects were included in our behavioral analysis so as not to bias the NT group against better performers. There were no significant differences in trial count between groups for either negative (average = ~ 18 trials) or positive (average = ~ 25 trials) feedback.

Measurement of the FRN and P3a were based on mean amplitudes at Fz from 250 to 298 ms and from 300 to 400 ms, respectively (Mangels *et al.*, 2006; Van Meel and Van Heijningen 2010). Although the LPP also emerges around 250 ms, we attempted to minimize its overlap with the P3a by focusing our analysis from 400 to 1000 ms, at its Cz maximum (Hajcak *et al.*, 2010).

Data analysis

We took both univariate and multivariate approaches to understand the relationship between the feedback response (ERP), tutor use and successful learning as a function of ST. Specific details of how each type of analysis was conducted are provided in each corresponding Results section. Here, we describe aspects of the analysis that were common across both approaches.

First, we defined successful learning in both analyses as the proportion of isomorphic retest items where the subject had: (i) made some attempt to use the tutor after making the error at first test (i.e. at least one click in the tutor environment) and (ii) successfully solved the isomorphic problem at retest with greater than guessing levels of confidence (i.e. confidence > 1). These criteria increased the likelihood that

accurate error correction would be linked to the successful application of problem solving strategies learned or refreshed during tutor exploration, rather than a lucky guess or retrieval-exclusive processes.

Math SAT and first test accuracy were included as control variables in both analyses. In the univariate analysis, Math SAT was served as a covariate in all analyses of first-test performance, tutor use and retest performance, given the potential for pre-existing knowledge to influence these variables. First-test performance was included as a covariate in analyses involving the ERPs of interest or retest performance, but not for analyses involving tutor use. The number of errors experienced at first test could have influenced both students' affective experience in the task and the amount of information they needed to learn and retrieve. Indeed, it served as a significant covariate in our analysis of overall retest accuracy ($P < 0.05$) and a marginally significant covariate in our analysis of successful learning ($P = 0.1$). First-test performance was included as a covariate for analyses of the FRN, P3a and LPP, given that the amplitude of these waveforms may be larger when the subjective probability of an error is lower (Holroyd and Coles, 2002; Butterfield and Mangels, 2003; Ito *et al.*, 2004; Iwanaga and Nittono, 2010). Indeed, it served as a significant covariate for all ERP analyses (all F 's > 4 , P 's < 0.05). In the multivariate model, these control variables (Math SAT and first-test performance) were included as separate paths to each of the relevant outcome variables.

RESULTS

Univariate analyses

Our univariate approach primarily assessed the overall effects of ST on our measures of interest. In addition, in order to capture how each of these measures may be related to successful learning and rebound from failure, these analyses included an individual differences measure of 'learning success', whereby subjects were characterized as 'better' or 'poorer' learners based on a post hoc median split of the residuals for retest error correction after regressing out Math SAT and first-test performance (with group means added back: ST median = 0.47; NT median = 0.48).

This individual differences variable of 'learning success' can be viewed as a between-subjects adaptation of the subsequent memory paradigm commonly used to isolate the neural correlates of successful encoding processes (Paller and Wagner, 2002). Thus, all univariate analyses were conducted with 2 (threat: ST vs NT) \times 2 (learning: better learners vs poorer learners) mixed-measures ANOVAs/ANCOVAs (see above for further description of covariates).

Gender and math-relevant individual differences

As shown in Table 1, Gender Identification did not differ as a function of either ST or subsequent learning success. Perceptions of others' stereotypes about females' abilities relative to males' in math (PEST) were also similar across

Table 2 First test and retest performance, error correction at retest, and tutor use

	Stereotype threat		Non-threat		Significance testing		
	Better learners	Poorer learners	Better learners	Poorer learners	Threat (T)	Learn (L)	TxL
First test ^a							
P(correct)	0.54 (0.02)	0.56 (0.02)	0.61 (0.02)	0.59 (0.02)	*		
Confidence (all)	2.7 (0.10)	2.7 (0.10)	2.8 (0.09)	2.8 (0.09)			
Confidence (errors)	2.2 (0.09)	2.2 (0.09)	2.2 (0.09)	2.1 (0.09)			
Time to solve problem (s)	36.5 (1.3)	37.6 (1.3)	36.9 (1.2)	37.1 (1.2)			
Retest ^b							
P (correct)	0.67 (0.02)	0.66 (0.02)	0.68 (0.02)	0.64 (0.02)			
P (corrected guesses excluded, tutor entered) ^c	0.60 (0.03)	0.30 (0.04)	0.60 (0.03)	0.29 (0.03)		**	
Tutor use (after errors only) ^a							
P (tutors entered) ^d	0.63 (0.07)	0.61 (0.07)	0.69 (0.06)	0.58 (0.06)			
P (clicks made tutor entry) ^e	0.72 (0.04)	0.72 (0.04)	0.83 (0.04)	0.67 (0.04)		*	*

Note: Mean (\pm s.e.m.), * $P < 0.05$, ** $P < 0.01$. ^aAnalyzed with 2 (ST vs NT) \times 2 (better learners vs poorer learners) ANCOVAs using math SAT as a covariate. ^bAnalyzed with 2 (ST vs NT) \times 2 (better learners vs poorer learners) ANCOVAs using math SAT and first-test accuracy as covariates. ^cProportion of items incorrect at first test and but corrected at retest with confidence greater than a guess, and where at least one click was made somewhere in the tutor environment before continuing to the next problem. This is the metric that was used to stratify subjects into better and poorer learners, and was used in the SEM. ^dProportion of all tutor opportunities following errors where subjects made at least one click somewhere in the tutor environment before continuing to the next problem. ^eProportion of combined 'steps' and 'mores' (i.e. clicks) out of all possible clicks for tutors where at least one click had been made (i.e. tutor entry). Note that only information from one 'more' can be present on the screen at any time, whereas all 'steps' can be present on the screen at the same time (i.e. a previous 'step' does not disappear from the screen when the next 'step' is clicked). As a result, the value for proportion of total clicks could be > 1 for a given item if a subject were to click on all 'steps' and 'mores', then return to one or more 'mores' that they visited earlier in the tutor experience.

groups, as were students' Math SAT and confidence in math ability. However, math identification demonstrated a significant interaction between ST and learning, $F(1, 64) = 3.9$, $P = 0.05$. Under NT, better learners reported marginally higher math identification than poorer learners ($P < 0.09$). In addition, although under ST, better and poorer learners did not differ from each other, the better learners under ST reported significantly lower math identification than better learners in the NT condition ($P < 0.04$). This latter result is generally consistent with past findings that higher math identification increases susceptibility to ST effects (e.g. Aronson *et al.*, 1999; Spencer *et al.*, 1999).

First-test and retest performance

Replicating the typical effect in the literature, ST suppressed females' overall performance on the initial test [Table 2; main effect of ST: $F(1, 63) = 4.3$, $P < 0.05$]. However, there was no effect of ST either on overall retest performance (Table 2; $P > 0.5$) or our measure of tutor-based error correction (i.e. proportion corrected at retest with greater than guessing level of confidence, following some use of the tutor; $P > 0.8$).¹

These results affirm that some females were able take advantage of tutorial opportunities and rebound from math failures even under threat (see also Good *et al.*, 2008). Indeed, within the ST group, some individuals achieved more successful outcomes than others, despite starting

from nearly identical levels of first-test impairment. Did the presence of ST create a specific emotional burden that some were able to overcome in order to successfully learn from the tutor? Was the emotional burden of the feedback more of a factor in determining rebound from error under ST than under NT? We addressed these questions by first examining the ERP response to performance feedback.

ERP responses to performance feedback

Given our interest in understanding how ST influences error processing, we focused our analyses on the difference waves created by subtracting positive from negative feedback (for raw waveforms see Supplementary Data). This subtraction highlights any differential salience associated with negative feedback by treating the positive feedback response as a baseline against which the negative feedback response is compared.

These difference waves clearly illustrate the negative-going FRN and positive-going P3a [t -tests against baseline: FRN_{diff} $t(78) = -2.79$, $P < 0.01$; P3a_{diff} $t(78) = 2.30$, $P < 0.05$; Figure 2A–C]. Yet, neither ST nor learning success were associated with significant modulations of these waveforms (main effects and interactions: all P 's > 0.3). Despite the appearance of an enhanced P3a_{diff} in poorer learners under ST, exploratory pair-wise comparisons between better and poorer learners as a function of threat condition confirmed that there were no significant differences as a function of learning success.

In contrast, learning success was directly related to the more sustained attention and emotional arousal processes as indexed by the LPP. An enhanced LPP_{diff} was found in poorer learners only [Figure 2D; main effect of learning:

¹ ST also did not affect the small likelihood of correcting items based either on a lucky guess (i.e. proportion corrected with guessing level of confidence; all F 's < 2 , P 's > 0.2 ; ST better learners: $M = 0.04$, ST poorer learners: $M = 0.05$; NT better learners: $M = 0.04$; NT poorer learners: $M = 0.07$) or on other, non-tutor based processes (i.e. proportion corrected with greater than guessing level of confidence without use of the tutor; all F 's < 1 , P 's > 0.3 ; ST better learners: $M = 0.19$, ST poorer learners: $M = 0.21$; NT better learners: $M = 0.14$; NT poorer learners: $M = 0.29$).

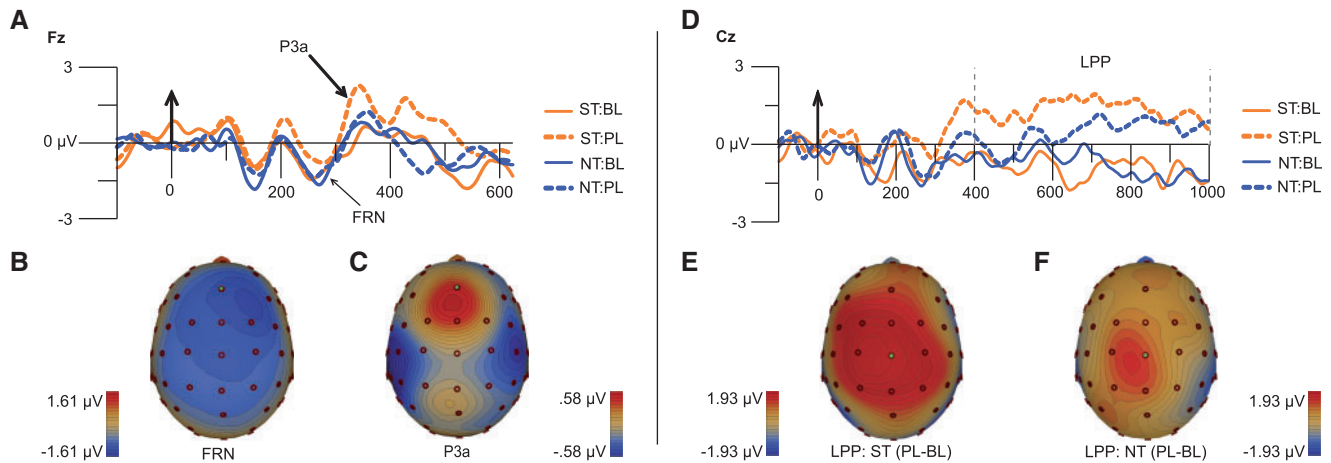


Fig. 2 Event-related potential responses to accuracy feedback as a function of stereotype threat (ST: stereotype threat; NT: non-threat) and learning success (BL: better learners; PL: poorer learners), highlighting the difference waves (negative feedback–positive feedback) for each group. (A) Grand mean waveforms for FRN and P3a difference waves at their Fz maximum. Positive voltages are plotted up. (B) Range-scaled voltage spline map of the scalp distribution of the FRN_{diff}, with the Fz electrode highlighted in green. (C) As in (B), but for the P3a_{diff}. (D) As in (A), but for the late positive potential (LPP), shown at its Cz maximum. Voltage spline maps of the LPP_{diff} for poorer learners under ST (E) and NT (F) conditions, with the Cz electrode highlighted in green.

$F(1, 61) = 9.0$, $P < 0.005$], supporting the general hypothesis that sustained emotional processing of negative feedback would interfere with learning success. Although no significant effects of ST nor interactions between threat and learning were found (F 's < 1 , P 's > 0.3), exploratory pair-wise comparisons indicated that ST was largely driving this effect ($P = 0.008$; Figure 2E), even though a similar trend could be found under NT ($P = 0.1$; Figure 2F).² This finding suggests that stereotypes may indeed selectively increase the emotional burden of negative feedback on learning.

Engagement with tutor following errors

Our analysis of tutor use took a hierarchical approach. First, we asked whether the frequency with which students made some attempt to investigate the tutor (i.e. by making at least one click in the tutor environment) differed as a function of threat or learning success. As shown in Table 2, there were no significant group differences in the basic measure of tutor entry. However, when we probed the depth of tutor engagement for tutors that were entered, we found that deeper exploration within the tutor led to greater overall success in solving isomorphic retest problems [main effect of learning success; $F(1, 63) = 4.10$, $P < 0.05$]. This effect was moderated by the presence of threat [interaction; $F(1, 63) = 4.45$, $P < 0.05$], in that the quantity of exploration was related to error correction only under NT ($P < 0.01$). Under ST, better and poorer learners appeared to make similar efforts to investigate the tutor, at a level that was marginally shallower than the better learners under NT ($P < 0.07$), but did not differ from the poorer learners under NT. This suggests

that poorer learners in the ST condition were retaining less information from their investigation of the tutor than were the better learners in this condition.

Structural equation modeling of retest success

To provide a more comprehensive picture of how emotion-cognition interactions influenced learning success and were influenced by ST, we conducted structural equation modeling (SEM) of the paths to retest success. Here, we focused on the relationships between feedback processing, tutor use and error correction, and as such, we did not include individual difference variables or post-test perceptions. Separate analyses were conducted on the ST and NT conditions in order to determine whether their paths to learning differed.

The main advantage of the SEM analysis was that we could independently test whether increased salience of the error was related to decreased error correction through a decrease in the quantity of tutor engagement (i.e. mediated model), as predicted by the disengagement hypothesis, or whether sustained attention to the thoughts and emotions elicited by this feedback directly interfered with learning success, as predicted by the interference hypothesis. Specifically, we predicted that disengagement would be evident in paths linking larger FRN_{diff} or P3a_{diff} effects to lower error correction that were significantly mediated through quantity of tutor use. In contrast, we predicted that interference from sustained processing would be reflected in a significant path linking a larger LPP_{diff} directly to decreased error correction, in the absence of a significant mediated path through quantity of tutor use. This pattern would indicate that distraction by sustained attention toward the negative feedback rendered the actual quantity of tutor engagement irrelevant.

² ANCOVAs of the 400–700 ms and 700–1000 ms segments yielded identical patterns of results to the 400–1000 ms epoch.

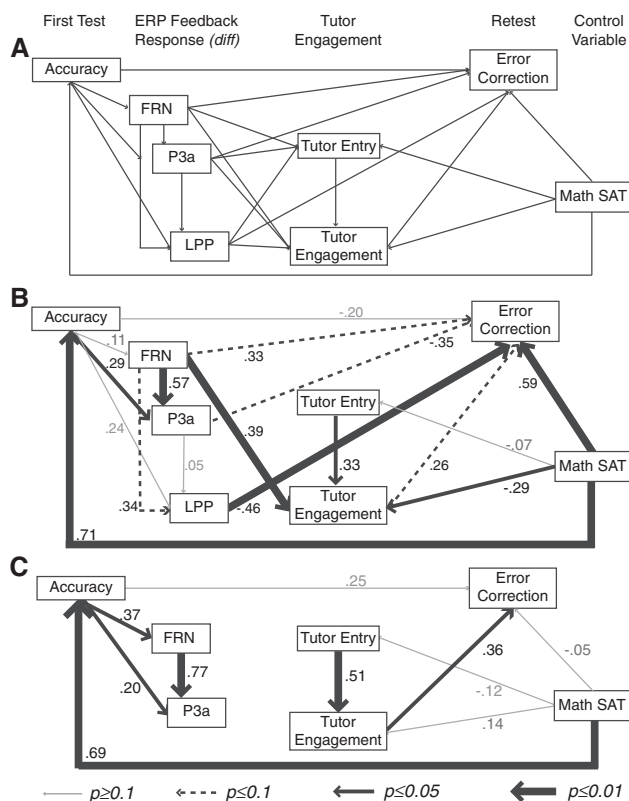


Fig. 3 Model testing. (A) Hypothesized model showing all paths tested in the more inclusive model (model 2; model 1 included all paths except those involving the FRN_{diff} and P3a_{diff}). All measures were defined in the same way as in the univariate analysis (see text). (B) Model 3, optimized for the ST condition, and showing all paths retained with standardized regression weights. Significance levels are illustrated graphically by line thickness. (C) As in (B), but for the NT condition.

With regard to choosing which ERP components to include in our model, Model 1 included only the LPP_{diff}, given that it was the only ERP component to demonstrate a significant relationship to learning in the univariate analysis. However, Model 2 (Figure 3A) additionally included the FRN_{diff} and P3a_{diff}, in order to determine if these components were related to learning success, but through quantity of tutor use, which could not have been assessed through the univariate approach. Both models included the control variables (first-test accuracy, Math SAT), and paths representing the hierarchical relationship between tutor entry and tutor engagement, given that engagement was measured only for the subset of tutors that had been entered. AMOS was used to conduct maximum likelihood parameter estimations on resulting variance-covariance matrices.

Multiple goodness-of-fit indices were used to determine whether Model 1 or 2 (Table 3) provided a better fit to the data. After finding that the more inclusive model (Model 2) was a better fit for the ST condition (RMSEA < 0.05), we further improved this model fit for each condition by eliminating any paths with $P > 0.1$. Given the exploratory nature of these models, we opted to retain marginally significant paths (Supplementary Table S1 for full details of Model 2

paths). The results of trimming are reflected in what we refer to as Model 3 for each condition (Figure 3B and C).

As can be seen by comparing the optimized models for ST (Figure 3B) and NT (Figure 3C), the response to negative feedback played a much greater role in determining error correction success under ST than under NT conditions. Under ST, the initial detection of the negative outcome, as indexed by the amplitude of the FRN_{diff}, significantly predicted how deeply females attempted to explore the tutor, with those who had a more negative-going FRN_{diff} withdrawing effort from the tutor earlier. Tutor engagement then marginally predicted error correction.³ Taken together, these results lend some support to the disengagement hypothesis, in that they revealed a mediated relationship between the FRN_{diff} and error correction that was not apparent in our univariate analyses. In contrast, the significant direct path from the LPP_{diff} to error correction under ST is more consistent with the interference hypothesis, and suggests that sustained emotional processing interfered with the quality of information extracted from the tutor.

These results suggest that ST can undermine learning through either reduced quality or quantity of tutor engagement, and furthermore, that these effects are linked to different components of the emotional response to negative feedback. Although we acknowledge that the FRN_{diff} and P3a_{diff} also demonstrated marginally significant direct paths to learning outcome under ST, it is also important to note that these paths accounted for much less variance than the direct path between the LPP_{diff} and error correction. Perhaps even stronger evidence that the FRN_{diff} and LPP_{diff} were related to different maladaptive learning strategies, however, is the finding that under ST, a more negative-going FRN_{diff} was associated with a smaller positive-going LPP_{diff}. Thus, ST did not simply increase all aspects of emotional processing of negative feedback uniformly. Rather, it was those individuals who demonstrated the largest initial response to the negative feedback who appear to have been the most motivated to disengage further attention from either this information, as indicated by their smaller LPP_{diff} or any other reminders of their poor performance on that problem (i.e. the tutor).

In contrast to the effects of the response to negative feedback on learning under ST, the only significant predictor of error correction under NT was the quantity of exploration within the tutor. This suggests that whatever parts of the tutor students in this condition opted to explore were encoded deeply, without the interference experienced by students in the ST condition. Furthermore, when these students opted to withdraw ‘earlier’ from the tutor it did not appear to be negatively related to their emotional response to the error, at least as measured by the ERPs of interest.

³ The relationship between quantity of tutor use and error correction may have been weakened somewhat by the concurrent presence of direct paths between the LPP_{diff} and error correction, which we argue relates to the quality of information extracted from the tutor.

Table 3 Model testing

	χ^2	DF	<i>P</i>	χ^2/DF	CFI	NFI	RMSEA	Pclose	AIC
Stereotype threat (<i>N</i> = 32)									
Model 1	5.08	4.00	0.28	1.27	0.97	0.89	0.09	0.32	51.08
Model 2	5.57	6.00	0.47	0.93	1.00	0.93	0.00	0.52	81.57
Model 3	7.10	11.00	0.79	0.65	1.00	0.91	0.00	0.83	73.10
Non-threat (<i>N</i> = 34)									
Model 1	3.10	4.00	0.54	0.78	1.00	0.94	0.00	0.58	49.10
Model 2	7.21	6.00	0.30	1.20	0.99	0.93	0.08	0.35	83.21
Model 3	9.61	11.00	0.57	0.87	1.00	0.90	0.00	0.63	57.61

Finally, we note that Math SAT was unrelated to learning success under NT conditions, but was the best predictor of outcome under ST. In the presence of ST, better Math SAT scores were also associated with less tutor use. One interpretation of these results is that greater pre-existing math ability necessitated less reliance on the tutor, and perhaps provided some protection against ST overall. However, given that the Math SAT cannot be viewed as an entirely context-free measure of ability (Walton and Spencer, 2009), its relationship to learning success may simply reflect similarity between the Math SAT and math testing under explicit ST conditions (Danaher and Crandall, 2008). Students who have adopted strategies enabling them to take tests well under ST could therefore demonstrate their ability on both the Math SAT and in the current study. Importantly, under NT, engagement with the tutor superseded the influence of prior standardized test performance in predicting the rebound from errors, even though math SAT was still a strong predictor of students' initial test performance.

DISCUSSION

In the presence of negative stereotypes about their math ability, females fulfilled this prophecy by underperforming on a math test compared to their unthreatened peers. Yet, an equivalent percentage of individuals in both ST and NT groups were able to demonstrate successful learning on a surprise retest a day later. Our main finding was that the paths to this successful learning differed as a function of ST, with threat increasing the dependence between females' emotional response to negative performance feedback and their error correction success. Specifically, females who demonstrated evidence of enhanced initial detection of negative outcomes (i.e. FRN_{diff}), perhaps stemming from greater overall vigilance for ability-impugning information, disengaged earliest from their exploration of subsequent learning opportunities, whereas those who demonstrated poorer regulation of their attention to and arousal from the negative feedback (i.e. LPP_{diff}) failed to receive significant learning benefits from any tutorial information they may have explored. Neutralizing threat not only liberated learning from dependence on these emotional responses, but also

from prior measures of ability (i.e. Math SAT), leaving learning success to be determined by other factors, such as intrinsic motivation to investigate new approaches to math problems, which may have been rooted in higher identification with the math domain.

These results are broadly consistent with recent fMRI findings that females' threat-induced performance deficits on math-relevant tasks are related to increased activity in neural regions associated with emotional conflict and its regulation, including the ventral anterior cingulate cortex (Wraga *et al.*, 2007; Krendl *et al.*, 2008). They also provide empirical support for theoretical models of ST that place emotional salience, and its consequent regulation, as mediators of the amount of the cognitive resources that remain available for performance on diagnostic tasks (Schmader *et al.*, 2008). Importantly, however, we demonstrate for the first time that these emotional responses can disrupt not only initial performance but also learning. These findings also lend support to general feedback theories that model negative feedback as harmful rather than helpful to learning if it directs attention away from the task and toward processing of the individual's affective experience (Kluger and DeNisi, 1996).

In addition, our findings begin to unpack the different mechanisms by which emotion might impact learning under ST. We found evidence for both disengagement and interference as disruptive mechanisms for learning, as well as evidence that these processes were associated with partially dissociable aspects of students' emotional response to negative feedback. The LPP, which has served as a sensitive index of up- or down-regulation of sustained emotional arousal in previous studies (Hajcak *et al.*, 2010), demonstrated the strongest relationship to interference with learning in the present study. The stronger relationship between interference and the LPP_{diff} , compared to the FRN_{diff} or $P3a_{diff}$, is highly consistent with a recent study finding that the LPP was better than the P3 at predicting the extent to which emotional stimuli interfered with processing of subsequent neutral targets (Weinberg and Hajcak, 2011). Our findings are also consistent with research showing that emotion regulation strategies that occur as a reaction to an emotionally-arousing event (i.e. rumination and

suppression), can impair working memory and long-term memory (Richards and Gross, 2000, 2006; Bonanno *et al.*, 2004). Moreover, these maladaptive strategies may be directly responsible for performance deficits under ST (Beilock *et al.*, 2007; Johns *et al.*, 2008).

More counterintuitive is the finding that the relatively fast, automatic processing of error salience indexed by the FRN predicted disengagement from learning. Although this finding fits well with vigilance-avoidance models of anxiety regulation (Mogg *et al.*, 2004), to the extent that the FRN indexes the magnitude of conflict between actual and desired task outcome, it is conceivable that it would have been positively related to the motivation to correct errors. Indeed, some studies with non-anxious individuals have found that the FRN predicts behavioral changes that result in improved performance, such as post-error slowing (Cavanagh *et al.*, 2010), or switching from incorrect to correct response patterns (Cohen and Ranganath, 2007). The error-related negativity (ERN), a related index of response-based conflict with a similar generator in anterior cingulate cortex (Potts *et al.*, 2010), also has demonstrated a positive relationship to self-regulatory processes including better everyday stress regulation (Compton *et al.*, 2008), and better academic grades (Hirsh and Inzlicht, 2010). To our knowledge, however, the FRN has only been found to predict beneficial behavioral adaptation in probabilistic learning tasks where the correct response is implicit in the error feedback and requires no knowledge-seeking effort. Indeed, past studies have often failed to find a significant relationship between the FRN and explicit seeking and learning of new information (Butterfield and Mangels, 2003; Mangels *et al.*, 2006; Chase *et al.*, 2010). In addition, we offer the suggestion that the ERN may be more predictive of beneficial self-regulation in various contexts because it is itself an index of the ability to internally monitor performance, whereas the FRN occurs in response to externally provided error signals.

For a female under ST, the amplitude of the FRN elicited by these external indicators of poor performance may be related to the degree of anxiety provoked by the stereotype-confirming message they send. Those stereotyped females who were most sensitive to this ability-impugning feedback may have then attempted to manage their anxiety by actively avoiding whatever aspects of the task they could—principally by limiting their exploration of the tutor. In contrast, for females under NT conditions, rapid registration of an error may not necessarily lead to an anxious response, and thus the amplitude of the FRN will have little influence on either their engagement with learning opportunities or their success of encoding this information into long-term memory (see also Mangels *et al.*, 2001; Butterfield and Mangels, 2003). However, we acknowledge that a disengagement strategy might have been less likely to occur overall if the learning information either had been passively presented, or students had known of the upcoming

retest. Nonetheless, a previous study found that even when they knew of an upcoming competition, athletes under ST sometimes actively avoided a learning opportunity (i.e. practice)—a self-handicapping strategy that they thought would help them cope with performance-related anxiety (Stone, 2002). By this view, both the FRN and LPP served as indices of maladaptive emotion regulation strategies, but the FRN was associated with avoidance of learning-relevant information (i.e. the tutor) that also put students at risk of reinforcing the threat of low ability, whereas the LPP was associated with excessive attention to the threat-relevant information itself (i.e. the negative feedback).

Understanding the emotion-cognition interactions responsible for compromised learning under ST can provide a starting point for interventions aimed at breaking the cycle of poor performance and poor learning that may fuel further stereotyping. In particular, our results suggest that individuals under ST could benefit from strategies aimed at helping individuals reappraise the arousal resulting from negative feedback as arising from something other than a lack of ability. In general, strategies that minimize the emotional response by preemptively altering the meaning of the stimulus and/or how attention is deployed to that stimulus are often the most effective in reducing interference from emotionally arousing events (Gross, 1998, 2002; Ochsner and Gross, 2005). Although these types of antecedent-focused strategies require some cognitive resources, they do not require the continuous self-monitoring and self-regulation associated with response-based strategies such as suppression, and therefore are not generally associated with cognitive costs (Richards and Gross, 2000; Dillon *et al.*, 2007). Correspondingly, recent studies have shown that task framing aimed at encouraging individuals to cognitively reappraise the source of their anxiety or arousal as stemming from something other than poor ability have been successful in reducing stereotype-threat induced performance decrements (Ben-Zeev *et al.*, 2005; Johns *et al.*, 2005, 2008). Indeed, it is possible that those individuals in the present study who had smaller LPPs were those who spontaneously used these types of antecedent-focused cognitive reappraisal strategies, allowing them to more rapidly redirect attention toward learning corrective information following negative feedback.

Finally, although it was beyond the scope of the present study to conduct a full investigation all the individual differences that might explain the variability we observed in stigmatized females' feedback responses, our data suggest that stronger identification with math led to poorer learning under ST (see also Aronson *et al.*, 1999; Spencer *et al.*, 1999), but fostered learning when threat was minimized. Under ST, greater identification would have likely increased the salience of signals of poor ability, and thus, the potential for such signals to disrupt subsequent learning. In contrast, under NT conditions, negative feedback may only signal failure to correctly solve a particular problem, rather than broader deficits

in ability, allowing investment in the domain to positively motivate an approach toward the solution to that problem.

The relative costs and benefits of math identification as a function of ST provide additional evidence for the importance of destigmatizing classroom environments, but leave open the question of how to mitigate the vulnerability of math-identified females in contexts where ST may be salient and unavoidable. However, if we consider that ST may temporarily induce a performance-focused mindset (Aronson et al., 2002; Good et al., 2003; Dar-Nimrod and Heine, 2006; Blackwell et al., 2007), then interventions that reinforce the belief that intelligence can be enhanced through learning-oriented effort may be effective. Indeed, this type of mindset may function as a type of antecedent cognitive reappraisal strategy that could both minimize the initial sting of negative outcomes and keep attention focused outwardly toward learning-relevant information, rather than inwardly toward one's affective response (see also Good et al., 2003; Mangels et al., 2006). Insofar as our present results demonstrate the powerful influence that students' emotional responses to negative feedback have on rebound from failure under ST, future research will be necessary to more directly address the strategies and mindsets that might promote effective cognitive reappraisal of arousal and the consequent benefits to learning in this context.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

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, 113–25.
- Aronson, J., Lustina, M.J., Good, C., Keough, K., Steele, C.M., Brown, J. (1999). When white men can't do math: necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35, 29–46.
- Beilock, S.L., Rydell, R.J., McConnell, A.R. (2007). Stereotype threat and working memory: mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, 136(2), 256–76.
- Ben-Zeev, T., Fein, S., Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, 41, 174–81.
- Blackwell, L.S., Trzesniewski, K.H., Dweck, C.S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: a longitudinal study and an intervention. *Child Development*, 78(1), 246–63.
- Bonanno, G.A., Papa, A., Lalande, K., Westphal, M., Coifman, K. (2004). The importance of being flexible: the ability to both enhance and suppress emotional expression predicts long-term adjustment. *Psychological Science*, 15(7), 482–7.
- Buckner, J.D., Dewall, C.N., Schmidt, N.B., Maner, J.K. (2010). A tale of two threats: social anxiety and attention to social threat as a function of social exclusion and non-exclusion threats. *Cognitive Therapy and Research*, 34(5), 449–55.
- Butterfield, B., Mangels, J.A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Brain Research Cognitive Brain Research*, 17(3), 793–817.
- Cavanagh, J.F., Frank, M.J., Klein, T.J., Allen, J.J. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, 49(4), 3198–209.
- Ceci, S.J., Williams, W.M., Barnett, S.M. (2009). Women's underrepresentation in science: sociocultural and biological considerations. *Psychological Bulletin*, 135(2), 218–61.
- Chase, H.W., Swainson, R., Durham, L., Benham, L., Cools, R. (2010). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *Journal of Cognitive Neuroscience*, 23, 936–46.
- Codispoti, M., Ferrari, V., De Cesarei, A., Cardinale, R. (2006). Implicit and explicit categorization of natural scenes. *Progress in Brain Research*, 156, 53–65.
- Cohen, G., Steele, C., Ross, L. (1999). The Mentor's dilemma: providing critical feedback across the racial divide. *Personality and Social Psychology Bulletin*, 25(10), 1302–18.
- Cohen, M.X., Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience*, 27(2), 371–8.
- Compton, R.J., Robinson, M.D., Ode, S., Quandt, L.C., Fineman, S.L., Carp, J. (2008). Error-monitoring ability predicts daily stress regulation. *Psychological Science*, 19(7), 702–8.
- Danaher, K., Crandall, C.S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology*, 38(6), 1635–55.
- Dar-Nimrod, I., Heine, S.J. (2006). Exposure to scientific theories affects women's math performance. *Science*, 314(5798), 435.
- Dillon, D.G., Ritchey, M., Johnson, B.D., LaBar, K.S. (2007). Dissociable effects of conscious emotion regulation strategies on explicit and implicit memory. *Emotion*, 7(2), 354–65.
- 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(3), 253–61.
- Friedman, D., Cycowicz, Y.M., Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews*, 25(4), 355–73.
- 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.
- Good, C., Aronson, J., Inzlicht, M. (2003). Improving adolescents' standardized test performance: an intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24, 645–62.
- Gross, J.J. (1998). Antecedent- and response-focused emotion regulation: divergent consequences for experience, expression and physiology. *Journal of Personality and Social Psychology*, 74, 224–37.
- Gross, J.J. (2002). Emotion regulation: affective, cognitive, and social consequences. *Psychophysiology*, 39(3), 281–91.
- Hajcak, G., MacNamara, A., Olvet, D.M. (2010). Event-related potentials, emotion, and emotion regulation: an integrative review. *Developmental Neuropsychology*, 35(2), 129–155.
- Hirsh, J.B., Inzlicht, M. (2010). Error-related negativity predicts academic performance. *Psychophysiology*, 47(1), 192–6.
- Holroyd, C.B., Coles, M.G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709.
- Ito, T.A., Thompson, E., Cacioppo, J.T. (2004). Tracking the timecourse of social perception: the effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin*, 30(10), 1267–80.
- Iwanaga, M., Nittono, H. (2010). Unexpected action effects elicit deviance-related brain potentials and cause behavioral delay. *Psychophysiology*, 47(2), 281–8.
- 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(4), 691–705.
- 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(3), 175–9.
- Kluger, A.N., DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–84.

- Krendl, A.C., Richeson, J.A., Kelley, W.M., Heatherton, T.F. (2008). The negative consequences of threat: a functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math. *Psychological Science*, 19(2), 168–75.
- Mangels, J.A., Butterfield, B., Lamb, J., Good, C., Dweck, C.S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social cognitive and affective neuroscience*, 1(2), 75–86.
- Mangels, J.A., Picton, T.W., Craik, F.I. (2001). Attention and successful episodic encoding: an event-related potential study. *Brain Research Cognitive Brain Research*, 11(1), 77–95.
- Mogg, K., Bradley, B., Miles, F., Dixon, R. (2004). Time course of the attentional bias for threat scenes: Testing the vigilance-avoidance hypothesis. *Cognition and Emotion*, 18, 689–700.
- Ochsner, K.N., Gross, J.J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242–9.
- Olvet, D.M., Hajcak, G. (2009). The stability of error-related brain activity with increasing trials. *Psychophysiology*, 46(5), 957–61.
- Paller, K.A., Wagner, A.D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6(2), 93–102.
- Potts, G.F., Martin, L.E., Kamp, S.M., Donchin, E. (2010). Neural response to action and reward prediction errors: comparing the error-related negativity to behavioral errors and the feedback-related negativity to reward prediction violations. *Psychophysiology*, in press; doi:10.1111/j.1469-8986.2010.01049.x.
- Richards, J.M., Gross, J.J. (2000). Emotion regulation and memory: the cognitive costs of keeping one's cool. *Journal of Personality and Social Psychology*, 79(3), 410–24.
- Richards, J.M., Gross, J.J. (2006). Personality and emotional memory: how regulating emotion impairs memory for emotional events. *Journal of Research in Personality*, 40, 631–51.
- Rydell, R.J., Rydell, M.T., Boucher, K.L. (2010a). The effect of negative performance stereotypes on learning. *Journal of Personality and Social Psychology*, 99(6), 883–96.
- Rydell, R.J., Shiffrin, R.M., Boucher, K.L., Van Loo, K., Rydell, M.T. (2010b). Stereotype threat prevents perceptual learning. *Proceedings of the National Academy of Sciences USA*, 107(32), 14042–7.
- Schmader, T., Johns, M., Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115(2), 336–56.
- Shih, M., Pittinski, T.L., Amabady, N. (1999). Stereotype susceptibility: shifts in quantitative performance from sociocultural identification. *Psychological Science*, 10, 81–4.
- Simons, R.F. (2010). The way of our errors: theme and variations. *Psychophysiology*, 47, 1–14.
- Spencer, S., Steele, C.M., Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28.
- Steele, C.M. (1997). A threat in the air: how stereotypes shape intellectual identify and performance. *American Psychologist*, 52(6), 613–29.
- 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, 1667–78.
- Van Meel, C.S., Van Heijningen, C.A. (2010). The effect of interpersonal competition on monitoring internal and external error feedback. *Psychophysiology*, 47(2), 213–22.
- Walton, G.M., Spencer, S.J. (2009). Latent ability: grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, 20(9), 1132–9.
- Weinberg, A., Hajcak, G. (2011). The late positive potential predicts subsequent interference with target processing. *Journal of Cognitive Neuroscience*, 23(10), 2994–3007.
- Wraga, M., Helt, M., Jacobs, E., Sullivan, K. (2007). Neural basis of stereotype-induced shifts in women's mental rotation performance. *Social Cognitive and Affective Neuroscience*, 2(1), 12–9.