

Stereotype-based stressors facilitate emotional memory neural network connectivity and encoding of negative information to degrade math self-perceptions among women

Chad E. Forbes, Rachel Amey, Adam B. Magerman, Kelly Duran, and Mengting Liu

Social Neuroscience Laboratory, Department of Psychological and Brain Sciences, University of Delaware, Newark, DE 19716, USA

Correspondence should be addressed to Chad E. Forbes, Department of Psychological and Brain Sciences, University of Delaware, 111 Wolf Hall, Newark, DE 19716, USA. E-mail: cforbes@psych.udel.edu.

Abstract

Stress engendered by stereotype threatening situations may facilitate encoding of negative, stereotype confirming feedback received during a performance among women in science, technology, engineering and mathematics (STEM). It is unclear, however, whether this process is comprised of the same neurophysiological mechanisms evident in any emotional memory encoding context, or if this encoding bias directly undermines positive self-perceptions in the stigmatized domain. A total of 160 men and women completed a math test that provided veridical positive and negative feedback, a memory test for feedback, and math self-enhancing and valuing measures in a stereotype threatening or neutral context while continuous electroencephalography activity and startle probe responses to positive and negative feedback was recorded. Indexing amygdala activity to feedback via startle responses and emotional memory network connectivity elicited during accurate recognition of positive and negative feedback via graph analyses, only stereotype threatened women encoded negative feedback better when they exhibited increased amygdala activity and emotional memory network connectivity in response to said feedback. Emotional memory biases, in turn, predicted decreases in women's self-enhancing, math valuing and performance. Findings provide an emotional memory encoding-based mechanism for well-established findings indicating that women have more negative math self-perceptions compared with men regardless of actual performance.

Key words: stress; emotional memory encoding; STEM attitudes; stereotype threat; women; social neuroscience; network neuroscience; graph theory; modularity

Introduction

When individuals find themselves in evaluative situations that prime their membership in a negatively stereotyped group they are likely to exhibit physiological and psychological markers of stress (Schmader *et al.*, 2008). Past research suggests that these

situations, termed stereotype threatening situations (Steele and Aronson, 1995), facilitate attention toward (Forbes *et al.*, 2008; Forbes and Leitner, 2014) and encoding of negative, stereotype confirming feedback received during the performance (Forbes *et al.*, 2015). These biases in encoding may also undermine

Received: 1 June 2017; Revised: 26 May 2018; Accepted: 18 June 2018

© The Author(s) (2018). Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

performance on future math tests and promote more general negative affective experiences. It is unclear, however, whether this process is comprised of the same neurophysiological mechanisms evident in any emotional memory encoding context, or if this encoding bias directly undermines positive self-perceptions that are otherwise critical for remaining engaged with the stigmatized domain.

The current study provides a direct link between neurophysiological emotional memory encoding processes, memory bias for negative feedback, performance and self-enhancing in the math domain among stereotype threatened women. We find evidence that while all individuals encode negative feedback more accurately, this occurs through different neural mechanisms for men compared with stereotype threatened women. Encoding feedback via the emotional route, in turn, prompts decreased self-enhancing in the math domain and math valuing among stereotype threatened women only.

Threatening STEM contexts prompt emotional memory encoding among women

When women find themselves in science, technology, engineering and mathematic (STEM) situations that prime the stereotype that they are inferior to men in math, they are likely to experience various identity threats including stereotype and social identity threat (Murphy *et al.*, 2007; Schmader *et al.*, 2008). These situations can arise in various ways, including having women indicate their gender prior to taking a diagnostic math test (DMT) such as the SAT or GRE (Spencer *et al.*, 1999), being outnumbered by men in a testing situation (Inzlicht and Ben-Zeev, 2000), or simply having a male instructor in their math class (Marx and Goff, 2005). Once primed, stereotype threat engenders a cascade of physiological, performance monitoring and appraisal processes that tax working memory resources necessary for optimal performance on difficult tasks (Schmader *et al.*, 2008).

Importantly identity threatening situations typically engender arousal and negative affective biomarkers indicative of stress, including increased cortisol, blood pressure, skin conductance and α amylase (Blascovich *et al.*, 2001; Osborne, 2006; Osborne, 2007; Schmader *et al.*, 2009; Townsend *et al.*, 2011), as well as explicit and implicit anxiety, math related worrying/concern and feelings of dejection (Osborne, 2001; Keller and Dauenheimer, 2003; Bosson *et al.*, 2004; Cadinu *et al.*, 2005; Delgado and Prieto, 2008; Forbes *et al.*, 2014; Johns *et al.*, 2008). We refer to this identity threat-based stress response as 'stereotype-based stress (SBS)'.

To the extent stereotype threatening contexts engender SBS, they provide all the necessary ingredients for a well-documented effect in the cognitive neuroscience literature referred to as emotional memory encoding. A large body of literature indicates that negative, emotionally arousing information receives privileged attention and as such is better encoded (Hamann, 2001), consolidated (LaBar and Phelps, 1998) and retrieved (Ochsner, 2000) in negatively arousing, stressful contexts (e.g. Levine and Burgess, 1997; Payne *et al.*, 2006; for a review see LaBar and Cabeza, 2006). The effects of emotional memory encoding are enduring as well (Hamann *et al.*, 1999; Canli *et al.*, 2000).

Better encoding, consolidating, and retrieving of events is largely due to arousal-based activation of the amygdala, the regulator of the stress response and hub for emotional processing in the brain, mediating the interaction between regions in the medial temporal lobe, such as the hippocampus, and areas

in the cortex, including dorsal and ventral aspects of anterior cingulate and prefrontal cortex, bilateral insular cortex and bilateral middle frontal gyri (Zald *et al.*, 1998; Bush *et al.*, 2000; LaBar and Cabeza, 2006; Murty *et al.*, 2010; Kim, 2011). Importantly, connectivity between these various regions may provide insight into the degree to which emotion, compared with more semantic, memory processes are involved in encoding of specific information. For instance, Steinmetz *et al.* (2010) presented individuals with positive and negatively valenced emotional pictures that varied in degree of arousal and then administered a surprise memory test for these pictures post-task. Effective connectivity analyses revealed that accurately identified negatively arousing pictures were associated with increased connectivity between the amygdala, inferior/middle frontal gyrus and occipital gyrus as well as other nodes of the emotional memory network; positively arousing pictures decreased connectivity between these regions. In the present study we utilized different 'modularity' analyses, or analyses that gauge the extent to which a set of nodes or brain regions in a functionally relevant network are interconnected and effectively communicate with one another compared with regions outside the network (Sporns, 2011), to index how connectivity between regions in the emotion and semantic memory networks modulates behavioral memory measures of interest.

Thus, greater connectivity or modularity between regions integral for emotional memory encoding exhibited in response to correctly identified negatively arousing stimuli could serve as evidence of the importance of emotion in the encoding process. With respect to stereotype threatened women, to the extent information associated with negative, stereotype confirming evidence, e.g. performance feedback, evokes a negative, viscerally arousing emotional reaction (evidenced by amygdala activity elicited in response to negative feedback), this information should be encoded better than positive, stereotype disconfirming information. Better encoding of negative information should, in turn, be a product of interactions between the amygdala and regions integral for emotion and memory encoding processes (evidenced by greater connectivity between emotion and memory regions on accurately identified stimuli during a memory test).

Emotional memory encoding processes have downstream consequences on women's performance and self-perceptions in STEM domains

Past research provides direct evidence that stereotype threatened women attend to and encode information associated with negative, stereotype confirming feedback better than positive, stereotype disconfirming feedback and non-threatened men and women (Forbes and Leitner, 2014; Forbes *et al.*, 2015). Furthermore, these attentional and encoding biases have negative downstream consequences on women's performance and self-perceptions in STEM domains. This suggests that the attentional and encoding bias exhibited toward negative feedback may be cognitively taxing in and of itself, depleting resources otherwise necessary for optimal performance on more cognitively demanding tasks, which in turn has negative ramifications for how women explicitly feel post-performance.

It is possible that negative memory biases may undermine tactics that women normally employ to maintain engagement with STEM domains as well, i.e. maintain a positive self-concept in the domain. Among other things, the vividness of past negative experiences may interfere with women's ability to reconstruct autobiographical memories in a manner that

embellishes current aspects of the self in STEM domains, and thus may affect the extent to which they view the domain as relevant to the self. Based on past findings among non-stigmatized students, autobiographical memory reconstruction processes maintain a positive academic self-concept via a three step interaction between past memories and one's domain-specific self-concept. Domain identified individuals tend to (i) engage in self-enhancing strategies that positively bias past academic memories, (ii) utilize the now more positive academic memories to positively influence their current domain self-concept and (iii) utilize the now more positive academic self-concept to bias past academic memories even more so to provide the individual with the belief that they have had primarily positive domain experiences (e.g. Wilson and Ross, 2003; Gramzow and Willard, 2006).

So what happens then when women have a disproportionate representation of vivid negative STEM memories? It is possible these memories undermine women's ability to successfully engage in the autobiographical memory reconstruction processes described above. If true, we would expect a link between emotional memory processes, biased encoding of negative feedback and stereotype threatened women's tendency to self-enhance in the math domain, i.e. perceive themselves as better as and more capable than their STEM peers. Given the critical importance of self-enhancing in domain engagement, i.e. maintaining the perception that one is capable, competent and has the ability to persevere through failures within a valued domain, we would also expect women's valuing and engagement of the domain to vary as a function of their ability to self-enhance in STEM domains; decreases in math self-enhancement should be associated with decreases in math valuing and engagement among stereotype threatened women to the extent they exhibit biased encoding of negative feedback.

Study overview

The current study placed women and men in contexts where they completed supposed DMTs (a stereotype threatening context for women) or problem solving tasks (a stereotype neutral context) while continuous electroencephalography (EEG) activity was recorded. Similar to Forbes *et al.* (2015), an initial math task provided participants with veridical positive and negative feedback yoked to unique fonts. To measure SBS and emotional reactions to feedback, amygdala activity was assessed by startle probes administered on random trials while participants viewed positive or negative feedback. Participants then completed a standard math test, a surprise memory test for fonts yoked to positive and negative feedback (and lures), and self-enhancing and math valuing questionnaires.

We hypothesized that stereotype threatened women would encode fonts associated with negative feedback better than positive feedback, particularly to the extent said feedback elicited emotional responses (amygdala activity in response to feedback) and facilitated encoding of feedback through more emotional memory encoding processes (connectivity within the emotional memory network exhibited during accurate identification of feedback on the memory test). This bias in encoding coupled with negative emotional reactions should in turn have downstream consequences on performance and perceptions in the math domain. Stereotype threatened women should underperform on the standard math test, perceive themselves as less competent in the math domain compared with their peers and value math less to the extent they exhibit biases in memory

encoding stemming from emotional memory encoding processes.

While we expect stereotype threatened women to exhibit more efficacious encoding of information associated with stereotype confirming evidence, it is also possible that individuals encode negative feedback more accurately in general. Even at very early stages of development individuals attend to, learn from, and use negative information far more than positive information (Vaish *et al.*, 2008; see also research on error related negativity). There are, however, different ways in which individuals can learn and encode information. While stereotype threatened women should encode negative information via neural regions integral for emotional processes, we hypothesized that given that diagnostic math or stereotype neutral contexts should be largely devoid of emotion for men and women, they may encode negative feedback through a different neural mechanism and via semantic memory processes specifically (bilateral fusiform gyrus, bilateral superior occipital gyrus, bilateral prefrontal cortex, bilateral superior temporal gyrus, bilateral inferior parietal gyrus, bilateral posterior cingulate cortex (PCC)/Precuneus; Martin and Chao, 2001; Patterson *et al.*, 2007; Binder *et al.*, 2009; Binder and Desai, 2011; Flegal *et al.*, 2014). Furthermore, given that negative information received in these contexts should not serve as confirmation of any given male stereotype, we further posited that any encoding bias toward negative feedback exhibited by men should not engender any downstream negative consequences like decreased self-enhancing, performance or valuing of the math domain and may be beneficial for them (e.g. associated with better performance).

Materials and methods

Participants

One hundred and sixty white participants (87 women) completed the study for payment. All participants were aware of the negative female math stereotype. Specifically, participants were recruited to participate in the study if they responded with a 3 or lower to the following question during a pre-study screening: '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).

Procedure

Participants were seated in front of a computer screen in a soundproofed chamber and were prepared for electroencephalographic recording. Participants were randomly assigned to either a stereotype threat/DMT condition or a control/problem-solving (PST) condition. In the DMT condition, participants were told that they would be completing tasks that were diagnostic of their math intelligence. In the PST condition, participants were told that they would be completing tasks that were diagnostic of different types of problem-solving techniques they prefer (Forbes and Leitner, 2014; Forbes *et al.*, 2015). To prime stereotype threat, participants completed demographic questions including a gender query in the DMT condition, all DMT sessions included at least one male experimenter and participants had pre-recorded instructions read aloud to them by a male experimenter through headphones. In contrast, participants in the PST condition completed demographic questions that excluded the gender query, sessions always contained all female experimenters and instructions were read aloud to them

by a female experimenter. After the instructions, participants completed a math feedback task for 34 min (described below) and a traditional performance measure consisting of difficult GRE math problems for 5 min. Next, participants were presented with a surprise memory test to assess the extent to which they encoded fonts associated with negative or positive feedback during the math feedback task (described below). Participants then answered a series of questions, were debriefed and compensated for their participation.

Math feedback task

Participants completed a 34-min math task consisting of standard multiplication and division problems (e.g. $10 \times 20 =$) that initial pilot tests confirmed varied in degree of difficulty (easy, medium and hard, ensuring all participants would solve problems correctly and incorrectly). During each trial, participants were given three answer choices below each problem (A, B or C), with the answer to each problem randomly presented in one of the three answer positions on each trial. Participants made all answer selections via a button box placed in their laps and did not have scratch paper. After each response participants received feedback for 2 s that indicated whether their answer was wrong or correct. To assess memory for feedback, the words 'Wrong' or 'Correct' were presented in a novel font on every trial (see Forbes *et al.*, 2015). Participants were given 16 s to solve each problem. If participants were unable to answer a problem within that time frame they would receive negative feedback. Participants completed an average of 83.9 problems. Math score accuracy was calculated by dividing the total number of correct responses by the total number of attempted problems.

Standard math task

The standard math task consisted of 15 difficult math word problems taken from the GRE (Forbes and Schmader, 2010; Forbes *et al.*, 2015). Participants received scratch paper and were given 5 min to solve as many problems as possible. Participants received no feedback during this task. Accuracy scores were created by dividing the number of problems answered correctly by the number of problems attempted and multiplying that outcome by 100.

Memory test

Similar to Forbes *et al.* (2015), after the standard math task, participants were presented with a surprise memory test containing 400 trials. Among the 400 trials, participants were presented with each font/feedback pairing they had previously seen during the math feedback task, with the remaining trials acting as 'lures'. During each trial participants were randomly presented with the words 'wrong' or 'correct' written in one of the 200 different fonts in the middle of a computer screen. A scale was presented below each font/feedback combination and participants were asked to indicate whether they had seen the combination during the math feedback task using a six-point scale (1 = I know I didn't see it, 4 = I think I saw it, 6 = I know I saw it). If participants were presented with a previously seen font, responses of four to six were classified as hits, and responses of one to three were classified as misses. If participants were presented with a novel font, responses of four to six were classified as false alarms, and responses of one to three were classified as correct rejections. Using these classifications we calculated d' to measure participants' ability to accurately discriminate seen

from unseen fonts. Prior research suggests that d' is a more sensitive assessment of memory effects that accounts for guessing (Wickens, 2002). To calculate d' scores, z scores for false alarm rates were subtracted from z scores for hit rates. Because z scores for 0 or 1 cannot be calculated, participants without hits were given scores of 0.1 and participants with perfect scores were given a score of 0.9. Therefore, larger d' values indicate that participants were better at discriminating between previously seen fonts and lures.

SAQ-S

Participants reported math self-assessments on a modified shortened version of Pelham and Swann's (1989) Self-Attributes Questionnaire (SAQ) ($\alpha = 0.68$). Our short version of the SAQ measured participant's perceptions of their own math ability compared with other University students, e.g. 'Rate your math ability to other college students your own age', on a 1 to 10 scale (1 indicating they fell in the bottom 5% of the population and 10 indicating that they fell within the top 5% of the population). This measure was used as a proxy for math self-enhancement, as it has in past research (Harrington and Liu, 2002), where larger numbers equate to greater math self-enhancement (perceiving one's self to be better at math compared with their peers).

Math devaluing

Participants completed five questions modified from Major and Schmader (1998) that measured the extent to which individuals devalued the math domain, e.g. 'It usually doesn't matter to me one way or the other how I do in math classes'. ($\alpha = 0.79$).

EEG recording

Continuous EEG activity was recorded using an ActiveTwo head cap and the ActiveTwo Biosemi system (BioSemi, Amsterdam, The Netherlands). Recordings were collected from 128 Ag-AgCl scalp electrodes and from bilateral mastoids. Two electrodes were placed next to each other 1 cm below the right eye to record startle eye-blink responses. A ground electrode was established by BioSemi's common Mode Sense active electrode and Driven Right Leg passive electrode. EEG activity was digitized with ActiView software (BioSemi) and sampled at 2048 Hz. Data were downsampled post-acquisition and analyzed at 512 Hz.

Startle acquisition

To elicit startle responses, a 40 ms burst of white noise (100 db) was presented during the math feedback task through headphones (Bose Quietcomfort 25). During each startle trial, the startle probe was elicited one second into feedback presentation. Startle probes were presented randomly throughout the math feedback task. The adaptive math task included 256 possible problems that participants could have seen. There were 425 different font feedback pairs (169 positive font-feedback pairs, 256 negative font-feedback pairs) that were randomly presented to indicate whether participants got a problem correct or wrong. Out of these feedback pairs there were 117 possible startle probes (66 for negative feedback, 51 for positive feedback); participants were exposed to approximately one startle probe every three problems. Thus participants heard an average of 31 startle probes throughout the task ($M_{\text{TotalProbe}} = 31.01$, $SD_{\text{TotalProbe}} = 2.41$; $M_{\text{NegativeProbe}} = 14.34$, $SD_{\text{NegativeProbe}} = 3.59$; $M_{\text{PositiveProbe}} = 16.67$, $SD_{\text{PositiveProbe}} = 4.03$). Startle responses were

obtained from electromyographic (EMG) recordings of the right orbicularis oculi muscle using two Biosemi FLAT electrodes (BioSemi). As with the EEG data, startle activity was digitized with ActiView software (Biosemi) and sampled at 2048 Hz. Data were downsampled post-acquisition and analyzed at 512 Hz.

EEG preprocessing

For feedback analyses, EEG signal was epoched and stimulus locked from 500 ms pre-feedback presentation to 2000 ms post-feedback presentation. For memory test analyses, EEG signal was epoched and stimulus locked from 500 ms pre-feedback presentation (previously seen font/feedback combinations or lures) to 1000 ms post-feedback presentation. EEG artifacts were removed via FASTER (Fully Automated Statistical Thresholding for EEG artifact Rejection; Nolan et al., 2010), an automated approach to cleaning EEG data that are based on multiple iterations of independent component and statistical thresholding analyses. Specifically, raw EEG data was initially filtered through a band-pass FIR filter between 0.3 and 55 Hz. Then EEG channels with significant unusual variance (absolute z scores larger than 3 s.d. from the average), mean correlations with other channels and Hurst exponents were removed and interpolated from neighboring electrodes using a spherical spline interpolation function. EEG signals were then epoched and baseline corrected; epochs with significant unusual amplitude range, variance and channel deviation were removed. The remaining epochs were then transformed through ICA. Independent components with significant unusual correlations with EOG channels, spatial kurtosis, slope in the filter band, Hurst exponent and median gradient were subtracted and the EEG signal was reconstructed using the remaining independent components. In the last step, EEG channels within single epochs with significant unusual variance, median gradient, amplitude range and channel deviation were removed and interpolated from neighboring electrodes within the same epochs.

Startle extraction

Eyeblink EMG data were filtered by a 28-Hz high-pass FIR filter and then rectified. Peaks appearing between 20 and 150 ms after administration of the acoustic startle probe were operationalized as startle responses (Blumenthal et al., 2005). To better detect the peak value of rectified EOG from background noise, we utilized a non-linear approach of ensemble empirical mode decomposition (EEMD; Wu and Huang, 2009; Wu et al., 2009), which enhances sensitivity to local peak values of the non-stationary and non-linear signals, as the low-pass filter. EEMD deconstructed the rectified EMG into a collection of intrinsic mode functions (IMFs), and high frequency noise in the EMG data was reduced through the elimination of functions IMF1–IMF3. To ensure the detected peak value was the actual startle response instead of background signal that may have peaked within the given time period, we established that identified peak values must be larger than the average of all local peak values across the entire temporal period of –500 to 1000 ms epoch. Startle probe elicited EMG were scored as non-response trials if the peak value was less than the average of all local peak values across the entire temporal period. Nonresponse trials were excluded from our analyses. In this study, participants elicited startle responses above threshold on more than one-third of the startle probe trials ($M_{\text{TotalStartle}} = 11.66$, $SD_{\text{TotalStartle}} = 6.27$; $M_{\text{NegativeStartle}} = 6.03$, $SD_{\text{NegativeStartle}} = 3.71$; $M_{\text{PositiveStartle}} = 5.63$, $SD_{\text{PositiveStartle}} = 3.41$).

Source reconstruction

All a priori sources used in network connectivity analyses were identified and calculated via forward and inverse models utilized by MNE-python (Gramfort et al., 2013, 2014). The forward model solutions for all source locations located on the cortical sheet were computed using a three-layers boundary element model (Hämäläinen and Sarvas, 1989) constrained by the default average template of anatomical MNI MRI. Cortical surfaces extracted with FreeSurfer were sub-sampled to approximately 10 240 equally spaced vertices on each hemisphere. The noise covariance matrix for each individual was estimated from the pre-stimulus EEG recordings after preprocessing. The forward solution, noise covariance and source covariance matrices were used to calculate the dynamic statistical parametric mapping (dSPM) estimated inverse operator (Dale et al., 1999, 2000). The inverse computation was done using a loose orientation constraint (loose = 0.2, depth = 0.8) (Lin et al., 2006). Using depth weighting and noise normalization approaches, dSPM inverse operators have been reported to help characterize distortions in cortical and subcortical regions, and improve the bias accuracy of neural generators in deeper structures, e.g. the insula (Attal and Schwartz, 2013). The cortical surface was divided into 68 anatomical regions (i.e. sources) of interest (ROIs; 34 in each hemisphere) based on the Desikan–Killiany atlas (Desikan et al., 2006) and signal within a seed voxel of each region was used to calculate the power within sources and phase locking (connectivity) between sources.

Connectivity analyses were conducted on a priori regions identified in the emotional memory and semantic memory networks. The sources and their coordinates are listed in Table 1 and displayed in Figure 2. It is worth noting that our semantic and emotional memory networks did have one overlapping region: PCC/precuneus. This should not be surprising as this region is integral for both emotional and semantic memory processes (e.g. Touryan et al. 2007; Herbert et al. 2011), including successful memory recall (Fletcher et al. 1995; Burianova et al., 2010; Binder and Desai, 2011). To ensure that our emotional and semantic memory networks assessed unique aspects of network connectivity and memory processes, correlational analyses were conducted on all network connectivity measures (i.e. modularity measures, see below) in all frequency bands. Modularity variables across all frequency bands in both the emotional and semantic memory networks had no meaningful correlation patterns, suggesting that these two networks assessed unique aspects of memory processing (Supplementary Tables 45–60).

Functional connectivity estimation

Phase locking values (PLV; Lachaux et al., 1999), which measures variability of phase between two signals across trials, and subsequent modularity values were utilized to define connectivity strength between sources and networks. PLV is defined as the absolute value of the mean phase difference between two signals in given ROIs, and is expressed as a complex unit-length vector (Aydore et al., 2013). If the two signals in ROIs are independent then their relative phase will also be independent and the PLV will be zero. Conversely, if the phases of the two signals are strongly coupled then the PLV will approach unity, which will be one. In this way, the PLV is not unlike a correlation coefficient where numbers closer to one indicate that two regions are communicating more strongly (or firing in phase) with one another and numbers closer to zero indicate the two regions are not communicating (or firing out of phase) with one another. The PLV were calculated within respective frequency bands

Table 1. Sources and coordinates utilized for emotional and semantic memory network analyses

Region	X	Y	Z	Network
Left precuneus	-11.6	-57.5	36.7	Emotional memory
Left dACC	-6.5	18.0	26.1	Emotional memory
Left vACC	-6.8	33.9	1.6	Emotional memory
Left insula	-34.2	-4.3	2.2	Emotional memory
Right insula	35.1	-3.9	2.2	Emotional memory
Left medial prefrontal	-8.0	44.8	-4.9	Emotional memory
Left MFG	-31.3	41.2	16.5	Emotional memory
Right MFG	32.3	40.9	17.3	Emotional memory
Left fusiform	-35.7	-43.3	-19.7	Semantic memory
Right fusiform	35.9	-43.0	-19.2	Semantic memory
Left superior occipital	-29.7	-86.9	-1.0	Semantic memory
Right superior occipital	30.3	-86.3	-0.5	Semantic memory
Left prefrontal	-42.4	30.6	2.3	Semantic memory
Right prefrontal	44.9	29.7	4.5	Semantic memory
Left superior temporal	-52.1	-17.8	-4.4	Semantic memory
Right superior temporal	53.0	-14.0	-5.5	Semantic memory
Left inferior parietal	-22.8	-60.9	46.3	Semantic memory
Right inferior parietal	22.6	-59.5	48.1	Semantic memory
Left PCC/precuneus	-9.0	-45.4	17.5	Semantic memory
Right PCC/precuneus	9.8	-44.7	16.9	Semantic memory

Coordinates for the emotional memory network and semantic memory network were defined via MNE (Gramfort et al. 2013, Gramfort et al. 2014).

(θ : 4–8 Hz, α : 9–14 Hz, β : 18–22 Hz, γ : 25–50 Hz), from 0 to 500 ms after the onset of feedback in the math feedback task, as well as after the onset of feedback/font pairings in the memory task. Graph analyses were based on the PLV between all pairs of ROIs. In other words, for every subject, condition, and frequency band, we obtained a full 68×68 adjacency matrix of PLV between each region/source.

Weighted graphs construction

We converted full PLV adjacency matrices into sparse, undirected, weighted graphs which can be analyzed with graph measures using the Brain Connectivity Toolbox (Rubinov and Sporns, 2010). Graphs are comprised of nodes, being systems' elements (which in this case are brain areas), and edges/connections, indicating interactions between elements (here PLVs). In order to obtain a sparse, weighted, undirected graph/network, full adjacency matrices were thresholded, so that all the values below the threshold were set to 0, and the values above the threshold retained the original values (weights). For each matrix the threshold was set statistically (Liu et al., 2011, 2013). The threshold of 0.25 was utilized for every single graph/network, meaning that only the largest 25% of values for edges out of all possible edge values were left for each graph. From this, two sub-networks were isolated in the labels of interest from the full sparse graph representing the emotional memory and semantic memory networks.

Graph measures

The brain is a highly complex system, consisting of multiple functional sub-systems/sub-groups/neighborhoods or *modules* that represent nearly decomposable units (Simon, 1962; Figure 1). In other words, modules are groups of brain regions with many intra-module links, or connections within a local neighborhood or sub-group, but fewer inter-modular links to external groups, or brain regions outside a given neighborhood.

Systems such as these have the ability to aggregate modules that can perform independent functions (Bassett et al., 2010). Recent advances in network neuroscience suggest that brain networks typically exhibit a highly modular architecture, often described as a set or hierarchy of modules (or 'modules-within-modules' structure, Meunier et al., 2009; Bassett et al., 2010). Modules located in middle levels of a hierarchy constitute components that work together for specific functions that contribute to upper levels of the hierarchy. Conversely, they have also been shown to constitute segregated components that perform sophisticated functions which themselves can be further organized into several subdivided lower level modules. A hierarchical module structure has long been proposed to confer advantages and facilitate adaptability or resolvability in diverse information processing systems (Félix and Wagner, 2008).

Modularity structures can be disrupted by cognitively demanding stimuli or tasks, prompting networks to reconfigure into a more efficient, higher inter-modular integrated system. Such an integrated system, although more complex and metabolically expensive with respect to synchronizing inter-modular regions, may provide a means for information to be exchanged between regions across whole brain networks more efficiently. Thus, brain networks are usually thought to switch their configuration between a more efficient but costly task-dependent global topology, and a more organized but modular baseline topology. Consistent with this, past research has demonstrated that better modular architecture (higher modularity) in a baseline/resting brain is associated with better performance on various cognitive tasks (Stevens et al., 2012; Stanley et al., 2014; Gallon et al., 2015).

In brain networks, topological modules are often made up of anatomically neighboring and/or functionally related cortical regions, whereas inter-module connections tend to appear in brain regions relatively farther apart from one another (Meunier et al., 2010). Brain regions, or the constituent nodes of topological modules, are often anatomically co-localized in the brain (Bertolero et al., 2015). This arrangement seems to be advantageous in terms of minimizing the connection distance or wiring cost for intra-modular edges. For example, modularity analyses conducted on whole brain networks typically find that fronto-temporal, central, parietal, occipital and default-mode modules represent the highest level of the hierarchy. Little is known, however, about how the topological modularity of large-scale brain networks is related to other aspects of modularity psychologically (Meunier et al., 2010). For example, a brain network that is psychologically/cognitively meaningful (e.g. the emotional memory network) can be widely distributed among several anatomical modules. Traditional modularity analyses have more difficulty addressing the question of how and whether regions in psychologically meaningful networks work together more efficiently during cognitive tasks, or how a given network may optimally reorganize or subdivide in to sub-modules to facilitate a given psychological process. To address this question of modular organization with respect to psychological variables of interest, we created two measures of modularity that extend the rationale of modularity in whole brain network analyses to a priori defined sub-networks thought to play an integral role in psychological processes of interest. We define these two new measures as (select network modularity) and (sub-network modularity).

Select network modularity

Select network modularity is a measure that theoretically compares within module connectivity (within an a priori network of

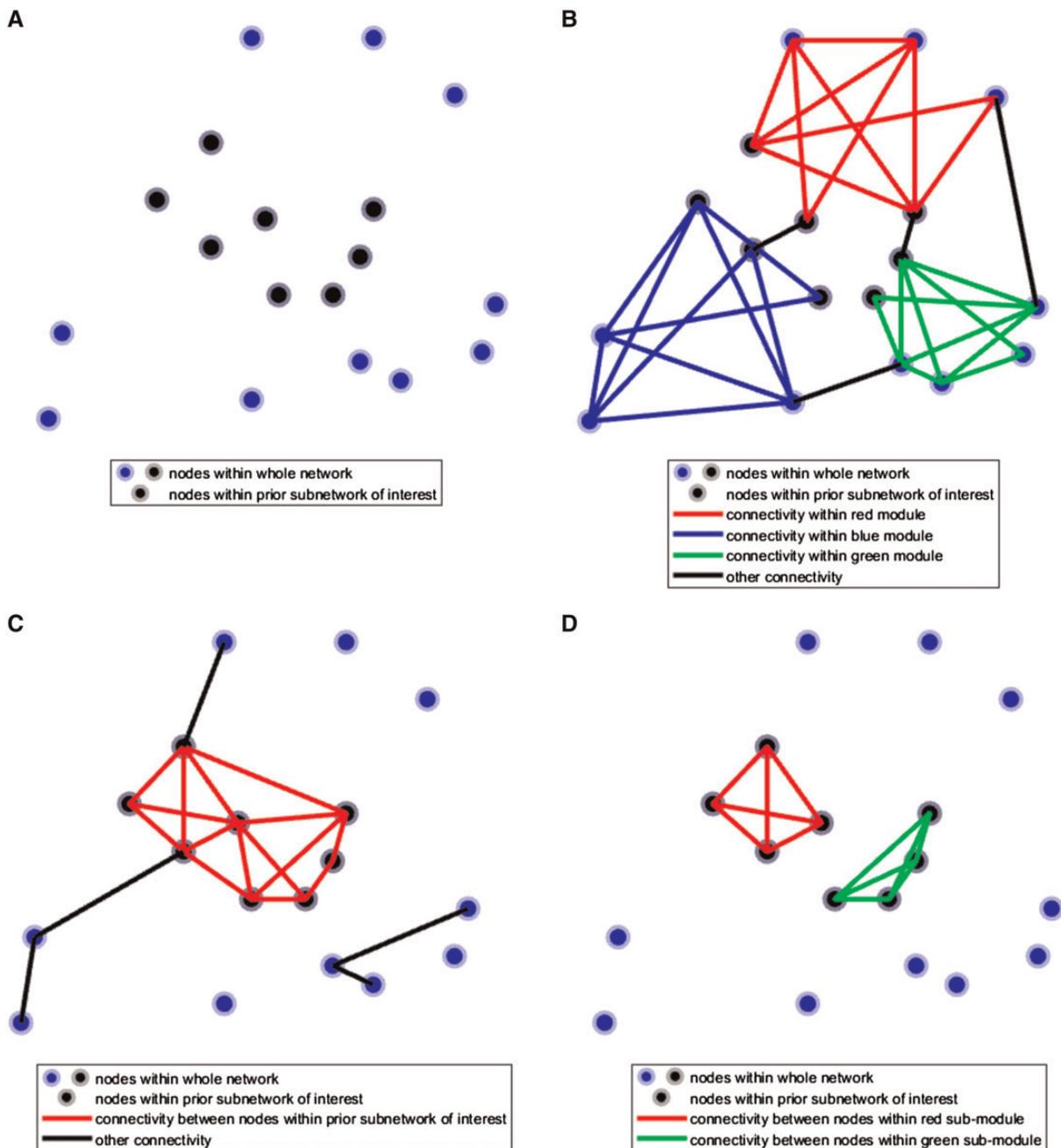


Fig. 1. A hypothetical whole brain network comprised of 18 nodes (blue nodes/circles), among which 8 nodes are of interest (an a priori network of interest, or APNI, which is denoted as black circles), e.g. brain regions selected from a meta-analysis that are hypothesized to be associated with a specific psychological function. (A) The nodes distributed throughout the brain sans connectivity. (B–D) Examples of network organizations with higher values for the three types of modularity discussed in this study. (B) Basic modularity analyses comprise a data-driven approach that mathematically parcellates the whole network into three modules (denoted as red, green and blue connectivity/lines) such that nodes (brain regions) are more effectively connected to nodes within its own given module compared with nodes within the other two modules (independent of how a priori regions of interest may interact during a given psychological process; Sporns, 2011). (C) Select network modularity essentially yields a ratio of the extent to which all regions within an APNI communicate with one another compared with the extent to which they communicate with other regions throughout the whole network. Greater select network modularity indicates that the nodes within the APNI are more efficiently interconnected with one another compared with other nodes in the whole brain, i.e. that which would be expected by chance. (D) Sub-network modularity measures the extent to which an APNI is organized into composite units or a collection of sub-modules. Higher sub-network modularity values indicate that a given number of sub-modules within the APNI work relatively independent or in parallel with one another compared with the entire network; this would indicate that the optimal state of a given APNI is better described as a function of connectivity within sub-networks of the larger APNI. Select network and sub-network modularity analyses thus could provide more nuanced information regarding how functionally relevant networks reorganize and work together during psychological processes of interest.

interest) to between module connectivity (operationalized as phase locking between pre-defined sub-networks or other regions across the whole brain network; Figure 1). Select network modularity was calculated using the equation below:

$$Q_{\text{select}} = \frac{1}{2m} \sum_{ij \in \text{select}} [A_{ij} - p_{ij}]$$

In the above equation, A is the actual PLV between i and j where, e.g. $i = \text{ACC}$ and $j = \text{DLPFC}$. $p_{ij} = k_i k_j / 2m$, where $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges attached to node i ; p_{ij} represents the connectivity value or PLV that i and j would have if connected by chance considering their connectivity with other regions across the brain. In the current study, p_{ij} is based on how i and j are connected with 66 other regions (68 total). Finally, $m = 1/2 \sum_{ij} A_{ij}$ and represents the summation of connectivity across the whole brain. Thus if the network of interest has a greater collection of hubs, select network modularity values would be more negative as hubs are interconnected with many regions across the brain, not just the two of interest. Select network modularity thus measures the integrated level or within network connectivity of an a priori network of interest with respect to either another network of interest, or the whole brain (in this study it is with respect to the whole brain). Higher select network modularity indicates that the nodes within a sub-network are more efficiently interconnected with one another compared with other nodes in the whole brain, i.e. that which would be expected by chance. This would suggest that this network was more active or efficient during a given cognitive process. Select network modularity is also valuable because some neural regions (i.e. hubs) are often integrated in many neural networks during different cognitive processes. Thus assessing the degree to which a region communicates with other ROIs within an a priori network of interest in relation to how much it communicates with regions in other networks in general provides an additional layer of understanding to network integrity and function as it relates to psychological processes of interest.

Sub-network modularity

For this measure, instead of conducting modularity analyses on the whole brain (the 68×68 adjacency matrices), we confined modularity analyses to pre-defined sub-networks of interest (e.g. modularity analyses were conducted within the emotional memory network only). Sub-network modularity decomposes a given a priori network of interest into mathematically defined secondary sub-networks that make the largest contributions to the overall modularity value (Figure 1). Sub-network modularity is calculated via the equation below.

$$Q_{\text{subnet}} = \frac{1}{2m^{\text{sub}}} \sum_{ij \in N} [A_{ij} - p_{ij}] \delta(c_i^{\text{sub}}, c_j^{\text{sub}})$$

where A_{ij} represents the weight of the edge between nodes i and j , c_i^{sub} is the secondary module to which node i is assigned, and $m^{\text{sub}} = 1/2 \sum_{ij \in \text{subnet}} A_{ij}$. The δ -function $\delta(u, v)$ is 1 if u and v are assigned to the same secondary module and 0 otherwise. Q_{subnet} can be calculated via connectivity within an a priori defined network. Given a sub-network and connectivity map within the sub-network, Q_{subnet} may vary according to how the secondary-sub-network/modules are constructed. In this study a heuristic method (Blondel et al., 2008) was used to optimize

the community structure within the sub-network, i.e. to maximize the Q_{subnet} value. Sub-network modularity thus measures how an a priori network of interest is organized into composite units or sub-modules. Greater sub-network modularity values indicate that the sub-modules within an a priori network of interest work relatively independent or parallel from one another compared with the entire network, indicating that a given network is better described as a function of sub-networks within the larger network. That is, the optimal state of the network is better described by connectivity within different (data-driven based) sub-networks compared with the connectivity between all nodes in a given a priori network of interest as a whole. This measure could provide valuable insight into the nature and degree to which larger neural networks of interest may reorganize and communicate during cognitively demanding tasks to facilitate a given psychological process of interest. The select network modularity and sub-network modularity analyses were conducted on the sources integral for the emotional and semantic memory networks outlined in Table 1.

Modularity analyses reveal common sub-networks for emotional but not semantic memory networks

To determine whether sub-network modularity analyses yielded systematic emotional and semantic memory sub-networks that were more strongly connected than the network as a whole we conducted initial modularity analyses on all participants' EEG activity in response to negative feedback. These analyses revealed a common set of sub-networks in the emotional memory network among most individuals that were better connected with one another compared with the emotional memory network as a whole (Figure 2). Specifically, 125 of the 152 participants yielded identical parcellation within the emotional memory network, regardless of condition. One sub-module consisted of the precuneus, dACC and bi-lateral Insula sources. The second sub-module consisted of vACC, mPFC and bi-lateral MFG sources. In the semantic memory network, however, no such parcellation patterns were found as the modular structure appeared to be more random. Among 152 participants, there were 38 different modular structures, some of which had two sub-modules and others that had three. Identical parcellation patterns exhibited regardless of condition thus might indicate the existence of intrinsic connections within the emotional memory network, whereas a more random modular structure suggests that there are more unstable connections between brain regions within the semantic memory network.

Results

Performance on the math feedback task

An initial 2 (Gender: Men or Women) \times 2 (Condition: DMT or PST) factorial ANOVA was conducted on participants' accuracy on the math feedback task (number correct/number attempted). This analysis yielded a main effect for gender, $F(1, 156) = 16.56$, $P < 0.001$, $d = 0.63$. There were no other main effects or interaction ($P_s > 0.38$). Given the well documented effects of stereotype threat on performance (for a review see Schmader et al., 2008), however, planned contrasts were also conducted to compare DMT women's performance on the math feedback task to the other three conditions. These analyses indicated that DMT women (i.e. those experiencing stereotype threat) performed worse on the math feedback task compared with the other three conditions, $t(1, 156) = 3.17$, $P = 0.002$, $d = 0.51$ (Table 2).

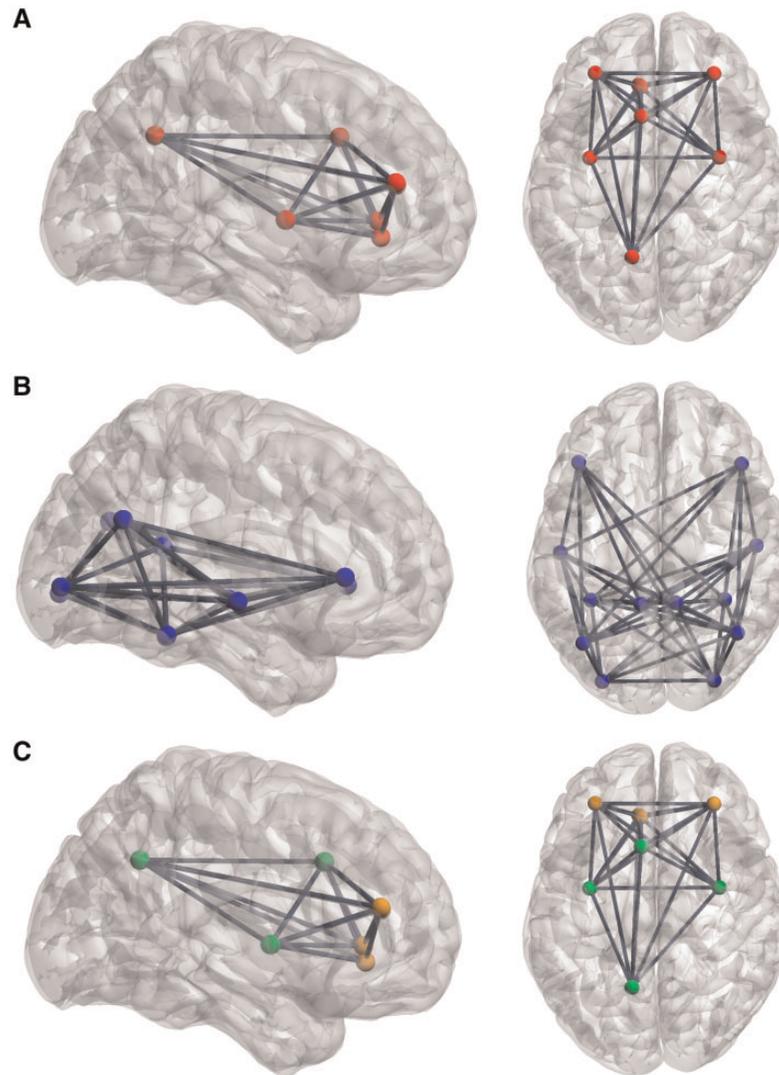


Fig. 2. Neural hubs utilized in the (A) emotional memory and (B) semantic memory connectivity analyses. (C) Results from sub-network modularity analyses conducted on emotional memory network revealed common sub-networks among precuneus, dACC and bi-lateral Insula sources (green dots) and the vACC, mPFC and bi-lateral MFG sources (yellow dots) in 125 out of 152 participants across gender and condition.

With respect to the role of task difficulty, given the nature of our math feedback task, which was lengthy and contained easy, medium and difficult questions, it is possible that these various problem types had variable effects on performance across groups. To examine this, a 2 (Gender: Men or Women) \times 2 (Condition: DMT or PST) \times 3 (Problem Type: easy, medium or hard) mixed factors ANOVA with repeated measures on the latter variable was conducted. This analysis yielded a main effect for gender, $F(1,156) = 12.38$, $P = 0.001$, $d = 0.55$, that was qualified by a problem type by condition interaction, $F(1, 156) = 4.44$, $P = 0.01$, $d = 0.35$, and problem type by gender interaction, $F(1,156) = 3.41$, $P = 0.03$, $d = 0.29$. Simple effects analyses using a Dunn–Sidak adjustment to control for multiple comparisons indicated that DMT women performed worse on easy, $F(1,156) = 4.17$, $P = 0.04$, $d = 0.33$, and difficult problems, $F(1,156) = 4.73$, $P = 0.03$, $d = 0.35$, compared with PST women. Men did not differ from one another with respect to condition, $P_s > 0.23$. DMT women also performed worse on easy, $F(1, 156) = 11.99$, $P = 0.001$, $d = 0.56$, and moderately difficult problems, $F(1, 156) = 3.77$, $P = 0.05$, $d = 0.31$, compared with DMT men. Women in the PST condition performed worse on easy,

$F(1, 156) = 6.51$, $P = 0.01$, $d = 0.41$, and moderate problems, $F(1, 156) = 5.64$, $P = 0.02$, $d = 0.38$, compared with men in the PST condition. Interestingly, only DMT women did not perform differently on easy, medium and difficult problem types, presumably because they underperformed across problem types in general ($P_s > 0.12$). All other groups showed the expected patterns, i.e. performing better on easy compared with moderate and difficult problems, and moderate compared with difficult problems ($P_s < 0.04$).

Performance on the standard math task

An additional 2 (Gender: Men or Women) \times 2 (Condition: DMT or PST) factorial ANOVA was conducted on participants' performance on the standard math task (i.e. the GRE problems). This analysis yielded no main effects ($P_s > 0.27$), however, there was the predicted interaction, $F(1, 156) = 4.43$, $P < 0.04$, $d = 0.35$ (Table 2). Simple effects analyses using a Dunn–Sidak correction to control for multiple comparisons were then conducted on gender; these analyses indicated that women in the DMT condition performed worse on the standard math test compared with

Table 2. Descriptive statistics for all primary variables of interest in the study

Variable	Condition	Gender	Mean	SD	N
Math feedback task score	DMT	Female	0.45	0.09	49
		Male	0.52	0.11	37
		Total	0.48	0.11	86
	PST	Female	0.47	0.12	38
		Male	0.53	0.11	36
		Total	0.50	0.12	74
	Total	Female	0.45	0.10	87
		Male	0.53	0.11	73
		Total	0.49	0.11	160
Standard math task score	DMT	Female	30.73	20.47	49
		Male	34.31	26.90	37
		Total	32.52	23.39	86
	PST	Female	43.77	30.31	38
		Male	30.62	25.12	36
		Total	37.37	28.50	74
	Total	Female	36.45	25.91	87
		Male	32.74	25.95	73
		Total	34.76	25.91	160
Memory accuracy for negative feedback	DMT	Female	0.19	0.19	48
		Male	0.19	0.21	37
		Total	0.19	0.20	85
	PST	Female	0.23	0.18	36
		Male	0.23	0.21	36
		Total	0.23	0.19	72
	Total	Female	0.21	0.18	84
		Male	0.21	0.21	73
		Total	0.21	0.20	157
Memory accuracy for positive feedback	DMT	Female	0.06	0.30	48
		Male	0.04	0.23	37
		Total	0.05	0.27	85
	PST	Female	0.08	0.23	36
		Male	0.03	0.26	36
		Total	0.06	0.25	72
	Total	Female	0.07	0.27	84
		Male	0.04	0.24	73
		Total	0.05	0.26	157

women in the PST condition, $F(1, 156) = 5.50$, $P = 0.02$, $d = 0.35$. There were no differences in performance between men in the DMT condition and PST condition. A comparable analysis on condition indicated that while there were no differences between men and women in the DMT condition ($P = 0.47$), women in the PST condition outperformed their male counterparts, $F(1, 156) = 4.876$, $P = 0.029$, $d = 0.35$. Planned contrasts indicated that women under stereotype threat trended toward performing worse on the math test compared with the other three conditions, $t(1, 156) = 1.28$, $P = 0.20$, $d = 0.20$.

It is worth noting that the other conditions (PST men and women specifically, not DMT women) unexpectedly varied in performance across math tasks, either better or worse on the math feedback and GRE tasks. It is possible that this pattern reflected the nature of the experiment (all participants performing individually in front of experimenters while brain activity is recorded and the fact that they were completing multiple math tasks that varied in length and difficulty over the course of an hour, which are far different circumstances compared with typical stereotype threat studies). To examine whether there were more stable performance differences across the course of the experiment we standardized accuracy variables from the two tasks, averaged those standardized scores together and conducted planned contrasts accordingly. A one-way ANOVA

conducted on the average standardized accuracy scores from both tasks yielded a significant between groups effect, $F(1, 156) = 5.45$, $P = 0.02$. Importantly, planned contrasts indicated that DMT women performed worse across both tasks compared with everyone else $t(1, 156) = -2.93$, $P = 0.004$, $d = 0.50$, and the means reflected a greater stability between differences across groups ($M_{DMTWomen} = -0.26$, $SD = 0.64$; $M_{PSTWomen} = 0.08$, $SD = 0.87$; $M_{DMTMen} = 0.14$, $SD = 0.73$; $M_{PSTMen} = 0.12$, $SD = 0.70$). Collectively, the data clearly indicate that DMT women performed worse than everyone else across tasks; thus these findings provide supporting evidence that the stereotype threat manipulation was successful.

Basic startle analyses

The following analyses excluded 11 individuals (1 woman in the DMT condition, 2 women in the PST condition, 5 men in the DMT condition and 3 men in the PST condition) identified as outliers via Grubbs test (also known as the extreme studentized deviate method) conducted on startle responses to positive and negative feedback on the math feedback task, i.e. these individuals had Z scores >3.5 (s.d.) from the grand mean ($P < 0.05$) of startle values. An additional seven participants were excluded for not having at least one startle response for fonts associated

Table 3. Descriptive statistics for all primary variables of interest in the study continued

Variable	Condition	Gender	Mean	SD	N
Average peak startle of negative feedback	DMT	Female	12.79	4.33	45
		Male	10.37	3.28	32
		Total	11.79	4.09	77
	PST	Female	12.90	5.00	34
		Male	11.79	4.56	31
		Total	12.32	4.79	65
	Total	Female	12.79	4.60	79
		Male	11.07	4.60	63
		Total	12.03	4.41	142
Average peak startle of positive feedback	DMT	Female	12.27	3.89	45
		Male	10.65	3.71	32
		Total	11.59	3.89	77
	PST	Female	12.05	4.22	34
		Male	10.72	4.09	31
		Total	11.41	4.17	65
	Total	Female	12.17	4.01	79
		Male	10.68	3.86	63
		Total	11.51	4.00	142
Math devaluing	DMT	Female	1.89	0.88	48
		Male	2.09	0.70	35
		Total	1.97	0.81	83
	PST	Female	1.92	0.97	37
		Male	2.19	0.88	35
		Total	2.05	0.93	72
	Total	Female	1.90	0.92	86
		Male	2.14	0.79	70
		Total	2.00	0.87	156
Self-enhancement	DMT	Female	6.48	1.14	46
		Male	6.80	1.14	33
		Total	6.61	1.14	79
	PST	Female	6.68	1.03	38
		Male	6.48	1.03	34
		Total	6.59	1.03	72
	Total	Female	6.57	1.10	84
		Male	6.63	1.09	67
		Total	6.60	1.09	151

with negative or positive feedback (three DMT women, two PST women, one DMT man, one PST man), and one PST male was excluded for experiencing an abnormal number of startle probes (4 SDs above the mean, denoting a software error). A 2 (Gender: Men, Women) \times 2 (Task Description: DMT, PST) \times 2 (Feedback Type: Negative, Positive) mixed factors ANOVA with repeated measures on the latter variable was then conducted on participant's startle values. These analyses revealed a marginal effect for feedback type indicating participants had larger startle responses to fonts associated with negative feedback compared with positive feedback ($P = 0.09$, $d = 0.29$), and a main effect for gender, $F(1, 137) = 5.972$, $P < 0.02$, $d = 0.41$, indicating that women had larger startle responses to feedback overall. No other effects were significant ($P_s > 0.20$, Table 3).

Exploratory non-linear amygdala responses to feedback over time

Amygdala activity has been shown to fluctuate over time and exhibit non-linear responses to threatening or arousing stimuli. In the aggregate, our startle results demonstrated that participants showed greater startle amplitudes to negative feedback overall, and that women elicited larger startle amplitudes to all feedback compared with men. However, given that startle

measures were derived with respect to common practices in the literature, which place stringent parameters on what constitutes a blink response, it is possible that meaningful fluctuations in amygdala responses to performance feedback were excluded from these analyses; this standard approach could then obscure meaningful non-linear amygdala responses to feedback. Thus in an attempt to determine whether ST women exhibited non-linear amygdala responses to performance feedback, and threatening feedback specifically (i.e. negative feedback) over time, all startle responses elicited to feedback (i.e. regardless of whether they were originally classified as a startle response) were examined across the entirety of the math feedback task.

To observe fluctuations in startle amplitudes over time in all groups, quadratic multilevel mixed models were run to account for within subject variability, specifying a first order autoregressive covariance matrix (Jongerling *et al.* 2015). All mixed models were run using SPSS 24 (IBM, New York, USA). The model consisted of condition and gender as categorical factors, in addition to time (represented as trial number), and a time by time interaction to allow for quadratic estimation, predicting startle amplitudes to all feedback. A full factorial model was run, excluding any interactions between time and the quadratic representation of time to avoid any cubic modeling. These analyses

included the same participants that were used in the aggregate startle analyses.

Results demonstrated a four-way interaction between condition, gender and the quadratic representation of time, $F(1, 872.26) = 5.91$, $P < 0.02$ (Supplementary Figure S1). Simple effects revealed that only DMT women demonstrated a unique quadratic relationship in their startle responses to feedback over time; all other groups exhibited amygdala patterns suggesting they habituated to feedback over time [$b = 0.003$, 95% CI (0.002, 0.004), $SE = 0.0007$, $t = 4.22$, $P < 0.001$]. Specifically, in the beginning of the task DMT women demonstrated a decreasing slope that curved into a quadratic function [$b = -0.222$, 95% CI (-0.305, -0.141), $SE = 0.0420$, $t = -5.311$, $P < 0.001$], but by the end of the task their slope uniquely increased, indicating a marginal increase in amygdala activity ($b = 0.062$, $P = 0.072$). Control women and all male participants did not demonstrate any increases in startle amplitude toward the end of the task ($P_s > 0.39$) or quadratic relationships in their startle amplitudes over time ($P_s > 0.605$).

Moreover, the non-linear amygdala response exhibited by DMT women was largely driven by stereotype confirming feedback, i.e. negative feedback. Running an identical analysis on negative feedback only, results demonstrated a four-way interaction between condition, gender and the quadratic representation of time $F(1, 624.07) = 4.40$, $P < 0.05$. Simple effects revealed that only DMT women demonstrated a unique quadratic relationship in their startle amplitudes to negative feedback over time compared with all other groups, which exhibited startle amplitudes that habituated to negative feedback [$b = 0.003$, 95% CI (0.001, 0.005), $SE = 0.0008$, $t = 3.56$, $P < 0.001$]. Again, in the beginning of the task DMT women exhibited a decreasing slope curving into a quadratic function [$b = -0.237$, 95% CI (-0.342, -0.132), $SE = 0.0535$, $t = -4.427$, $P < 0.001$]. Unlike responses to all performance feedback, DMT women's amplitudes did not significantly increase in response to negative feedback toward the end of the task ($P = 0.14$). However, DMT women's startle amplitudes to negative feedback were marginally larger compared with PST women ($P = 0.082$). PST women and all male participants did not demonstrate any increases in startle amplitude toward the end of the task ($P_s > 0.39$) or quadratic relationships in their startle amplitudes over time ($P_s > 0.656$). Models run with correct feedback demonstrated no quadratic relationships among any group ($P = 0.166$). Thus, when accounting for non-linear effects, these findings reveal a more complex, nuanced stress response to negative feedback among women in SBS contexts where the stress response is more intense at the beginning of the threatening task, plateaus more quickly, but then rebounds toward the end of the task. These findings are consistent with recent examinations of performance differences among DMT women that indicate performance decrements are more pronounced at the beginning as opposed to the middle and end of a math task, especially when using a math task like the one used in this study, which is much longer than math tasks traditionally used in the ST literature (Liu et al., 2017).

Performance on the memory test

An initial examination of the memory data revealed that three participants (one woman in the DMT condition and two women in the PST condition) were identified as outliers in a Grubbs' test conducted on positive and negative feedback memory scores. These individuals were excluded from any analyses involving memory scores. Initial one-sample t-test analyses revealed that the means for our d' measures were significantly above chance.

Specifically, given that with respect to d' a score of 0 is considered chance, one sample t-tests across conditions revealed that both d' scores for positive [$t(1,156) = 2.64$, $P = 0.009$, $d = 0.16$] and negative feedback [$t(1,156) = 13.14$, $P < 0.001$, $d = 1.05$] were significantly above chance. To examine behavioral effects of SBS on memory for font/feedback pairings seen during the math feedback task, a 2 (Gender: Men, Women) \times 2 (Task Description: DMT, PST) \times 2 (Feedback Type: Negative, Positive) mixed factors ANOVA with repeated measures on the latter variable was conducted on participant's memory scores for fonts. These analyses only revealed a main effect for feedback type indicating participants had better recall for fonts associated with negative feedback compared with positive feedback, $F(1,153) = 44.78$, $P < 0.001$, $d = 1.09$. No other effects were significant ($P_s > 0.43$, Table 2).

Unexpectedly, stereotype threat did not uniquely enhance recall for information associated with solving problems incorrectly, at least behaviorally. This finding is not surprising given the well documented negativity bias found in the literature (Vaish et al., 2008). Thus, when in a motivated performance situation such as the one primed in this study, individuals may have attended to and encoded negative information better than positive information overall. However, there are multiple routes to encoding information, including via emotion-based and/or cognitive-based pathways. We posited that only women in SBS contexts should exhibit a relationship between stress-based amygdala responses to negative feedback, neural indices of better memory encoding of negative feedback and behavioral indices of better encoding of negative feedback. To do this we focused on three variables: (V1) startle responses (amygdala activity) elicited to negative font/feedback pairings during the math feedback task, (V2) connectivity between regions in the emotional memory network (modularity values) during negative font/feedback hits (accurately identified font/feedback pairings during memory test) and (V3) memory for the negative font/feedback pairings themselves (d' for positive and negative font/feedback pairings). We tested these hypotheses directly via a series of path analyses.

Stereotype threatened women encode negative feedback through the emotional memory network

The following analyses excluded all participants excluded for basic memory and startle analyses in addition to 35 participants who did not have enough valid trials to compute modularity values (10 DMT women, 8 PST women, 6 DMT men, 11 PST men). All basic modularity analyses are reported in the Supplementary Results p. 3–8. To determine whether there was a link between V1, V2 and V3 described above, double moderated mediation analyses were conducted. We tested for double moderated mediation by deriving unstandardized regression coefficients and 95% bias-corrected confidence intervals (CIs) from 10000 bootstrap estimates (Hayes, 2013; model 72). 95% CIs are considered significant if the interval [e.g. (0.5, 0.9)] does not contain zero (Cumming, 2008). In separate models for negative and positive font-related variables and emotion network connectivity within each of the four frequency bands (sub-network and select network modularity values) on trials in the memory test associated with hits for negative and positive font/feedback pairings, we entered variables in a manner that most closely represented the temporal nature in which the dependent variables were collected. Thus, startle responses (V1) were utilized as a predictor, emotion network connectivity in response to fonts on hit trials (V2) served as the mediator, d' for

negative or positive font/feedback pairings (V3) was the outcome variable and condition and gender served as the two moderator variables. Multiple comparisons were controlled for using the Benjamini–Hochberg False Discovery Rate (FDR) procedure (Benjamini and Hochberg, 1995), using a standard q -level of 0.1 (Singh and Phillips, 2010; Ewald et al., 2012; Pintzinger et al., 2017). Given that the following double moderated mediation analyses inherently account for comparisons made between condition and gender in a given model, we accounted for the number of frequency band comparisons made in each respective group of analyses (four total comparisons for each set of independent analyses).

These analyses revealed that the indirect pathway was only significant for women in the DMT condition in the β band on correctly identified negative font/feedback pairings, $b=0.009$, 95% CI [0.0013, 0.0185] (Figure 3). The relationship between stereotype threatened women's startle responses and memory for fonts associated with negative feedback was mediated by better connectivity within the emotion network on memory test trials where they accurately identified negative font/feedback pairings seen during the math feedback task. Similar patterns were found in θ , α and γ frequency bands, however, significance was not reached. These patterns were not evident among men in the DMT condition or women and men in the PST condition or when variables associated with correct font/feedback pairings were included (all CIs contained zero). These relationships also were not evident when using select network modularity values as the index for network connectivity (Supplementary Results p. 9, Tables 1–8), suggesting that connectivity between two selective sub-regions within the emotional memory network were driving behavioral memory effects for negative feedback as a function of initial amygdala activity elicited in response to negative feedback.

To determine what exact regions were driving this effect, exploratory post-hoc analyses were conducted to approximate the composition of the sub-networks among individuals who exhibited d' scores for negative font-feedback pairings that were >1 s.d. above the mean. These analyses indicated that among stereotype threatened women, connectivity between DACC, VACC, MPFC and bilateral MFG was particularly robust, i.e. these regions comprised the dominant module in the sub-network among individuals exhibiting more accurate memory for negative feedback. Conversely, among all other individuals, the most dominant module consisted of DACC and bilateral MFG (but the modularity, i.e. connectivity between these regions, was much lower). This suggests that VACC may play a particularly important role in interacting with other regions known for basic attentional and memory encoding processes to facilitate more efficacious encoding of emotional stimuli under stress.

Men's memory for negative feedback modulated by the semantic memory network

The following analyses excluded the same participants that were excluded from emotional memory network moderated mediation analyses. Behavioral findings indicated that all individuals encode negative feedback more accurately than positive feedback, however, given that diagnostic math or stereotype neutral contexts should be largely devoid of emotion for men, it is possible that they may encode negative feedback through a different neural mechanism and through semantic memory processes specifically.

Initial double moderated mediation analyses on semantic memory connectivity (selective and sub-network modularity)

similar to those conducted for emotional memory network connectivity did not provide direct evidence that men encoded negative feedback through the semantic memory network. We then examined whether there was any relationship in general between semantic memory network connectivity and memory. In separate models, double moderated regression analyses for semantic network connectivity within each of the four frequency bands (sub-network and select network modularity values) on trials in the memory test associated with hits for negative and positive font/feedback pairings and memory for font/feedback pairings were conducted. Multiple comparisons between frequency bands were controlled for using the Benjamini–Hochberg FDR procedure (Benjamini and Hochberg, 1995), using a q -level of 0.1 (Singh and Phillips, 2010; Ewald et al., 2012; Pintzinger et al., 2017). We tested for double moderation by deriving unstandardized regression coefficients from 10 000 bootstrap estimates (Hayes, 2013). Using model 3 in PROCESS, gender (women = 0, men = 1) and condition (DMT = 0, PST = 1) were entered as M and W in the model and semantic memory connectivity elicited in response to font/feedback pairings on the memory test was entered as X , predicting memory accuracy for font/feedback pairings on the math feedback task.

These analyses yielded a main effect for gender in the γ band, $b=0.10$, $P = 0.05$ (Figure 4). Conditional effects indicated that the slopes for the relationship between memory accuracy and semantic memory network connectivity were significant for men (DMT men: $b = 11.29$, $P < 0.02$; PST men: $b = 14.01$, $P < 0.01$) but not women ($P > 0.41$). These relationships were not evident among any other variable combinations. Thus, analyses suggest that men encoded negative feedback more accurately to the extent they exhibited greater connectivity within the semantic network (select network modularity, i.e. connectivity between the semantic memory network in relation to whole brain connectivity) in the γ band during trials in which they accurately identified previously seen fonts associated with negative feedback. Similar patterns were evident but marginal with respect to sub-network modularity (Supplementary Results p. 9–11, Tables 9–16).

Downstream consequences of emotional memory encoding on math self-perceptions

We next examined the downstream consequences of emotional memory encoding on self-perceptions within the math domain to determine if biased encoding processes undermined math self-enhancement and math valuing. The same participants excluded from startle, memory and modularity analyses were excluded from the following analyses in addition to eight participants who did not complete math self-enhancement/valuing measures (two females in the DMT condition, four males in the DMT condition and one male in the PST condition). To compare the effects of emotional memory encoding processes on STEM self-perceptions in men and women in the DMT and PST conditions exploratory multi-group structural equation modeling analyses with 10 000 bootstraps was performed in AMOS (Arbuckle, 1997). An initial analysis conducted on the path best reflecting the temporal nature of the experiment (startle responses to negative feedback \rightarrow performance on standard math task \rightarrow memory connectivity to accurately identified negative feedback $\rightarrow d'$ for negative feedback \rightarrow math self-enhancement/valuing) yielded suboptimal model parameters, CFI = 0.757, RMSEA = 0.064. Next, we replicated the fully saturated serial mediation PROCESS model in AMOS (Hayes, 2013; model 6) and added all pathways to our variables of interest

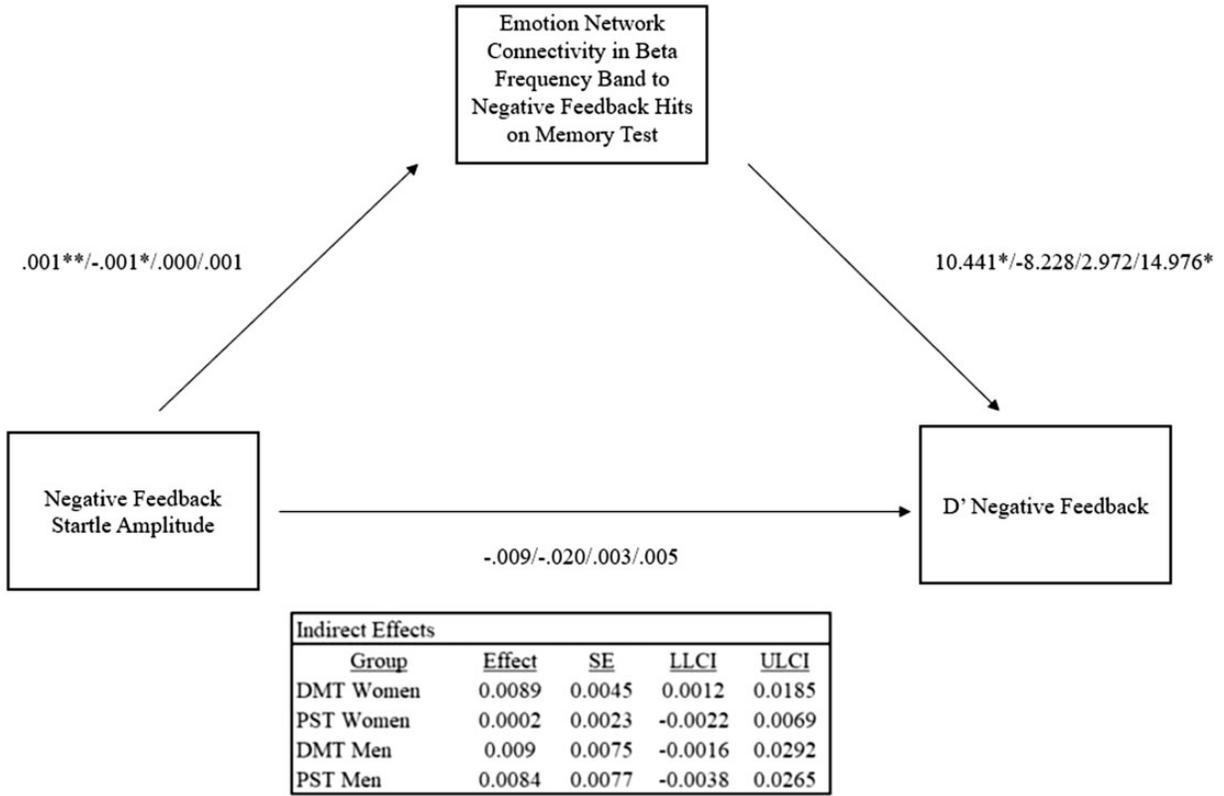


Fig. 3. Moderated mediation depicting the link between negative feedback startle amplitudes and memory accuracy for negative font feedback pairs mediated by emotional memory network connectivity elicited on trials where women accurately identified previously seen negative font-feedback pairings. Order of β s reported for each path correspond to the DMT women/DMT men/PST women/PST men conditions, respectively. * $P < 0.05$, ** $P < 0.01$.

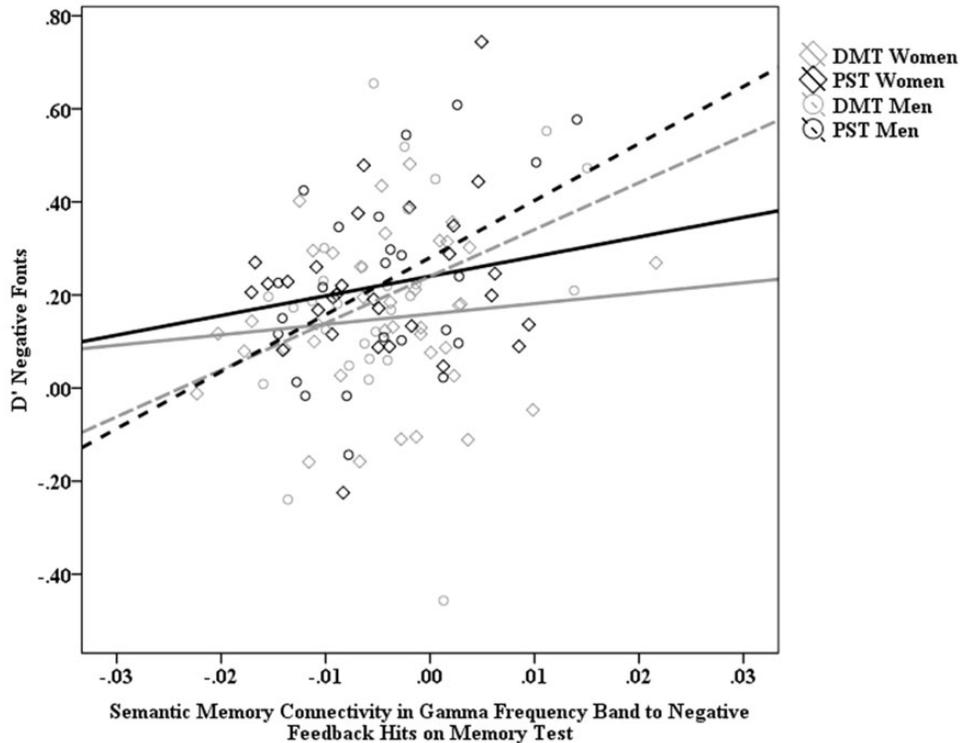


Fig. 4. Scatter plot depicting relationship between men's negative font/feedback d' scores and semantic memory select network modularity (connectivity) elicited on trials where men accurately identified previously seen negative font/feedback pairings.

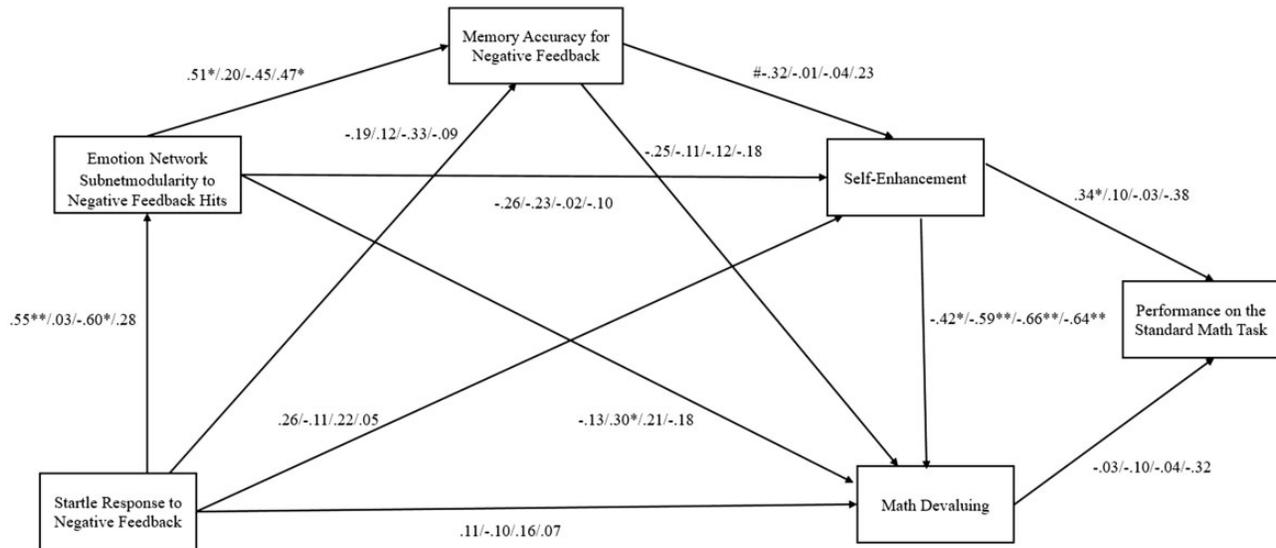


Fig. 5. Path model for link between stereotype threatened women's amygdala activity in response to negative feedback, emotional memory network connectivity elicited on trials where women accurately identified previously seen negative font/feedback pairings, their negative font/feedback d' scores, math self-enhancement and math valuing scores and performance. Order of β s reported for each path corresponds to the DMT women/DMT men/PST women/PST men conditions, respectively. # $P = 0.07$; * $P < 0.05$.

(math performance on the standard math task, math self-enhancement, and math valuing). Performance on the standard math task was moved to solely relate to self-enhancement and math valuing and pathways were trimmed from this model until adequate fit was achieved. According to fit criteria, our final model with all parameters freely estimated within the four groups fit the data well (CFI = 0.99, RMSEA = 0.02) and was therefore able to reproduce the data (Hu and Bentler, 1999; Measuring Model Fit, Kenny, 2015): startle responses to negative feedback → memory connectivity to accurately identified negative feedback → d' for negative feedback → math self-enhancement/valuing → performance on standard math task (Figure 5).

Nested model comparisons revealed that all structural weights between the four groups differed, $\chi^2(36) = 57.42$, $P < 0.02$, suggesting that the downstream consequences of emotional memory encoding processes on self-perceptions and performance in STEM domains differed across groups (men and women in the DMT and PST conditions). An examination of the indirect effects revealed that women in the DMT condition exhibited an indirect relationship between emotional memory encoding and math self-enhancement [$b = -0.02$, 95% CI (-0.089, -0.002), $P < 0.03$], math valuing [$b = 0.01$, 95% CI (0.0003, 0.038), $P < 0.03$], and performance on the standard math task [$b = -0.16$, 95% CI (-0.793, -0.006), $P < 0.04$; Table 4]. These patterns were not found in models with men in the DMT condition or women and men in the PST condition (all CIs contained 0; Table 4) suggesting that these downstream consequences are specific to stereotype threatened women.

Discussion

Findings from this study indicate that in addition to undermining performance in general, stereotype threatening contexts, and SBS specifically, engender encoding of negative, stereotype confirming feedback through emotional memory encoding processes. Consistent with a general negativity bias (Vaish et al., 2008), while all participants encoded fonts associated with negative feedback more efficaciously compared with positive

feedback (behaviorally), only stereotype threatened women exhibited a non-linear stress response and a link between this stress response and the encoding of negative feedback via neural markers of emotional memory processes. Stereotype threatened women's increased amygdala activity (startle responses) elicited to negative feedback during the performance was associated with increased connectivity between regions integral for emotional memory encoding during memory test trials where they accurately recognized negative font-feedback pairings seen during the performance. This increased connectivity between emotional memory regions, in turn, predicted better memory accuracy for negative font-feedback pairings seen during the performance. These patterns were not evident among correct feedback, men in the DMT condition, men or women in the PST condition, or other types of memory test variables (i.e. misses and false alarms, see Supplementary Results p. 9, Tables 1–4). This suggests that stereotype threatened women remembered negative feedback better via stress-induced emotional memory processes.

The emotional memory bias toward negative feedback, in turn, had downstream consequences on women's perceptions and performance within the math domain. Stereotype threatened women reported having less math ability in relation to their STEM peers, decreases in the importance of the math domain to their self-concept (who they are as a person) and exhibited decreased performance on a standard math test to the extent they encoded negative feedback more accurately via an amygdala-based emotional memory encoding process. These patterns were not evident among non-threatened women or men in either condition, suggesting SBS has particularly deleterious consequences for women in STEM domains. These findings are consistent with past research indicating that encoding biases are specific to negative feedback in stereotype threatening contexts, and can have downstream consequences on both performance and the experience of negative affect in general (Mangels et al., 2012; Forbes and Leitner, 2014; Forbes et al., 2015).

These findings extend upon past studies by providing a direct link between this bias and emotional memory encoding

Table 4. Results from path analysis on participants' amygdala activity in response to negative feedback (B1), emotional memory network connectivity elicited on trials where individuals accurately identified previously seen negative font/feedback pairings (B2), their negative font/feedback d' scores (B3), math self-enhancement scores and math valuing scores (B4)

Indirect path	DMT women		PST women		DMT men		PST men	
	LLCI	ULCI	LLCI	ULCI	LLCI	ULCI	LLCI	ULCI
B1 → B2 → B3 → B4	0.0003*	0.038*	-0.002	0.002	-0.019	0.037	-0.050	0.004
B1 → B2 → B3	-0.089*	-0.002*	-0.003	0.003	-0.092	0.037	-0.010	0.060

Table 5. Descriptive statistics for emotional network sub-network modularity to negative hits during the memory task in the main analyses

Variable	Condition	Gender	Mean	SD	N	
Emotional network sub-network modularity negative hits θ	DMT	Female	0.0880	0.0079	38	
		Male	0.0881	0.0091	28	
		Total	0.0880	0.0083	66	
	PST	Female	0.0875	0.0097	30	
		Male	0.0832	0.0102	24	
		Total	0.0856	0.0101	54	
	Total	Female	0.0877	0.0087	68	
		Male	0.0858	0.0098	52	
		Total	0.0869	0.0092	120	
	Emotional network sub-network modularity negative hits α	DMT	Female	0.0900	0.0099	38
			Male	0.0915	0.0153	28
			Total	0.0904	0.0124	66
PST		Female	0.0925	0.0090	30	
		Male	0.0886	0.0107	24	
		Total	0.0908	0.0099	54	
Total		Female	0.0909	0.0096	68	
		Male	0.0902	0.0133	52	
		Total	0.0906	0.0113	120	
Emotional network sub-network modularity negative hits β		DMT	Female	0.0883	0.0084	38
			Male	0.0908	0.0095	28
			Total	0.0894	0.0089	66
	PST	Female	0.0886	0.0094	30	
		Male	0.0874	0.0062	24	
		Total	0.0880	0.0081	54	
	Total	Female	0.0884	0.0088	68	
		Male	0.0893	0.0083	52	
		Total	0.0888	0.0085	120	
	Emotional network sub-network modularity negative hits γ	DMT	Female	0.0921	0.0110	38
			Male	0.0884	0.0130	28
			Total	0.0905	0.0119	66
PST		Female	0.0859	0.0099	30	
		Male	0.0882	0.0082	24	
		Total	0.0872	0.0089	54	
Total		Female	0.0904	0.0099	68	
		Male	0.0873	0.0116	52	
		Total	0.0890	0.0108	120	

Note that for Tables 5–8, 40 participants were excluded for not having enough valid EEG trials for hits during the memory test to accurately calculate modularity values for memory test hits. Five of these participants were also either outliers on our startle measure or memory d' measure.

processes. They also highlight a potential mechanism for why women report more negative math attitudes, less interest in STEM domains and less math confidence in general compared with men (e.g. Nosek *et al.*, 2002; Else-quest *et al.*, 2010). If SBS biases women toward encoding information that is consistent with the negative group stereotype over and above positive, stereotype inconsistent information, it paves the way for a vicious, potentially ruminative cycle in STEM domains. In addition to undermining performance, this bias may undermine self-enhancing strategies that are necessary to cope with variations in performance in valued domains, as well as the

importance of the valued domain to the self. Through this process it is clear to see how these subtle biases can foster more negative STEM perceptions and interest in STEM domains over time compared with men.

Furthermore, these findings suggest an enhanced role for more indirect, non-conscious processes in stereotype threatening contexts. Specifically, there appeared to be a basic disconnect between how the brain was responding to threatening feedback within milliseconds of exposure to said feedback and explicit math self-perceptions reported post-experiment. For instance, with respect to math perceptions, ST women did not

Table 6. Descriptive statistics for semantic network select network modularity to negative hits during the memory task in the main analyses

Variable	Condition	Gender	Mean	SD	N	
Semantic network select network modularity negative hits θ	DMT	Female	-0.0036	0.0061	38	
		Male	-0.0061	0.0061	28	
		Total	-0.0076	0.0062	66	
	PST	Female	-0.0077	0.0081	30	
		Male	-0.0060	0.0095	24	
		Total	-0.0070	0.0087	54	
	Total	Female	-0.0083	0.0070	68	
		Male	-0.0061	0.0078	52	
		Total	-0.0073	0.0073	120	
	Semantic network select network modularity negative hits α	DMT	Female	-0.0059	0.0086	38
			Male	-0.0049	0.0089	28
			Total	-0.0054	0.0087	66
PST		Female	-0.0009	0.0076	29	
		Male	0.0053	0.0092	24	
		Total	-0.0028	0.0086	54	
Total		Female	-0.0036	0.0085	68	
		Male	-0.0051	0.0090	52	
		Total	-0.0043	0.0087	120	
Semantic network select network modularity negative hits β		DMT	Female	-0.0036	0.0074	38
			Male	-0.0027	0.0089	28
			Total	-0.0032	0.0080	66
	PST	Female	-0.0052	0.0065	30	
		Male	-0.0046	0.0070	24	
		Total	-0.0049	0.0067	54	
	Total	Female	-0.0043	0.0070	68	
		Male	-0.0036	0.0081	52	
		Total	-0.0040	0.0074	120	
	Semantic network select network modularity negative hits γ	DMT	Female	-0.0043	0.0085	38
			Male	-0.0046	0.0078	28
			Total	-0.0044	0.0082	66
PST		Female	-0.0048	0.0076	30	
		Male	-0.0044	0.0076	24	
		Total	-0.0046	0.0076	54	
Total		Female	-0.0045	0.0081	68	
		Male	-0.0045	0.0077	52	
		Total	-0.0045	0.0079	120	

differ from men and women in the other conditions in the extent to which they explicitly valued math or thought they were the same as, if not better than, their male and female math peers post-task (Supplementary Results p. 1). However, the deleterious consequences of ST were evident when taking into consideration what's going on 'under the hood', i.e. in the brain, with respect to neural regions integral for emotion and emotional memory processes. Here we found an indirect link between emotional reactions to negative feedback (non-linear startle, i.e. amygdala, responses predicting increases in emotional memory network connectivity) and downstream decreases in math devaluing and self-enhancement. In other words, much like with many effects in social psychology, there was a difference between what people consciously report and experience and how they actually reacted to the situation as it unfolded, a la Nisbett and Wilson (1977), and more appropriately, e.g. Steele and Aronson (1995), who found that ST effects occurred in spite of participants not reporting any knowledge of experiencing anxiety or stereotype activation associated with the manipulation. A better understanding of the role that more non-conscious indirect processes play in ST contexts could ultimately help explain current discrepancies found in the literature between women's underperformance on math tests of importance (like the GRE) and decisions to leave STEM fields at

disproportionate rates compared with men and self-reported decreased feelings of STEM marginalization or outperforming men on academic tasks in general (Deary et al., 2007).

With respect to the network analyses, findings also suggest connectivity between two sub-networks within the emotional memory network were particularly important for emotional memory encoding processes with respect to women's accuracy for negative font-feedback pairings on the memory test itself. These analyses provide insight on the nature of these neural processes with respect to behavioral measures. Among other things, they suggest that neural network interactions (e.g. emotion and memory networks) may be more nuanced, with certain regions in each network playing a more integral role in instantiating a given psychological process than others and the vACC playing a particularly unique role in this process in stressful contexts. Given the nature of our sub-network modularity variable, where connectivity between sub-networks in a hierarchy are compared with connectivity between regions in the network as a whole, our findings suggest that emotional memory processes may arise via more bottom-up sub-network interactions as opposed to more top-down hierarchical network interactions. This is not surprising as emotional memory processes inherently rely on the coordination between emotion and memory networks, but our findings provide insight in to these

Table 7. Descriptive statistics for emotional network select network modularity to negative hits during the memory task

Variable	Condition	Gender	Mean	SD	N	
Emotional network select network modularity negative hits θ	DMT	Female	-0.0036	0.0058	38	
		Male	-0.0036	0.0069	28	
		Total	-0.0036	0.0062	66	
	PST	Female	-0.0030	0.0070	30	
		Male	-0.0050	0.0081	24	
		Total	-0.0038	0.0074	54	
	Total	Female	-0.0033	0.0063	68	
		Male	-0.0042	0.0074	52	
		Total	-0.0037	0.0068	120	
	Emotional network select network modularity negative hits α	DMT	Female	-0.0011	0.0061	38
			Male	-0.0017	0.0075	28
			Total	-0.0013	0.0067	66
PST		Female	-0.0008	0.0055	30	
		Male	-0.0012	0.0058	24	
		Total	-0.0010	0.0056	54	
Total		Female	-0.0009	0.0058	68	
		Male	-0.0015	0.0067	52	
		Total	-0.0012	0.0061	120	
Emotional network select network modularity negative hits β		DMT	Female	0.0010	0.0070	38
			Male	-0.0006	0.0048	28
			Total	0.0003	0.0062	66
	PST	Female	0.0012	0.0070	30	
		Male	-0.0005	0.0054	24	
		Total	0.0004	0.0063	54	
	Total	Female	0.0011	0.0070	68	
		Male	-0.0005	0.0050	52	
		Total	0.0004	0.0062	120	
	Emotional network select network modularity negative hits γ	DMT	Female	-0.0011	0.0062	38
			Male	-0.0020	0.0059	28
			Total	-0.0015	0.0061	66
PST		Female	0.0003	0.0080	30	
		Male	-0.0028	0.0075	24	
		Total	-0.0010	0.0078	54	
Total		Female	-0.0005	0.0070	68	
		Male	-0.0024	0.0067	52	
		Total	-0.0013	0.0068	120	

interactions by highlighting what regions within each network work better together and how to facilitate emotional memory encoding processes in general and when individuals are placed in more stressful, emotionally arousing contexts. Importantly, future neuroimaging studies could utilize the sub-network and select network modularity equations developed for this study to help better understand more nuanced interactions between brain regions involved in a variety of cognitive processes and in relation to top-down and bottom-up hierarchical processing in general.

Conversely, other findings revealed that men's accuracy for negative font-feedback pairings (in both conditions) was modulated by increased connectivity between regions integral for semantic memory processes during memory test trials where they accurately recognized negative font-feedback pairings. Given the nature of the modularity measure involved in these analyses (select network modularity), findings suggest connectivity within the semantic network as a whole was more pronounced in comparison to all other possible connections in the brain that would be expected to occur by chance (as opposed to the sub-network modularity analysis that identifies integral sub-networks within a larger network, although similar patterns were evident with this modularity measure as well). This suggests all regions implicated in the semantic network played

a role in encoding negative feedback in non-stressful situations among men, possibly in a more hierarchical, top-down manner given the nature of the select network modularity variable. Importantly, men also did not associate this enhanced encoding of negative feedback with perceptions of the self in the math domain as there was no relationship between memory accuracy for negative feedback and math self-enhancement or valuing, and may have even performed better to the extent they encoded negative feedback via semantic memory network connectivity (Supplementary Results p. 11, Tables 17, 18). In conjunction with superior performance on the math feedback task in general, it appeared men were able to utilize this information in a manner that did not undermine performance, e.g. they did not suffer setbacks as a function of receiving potentially self-threatening information; in fact, it was quite the opposite.

It is unclear why women performing in theoretically stereotype neutral contexts exhibited more efficacious encoding of negative feedback that was not linked to either emotional or semantic memory encoding processes. In light of their suboptimal performance on the math feedback task (i.e. women overall performed worse on the math feedback task), but then superior performance on the standard math task, one possibility is that these women were experiencing elements of an initial, implicit stereotype threat inadvertently primed by our procedure or by

Table 8. Descriptive statistics for emotional network select network modularity to negative hits during the memory task

Variable	Condition	Gender	Mean	SD	N	
Semantic memory network sub-network modularity negative hits θ	DMT	Female	0.1019	0.0104	38	
		Male	0.0994	0.0121	28	
		Total	0.1008	0.0111	66	
	PST	Female	0.0974	0.0104	30	
		Male	0.1030	0.0091	24	
		Total	0.0999	0.0102	53	
	Total	Female	0.0999	0.0106	68	
		Male	0.1011	0.0109	52	
		Total	0.1004	0.0107	120	
	Semantic memory network sub-network modularity negative hits α	DMT	Female	0.1052	0.0102	38
			Male	0.1007	0.0122	28
			Total	0.1033	0.0113	66
PST		Female	0.1033	0.0117	30	
		Male	0.1025	0.0117	24	
		Total	0.1029	0.0116	53	
Total		Female	0.1044	0.0109	68	
		Male	0.1015	0.0119	52	
		Total	0.1031	0.0114	120	
Semantic memory network sub-network modularity negative hits β		DMT	Female	0.1023	0.0091	38
			Male	0.0993	0.0086	28
			Total	0.1010	0.0089	66
	PST	Female	0.1035	0.0113	30	
		Male	0.1016	0.0088	24	
		Total	0.1027	0.0102	53	
	Total	Female	0.1028	0.0101	68	
		Male	0.1004	0.0087	52	
		Total	0.1018	0.0095	120	
	Semantic memory network sub-network modularity negative hits γ	DMT	Female	0.1020	0.0117	38
			Male	0.1036	0.0111	28
			Total	0.1027	0.0114	66
PST		Female	0.1018	0.0115	30	
		Male	0.1058	0.0119	24	
		Total	0.1036	0.0118	53	
Total		Female	0.1019	0.0116	68	
		Male	0.1046	0.0114	52	
		Total	0.1031	0.0115	120	

receiving negative feedback after solving math problems in general, that was then nullified over time. Nevertheless, despite biased encoding of negative feedback and variable performance, women in stereotype neutral contexts did not associate negative feedback with perceptions of the self in the math domain or exhibit any links between encoding biases and performance. This provides yet another example of how placing women in more stereotype neutral contexts can have positive ramifications for their performance and perceptions in STEM domains as opposed to the status quo.

One concern associated with assessing memory for feedback received on a math task is that given that stereotype threat undermines performance, women in the DMT condition may be exposed to more negative feedback in general. This was not a concern in our study, however, for a number of reasons. One, d' scores are ultimately standardized in a specific effort to account for variability in the quantity of stimuli individuals evaluate. Two, the main effect for gender on the math feedback task indicated that all women were exposed to more negative feedback compared with men, however, only stereotype threatened women exhibited a link between encoding of negative feedback, emotional memory encoding processes, performance, math self-enhancing and math devaluing. Finally, all individuals encoded negative feedback more efficaciously, indicating that

despite women being presented with more negative feedback overall, this did not result in discrepant memory accuracy between the conditions in any way.

It should also be noted that while results were specific to the β band for emotional memory encoding and the γ band for semantic memory encoding, similar patterns were found across frequency bands with respect to these findings. That said, both β and γ frequencies are believed to play important roles in memory encoding processes in general (Benchenane et al., 2011). It is also always important with respect to EEG to stress caution in interpretation of results related to specific brain regions given the spatial limitations inherent in the methodology. Nevertheless, by using a high density electrode array, employing an advanced source Bayesian analytic approach including dSPM inverse operators, and confining analyses to regions closer to the cortical surface, it is possible to make accurate assumptions about contributions from specific brain regions (Cohen, 2014).

Findings from this study suggest that SBS may prompt women to leave STEM performance contexts with more than just a poor score. Indeed these contexts may bias women in STEM toward focusing on negative, stereotype confirming feedback in lieu of positive feedback and encoding this information via emotional memory encoding processes; a process that

typically engenders enduring, vivid memories about specific autobiographical moments in time. These emotional memories, in turn, have downstream ramifications on how women view themselves in the STEM domain and in relation to their STEM peers. Men may show a similar accuracy for negative feedback but do so through semantic memory processes typically employed in neutral, non-emotional contexts that have no effect on how they perceive themselves in STEM domains. Such findings suggest that the deleterious effects of negative group expectations extend beyond the performance to forge a more insidious foundation of negative memories and self-perceptions within stigmatized domains that may ultimately help explain why women opt out of STEM domains at disproportionate rates compared with men.

Supplementary data

Supplementary data are available at SCAN online.

Conflict of interest. None declared.

Acknowledgements

All aspects of this study and article were supported by National Science Foundation grant #1329281 awarded to C.E.F.

References

- Arbuckle, J.L., Users' Guide, A. (1997). Version 3.6. Chicago, IL: SmallWaters Corporation.
- Attal, Y., Schwartz, D. (2013). Assessment of subcortical source localization using deep brain activity imaging model with minimum norm operators: a MEG study. *PLoS One*, *8*(3), e59856.
- Aydore, S., Pantazis, D., Leahy, R.M. (2013). A note on the phase locking value and its properties. *Neuroimage*, *74*, 231–44.
- Bassett, D.S., Greenfield, D.L., Meyer-Lindenberg, A., Weinberger, D.R., Moore, S.W., Bullmore, E.T. (2010). Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Computational Biology*, *6*(4), e1000748.
- Benchenane, K., Tiesinga, P.H., Battaglia, F.P. (2011). Oscillations in the prefrontal cortex: a gateway to memory and attention. *Current Opinion in Neurobiology*, *21*(3), 475–85.
- Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society, Series B (Methodological)*, 289–300.
- Bertolero, M.A., Yeo, B.T., D'Esposito, M. (2015). The modular and integrative functional architecture of the human brain. *Proceedings of the National Academy of Sciences*, *112*(49), E6798–807.
- Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–96.
- Binder, J.R., Desai, R.H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–36.
- BioSemi, B. (2011). *BioSemi ActiveTwo*. [EEG System]. Amsterdam: BioSemi.
- Blascovich, J., Spencer, S.J., Quinn, D., Steele, C. (2001). African Americans and high blood pressure: the role of stereotype threat. *Psychological Science*, *12*(3), 225–9.
- Blondel, V.D., Guillaume, J., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.
- Blumenthal, T.D., Cuthbert, B.N., Filion, D.L., Hackley, S., Lipp, O.V., Van Boxtel, A. (2005). Committee report: guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*(1), 1–15.
- 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*(2), 247–55.
- Burianova, H., McIntosh, A.R., Grady, C.L. (2010). A common functional brain network for autobiographical, episodic, and semantic memory retrieval. *Neuroimage*, *49*(1), 865–74.
- Bush, G., Luu, P., Posner, M.I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215–22.
- 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*(7), 572–8.
- Canli, T., Zhao, Z., Brewer, J., Gabrieli, J.D., Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience*, *20*(19), RC99.
- Cohen, M.X. (2014). Fluctuations in oscillation frequency control spike timing and coordinate neural networks. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(27), 8988–98.
- Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *3*(4), 286–300.
- Dale, A.M., Fischl, B., Sereno, M.I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179–94.
- Dale, A.M., Liu, A.K., Fischl, B.R., et al. (2000). Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, *26*(1), 55–67.
- Deary, I.J., Strand, S., Smith, P., Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21.
- Delgado, A.R., Prieto, G. (2008). Stereotype threat as validity threat: the anxiety–sex–threat interaction. *Intelligence*, *36*(6), 635–40.
- Desikan, R.S., Ségonne, F., Fischl, B., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968–80.
- 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.
- Ewald, A., Aristei, S., Nolte, G., Rahman, R.A. (2012). Brain oscillations and functional connectivity during overt language production. *Frontiers in psychology*, *3*, 166.
- Félix, M.A., Wagner, A. (2008). Robustness and evolution: concepts, insights and challenges from a developmental model system. *Heredity*, *100*(2), 132–40.
- Flegal, K.E., Marín-Gutiérrez, A., Ragland, J.D., Ranganath, C. (2014). Brain mechanisms of successful recognition through retrieval of semantic context. *Journal of Cognitive Neuroscience*, *26*(8), 1694–704.
- Fletcher, P.C., Frith, C.D., Baker, S.C., Shallice, T., Frackowiak, R.S.J., Dolan, R.J. (1995). The mind's eye—precuneus activation in memory-related imagery. *Neuroimage*, *2*(3), 195–200.

- Forbes, C.E., Leitner, J.B. (2014). Stereotype threat engenders neural attentional bias toward negative feedback to undermine performance. *Biological Psychology*, *102*, 98–107.
- 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*(5), 740.
- Forbes, C.E., Leitner, J.B., Duran-Jordan, K., Magerman, A.B., Schmader, T., Allen, J.J. (2014). Spontaneous default mode network phase-locking moderates performance perceptions under stereotype threat. *Social Cognitive and Affective Neuroscience*, *10*(7), 994–1002.
- Forbes, C.E., Duran, K.A., Leitner, J.B., Magerman, A. (2015). Stereotype threatening contexts enhance encoding of negative feedback to engender underperformance and anxiety. *Social Cognition*, *33*(6), 605.
- 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.
- Gramfort, A., Luessi, M., Larson, E., et al. (2013). MEG and EEG data analysis with MNE-python. *Frontiers in Neuroscience*, *7*, 267.
- Gramfort, A., Luessi, M., Larson, E., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–60.
- Gramzow, R.H., Willard, G. (2006). Exaggerating current and past performance: motivated self-enhancement versus reconstructive memory. *Personality & Social Psychology Bulletin*, *32*(8), 1114–25.
- Hämäläinen, M.S., Sarvas, J. (1989). Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data. *IEEE Transactions on Biomedical Engineering*, *36*(2), 165–71.
- Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*(9), 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*(3), 289–93.
- Harrington, L., Liu, J.H. (2002). Self-enhancement and attitudes toward high achievers: A bicultural view of the independent and interdependent self. *Journal of Cross-Cultural Psychology*, *33*(1), 37–55.
- Hayes, A.F. (2013). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. *Journal of Educational Measurement*, *51*(3), 335–7.
- Herbert, C., Pauli, P., Herbert, B.M. (2011). Self-reference modulates the processing of emotional stimuli in the absence of explicit self-referential appraisal instructions. *Social Cognitive and Affective Neuroscience*, *6*(5), 653–61.
- Hu, L., Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.
- Inzlicht, M., Ben-Zeev, T. (2000). A threatening intellectual environment: why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, *11*(5), 365–71.
- 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.
- Jongerling, J., Laurenceau, J.P., Hamaker, E.L. (2015). A multilevel AR (1) model: allowing for inter-individual differences in trait-scores, inertia, and innovation variance. *Multivariate Behavioral Research*, *50*(3), 334–49.
- Keller, J., Dauenheimer, D. (2003). Stereotype threat in the classroom: dejection mediates the disrupting threat effect on women's math performance. *Personality & Social Psychology Bulletin*, *29*(3), 371–81.
- Kenny, D. (2015). *Measuring Model Fit*. Available: <http://davidakenny.net/cm/fit.htm>.
- Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *Neuroimage*, *54*(3), 2446–61.
- LaBar, K.S., Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*, *7*(1), 54–64.
- LaBar, K.S., Phelps, E.A. (1998). Arousal-mediated memory consolidation: role of the medial temporal lobe in humans. *Psychological Science*, *9*(6), 490–3.
- Lachaux, J., Rodriguez, E., Martinerie, J., Varela, F.J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, *8*(4), 194–208.
- Levine, L.J., Burgess, S.L. (1997). Beyond general arousal: effects of specific emotions on memory. *Social Cognition*, *15*(3), 157.
- Lin, F., Belliveau, J.W., Dale, A.M., Hämäläinen, M.S. (2006). Distributed current estimates using cortical orientation constraints. *Human Brain Mapping*, *27*(1), 1–13.
- Liu, M., Kuo, C., Chiu, A.W. (2011). Statistical threshold for non-linear granger causality in motor intention analysis. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 5036–39.
- Liu, M., Kuo, C., Chiu, A.W. (2013). Non-linear Granger causality and its frequency decomposition in decoding human upper limb movement intentions. *International Journal of Biomedical Engineering and Technology*, *34*, *12*(1), 1–25.
- Liu, M., Amey, R.C., Forbes, C.E. (2017). On the role of situational stressors in the disruption of global neural network stability during problem solving. *Journal of cognitive neuroscience*, *29*(12), 2037–53.
- Major, B., Schmader, T. (1998). Coping with stigma through psychological disengagement. In: *Prejudice*, 219–41.
- Mangels, J.A., Good, C., Whiteman, R.C., Maniscalco, B., Dweck, C.S. (2012). Emotion blocks the path to learning under stereotype threat. *Social Cognitive and Affective Neuroscience*, *7*(2), 230–41.
- Marx, D.M., Goff, P.A. (2005). Clearing the air: the effect of experimenter race on target's test performance and subjective experience. *British Journal of Social Psychology*, *44*(Pt 4), 645–57.
- Martin, A., Chao, L.L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, *11*(2), 194–201.
- Meunier, D., Achard, S., Morcom, A., Bullmore, E. (2009). Age-related changes in modular organization of human brain functional networks. *Neuroimage*, *44*(3), 715–23.
- Meunier, D., Lambiotte, R., Bullmore, E.T. (2010). Modular and hierarchically modular organization of brain networks. *Frontiers in Neuroscience*, *4*, 200.
- 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*(10), 879–85.
- Murty, V.P., Ritchey, M., Adcock, R.A., LaBar, K.S. (2010). fMRI studies of successful emotional memory encoding: a quantitative meta-analysis. *Neuropsychologia*, *48*(12), 3459–69.
- Nisbett, R.E., Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, *84*(3), 231.

- Nolan, H., Whelan, R., Reilly, R. (2010). FASTER: fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods*, *192*(1), 152–62.
- Nosek, B.A., Banaji, M.R., Greenwald, A.G. (2002). Math=male, me=female, therefore math≠me. *Journal of Personality and Social Psychology*, *83*(1), 44.
- Ochsner, K.N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General*, *129*(2), 242.
- Osborne, J.W. (2001). Testing stereotype threat: does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, *26*(3), 291–310.
- Osborne, J.W. (2006). Gender, stereotype and threat anxiety: Psychophysiological and cognitive evidence. *Journal of Research in Educational Psychology*, *8*, 109–38.
- Osborne, J.W. (2007). Linking stereotype threat and anxiety. *Educational Psychology*, *27*(1), 135–54.
- Patterson, K., Nestor, P.J., Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–87.
- Payne, J., Jackson, E., Ryan, L., Hoscheidt, S., Jacobs, J., Nadel, L. (2006). The impact of stress on neutral and emotional aspects of episodic memory. *Memory*, *14*(1), 1–16.
- Pelham, B.W., Swann, W.B. (1989). From self-conceptions to self-worth: on the sources and structure of global self-esteem. *Journal of Personality and Social Psychology*, *57*(4), 672.
- Pintzinger, N.M., Pfabigan, D.M., Pfau, L., Kryspin-Exner, I., Lamm, C. (2017). Temperament differentially influences early information processing in men and women: preliminary electrophysiological evidence of attentional biases in healthy individuals. *Biological psychology*, *122*, 69–79.
- Rubinov, M., Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*(3), 1059–69.
- Schmader, T., Johns, M., Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*(2), 336.
- Schmader, T., Forbes, C.E., Zhang, S., Mendes, W.B. (2009). A metacognitive perspective on the cognitive deficits experienced in intellectually threatening environments. *Personality & Social Psychology Bulletin*, *35*(5), 584–96.
- Simon, H.A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, *106*(6), 467–82.
- Singh, A.K., Phillips, S. (2010). Hierarchical control of false discovery rate for phase locking measures of EEG synchrony. *NeuroImage*, *50*(1), 40–7.
- Spencer, S.J., Steele, C.M., Quinn, D.M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28.
- Sporns, O. (2011). The human connectome: a complex network. *Annals of the New York Academy of Sciences*, *1224*(1), 109–25.
- Stanley, M.L., Dagenbach, D., Lyday, R.G., Burdette, J.H., Laurienti, P.J. (2014). Changes in global and regional modularity associated with increasing working memory load. *Frontiers in Human Neuroscience*, *8*, 954.
- Steele, C.M., Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797.
- Steinmetz, K.R.M., Addis, D.R., Kensinger, E.A. (2010). The effect of arousal on the emotional memory network depends on valence. *Neuroimage*, *53*(1), 318–24.
- Stevens, A.A., Tappon, S.C., Garg, A., Fair, D.A. (2012). Functional brain network modularity captures inter- and intra-individual variation in working memory capacity. *PLoS One*, *7*(1), e30468.
- Touryan, S.R., Marian, D.E., Shimamura, A.P. (2007). Effect of negative emotional pictures on associative memory for peripheral information. *Memory*, *15*(2), 154–66.
- Townsend, S.S.M., Major, B., Gangi, C.E., Mendes, W.B. (2011). From “In the Air” to “Under the Skin”: cortisol responses to social identity threat. *Personality & Social Psychology Bulletin*, *37*(2), 151–64.
- Vaish, A., Grossmann, T., Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological Bulletin*, *134*(3), 383.
- Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 159–77.
- Wilson, A., Ross, M. (2003). The identity function of autobiographical memory: Time is on our side. *Memory*, *11*(2), 137–49.
- Wu, Z., Huang, N.E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, *1*(1), 1–41.
- Wu, Z., Huang, N.E., Chen, X. (2009). The multi-dimensional ensemble empirical mode decomposition method. *Advances in Adaptive Data Analysis*, *1*(3), 339–72.
- Zald, D.H., Lee, J.T., Fluegel, K.W., Pardo, J.V. (1998). Aversive gustatory stimulation activates limbic circuits in humans. *Brain: A Journal of Neurology*, *121*(6), 1143–54.