Negative Overgeneralization is Associated with Pattern Completion in Peripubertal Youth

Dana L. McMakin[1,2,4], Adam Kimbler[1], Nicholas J. Tustison[3], Jeremy W. Pettit[2,4], Aaron T. Mattfeld[1,4,5]

[1]Cognitive Neuroscience Program, Department of Psychology, Florida International University, Miami, FL 33199 USA
[2]Clinical Science Program, Department of Psychology, Center for Children and Families, Florida International University, Miami, FL 33199 USA
[3]Department of Radiology and Medical Imaging, University of Virginia, Charlottesville, VA 22903 USA
[4]Center for Children and Families, Florida International University, Miami, FL 33199 USA

[5]Corresponding author:
Aaron T. Mattfeld, Ph.D.
Department of Psychology, Florida International University
11200 SW 8th Street, AHC4-462, Miami, FL 33199
email: amattfel@fiu.edu

Keywords: anxiety, adolescence, generalization, hippocampus, mPFC, fMRI, sleep

Word count:
Abstract – 149
Introduction, Results, Discussion – 4497
Methods and Materials – 2223
References – 54

ABSTRACT

Negative overgeneralization is a dimension of anxiety whereby responses to one aversive situation (e.g., severe weather) spread to others that share contextual features (e.g., breezy day). We aim to sharpen mechanistic understanding of negative overgeneralization. In peripuberty – a time when changes in neurodevelopment potentiate generalization of salient experiences – these mechanisms may shape behavior and contribute to emotional health. In an emotional mnemonic similarity task conducted with youth, negative, relative to neutral, scene images were generalized more frequently. Negative overgeneralization was related to both greater and more similar patterns of activation in the CA1 hippocampal subfield and medial prefrontal cortex (mPFC) for negative relative to neutral stimuli. At encoding, the amygdala increased functional coupling with CA1 and mPFC during negative items that were later generalized. Negative overgeneralization is rooted in mechanisms of modulation at encoding and pattern completion at retrieval. Targeting these mechanisms during peripuberty could positively shape emotional health.

INTRODUCTION

The ability to generalize information allows organisms to behave efficiently within a complex, ever-changing environment. However, when unconstrained, excessive generalization can lead to suboptimal behavior. For example, overgeneralization of negative information is a defining feature of anxiety disorders (Lissek et al., 2014) that leads to avoidance of contexts that weakly resemble a previous aversive experience. The transition from childhood to adolescence (“peripuberty”) is a neurodevelopmental window when neural networks associated with generalization (Bowman and Ziethmova, 2018) and emotional processing (Phelps and LeDoux, 2005) undergo dynamic change. At the same time, disorders of emotion, such as anxiety increase (Beesdo et al., 2009), putting youth at a higher risk for escalating mental health problems (e.g. depression) in later adolescence and adulthood (Pine et al., 1998). Therefore, characterizing neurobiological mechanisms of negative overgeneralization in peripuberty ultimately may help to explain rising rates of anxiety in this developmental window.

Negative overgeneralization is evident across contexts and conditions. Aversive classical conditioning shows that negative overgeneralization is characterized by a shallow response decay gradient, whereby conditioned responses are evoked by stimuli that are increasingly dissimilar from the conditioned stimulus (Cha et al., 2014; Dymond et al., 2015; Greenberg et al., 2013). While classical conditioning provides insights into the neurobiology of negative overgeneralization and anxiety disorders, it cannot account for the complex and multi-faceted nature of most experiences. Declarative or episodic-like memory paradigms approximate complex experiences, and researchers can use them to further elucidate mechanisms of negative overgeneralization, and refine etiological models of anxiety.

Distinct computational processes are critical to generalization of declarative memories. Pattern completion, a process by which partial or degraded cues reinstate previously stored representations, is associated with behavioral generalization (McClelland et al., 1995); while pattern separation, a process by which overlapping representations are made distinct, is associated with behavioral discrimination (Yassa and Stark, 2011). Subfields of the hippocampus differentially support these computational processes, with the dentate gyrus (DG) playing a disproportionate role in pattern separation; while the Cornu Ammonis (CA)3 subfield is important for pattern completion (Marr, 1971). In human fMRI studies, the DG and CA3 regions cannot be anatomically separated, thus they are combined to investigate neurobiological signals of pattern separation while activations in the CA1 subfield, which receives input from CA3, reflect pattern completion (Bakker et al., 2008).

Negative overgeneralization may arise from enhanced pattern completion or a deficit in pattern separation. Theoretical models have emphasized poor pattern separation and related behavioral discrimination as a likely mechanism of negative overgeneralization in adult clinical populations (Leal and Yassa, 2018). A recent study induced “anxiety” in healthy adults and observed enhanced generalization of neutral stimuli when learning occurred under threat (Starita et al., 2019). However, there are no available neurobiological data in anxious clinical populations to directly examine neural correlates of pattern separation, and behavioral effects have only been examined using neutral object stimuli in conditions of threat and safety leaving unanswered the question concerning what information processing operations support negative overgeneralization, and particularly in peripuberty when relevant systems are in flux.

Rather than a deficit in pattern separation, negative overgeneralization may be governed by enhanced pattern completion, especially during peripuberty. The DG/CA3 exhibits a protracted developmental trajectory in comparison to other subfields of the hippocampus (e.g., CA1) (Lavenex and Banta Lavenex, 2013). Commensurate with this protracted development, studies in youth in the peripubertal age range of 9 to 11 years have shown reduced memory specificity that is correlated with both DG/CA3 volume (Lee et al., 2014) and multivariate measures of hippocampal volume (Keresztes et al., 2017). Thus, we hypothesize that maturational changes in the hippocampus contribute to weak pattern separation signals leaving pattern completion to drive neurobiological mechanisms of negative overgeneralization during this developmental window.

Generalization can also arise through modulation of target regions, like the hippocampus, by neural regions involved in mnemonic control (e.g., mPFC) and emotional processing (e.g., amygdala). The mPFC can be split into regions that potentiate (dorsal mPFC; dmPFC) and inhibit (ventral mPFC; vmPFC) fear memory (Giustino & Maren, 2015; Spalding, 2018). Both have been implicated in negative generalization during fear conditioning (e.g., Antoniadis and McDonald, 2006; Cullen et al., 2015; Dymond et al., 2015; Onat and Büchel, 2015). Amygdala-based salience detection has been shown to enhance the creation of long lasting memories (McGaugh, 2013). The amygdala, specifically the basolateral nucleus, shares direct projections to both the CA1 and the mPFC (Bacon et al., 1996). Increases in generalization are reliably identified following negative (Schechtman et al., 2010), aversive (Resnik et al., 2011), or threatening (Starita et al., 2019) events. Moreover, lesions of the amygdala result in selective deficits of gist but not detailed memories (Adolphs et al., 2005). This modulatory influence of emotional arousal and related amygdala activation is particularly noteworthy in peripubertal anxiety because youth with anxiety show increased activation of the amygdala in response to threatening or ambiguously threatening cues, peripuberty is associated with maturational changes in amygdala response to emotional stimuli (Forbes et al., 2011), and escalating trajectories of anxiety symptoms and emotional disorders are prevalent during this developmental window (Dahl and Gunnar, 2009).

To investigate whether mechanisms of negative overgeneralization in peripubertal youth arise as a product of impaired pattern separation or enhanced pattern completion we used an emotional mnemonic similarity task (Leal and Yassa, 2014). The study consisted of three phases. The first phase (Intake) comprised a diagnostic intake where study eligibility was determined. During the second phase (Study) in the MRI scanner, participants performed an incidental learning task during which they rated each stimulus (2 s; 2-6 s inter-stimulus-interval [ISI]) as negative, neutral, or positive. Approximately twelve hours later, participants returned to the scanner for a surprise memory test (third phase, Test). Critically, in addition to target and foil images, similar but not identical lure stimuli were presented at Test (2 s; separated by 2-6 s ISI). Participants were instructed to only endorse images as 'old' if the image at Test was *exactly* the same as the image presented at Study. Generalization, both behaviorally and neurobiologically, was assessed through responses to similar lure stimuli. We expected to observe an increased likelihood to false alarm to negative relative to neutral lures. Consistent with a pattern completion signal, we predicted that at Test the CA1 and mPFC would exhibit greater activations for negative compared to neutral lures that were false alarmed. Lastly, we hypothesized that functional coupling with the amygdala at Study would be elevated for items that were subsequently false alarmed.

This study was drawn from a larger project focused on how sleep-related memory consolidation impacts negative overgeneralization in youth across a spectrum of anxious symptoms. Therefore, participants (N = 34; 11.4 +- 2.0 years) were recruited from both clinical and community samples to maximize variability in negative overgeneralization, our construct of interest. Eligible participants were randomized to either a sleep (N = 16) or wake (N = 18) condition. In the current report, we have collapsed the sleep and wake conditions into a single group to maximize power and focus our investigation on general mechanisms of negative overgeneralization. Future reports will detail the specific impact of sleep on these mechanisms, but are outside of the scope of this study.

Here, we aim to deepen mechanistic understanding of negative overgeneralization by using a task that can disambiguate two distinct neural contributions (pattern separation, pattern completion) to the same behavioral phenotype (negative overgeneralization). Ultimately, the goal is to identify targets for intervention in order to offset developmental risk trajectories during the sensitive period of peripuberty. For example, approaches that target consolidation (e.g., targeted memory reactivation) can be employed to alter negative overgeneralization reliant on pattern completion, while interventions such as physical activity can upregulate hippocampal neurogenesis (van Praag et al., 2005) and improve negative overgeneralization by increasing discrimination through enhanced pattern separation.

RESULTS

Increased false alarms to negative lures in peripubertal youth

We collected 'old'/'new' recognition memory performance following an emotional mnemonic similarity task (**Fig. 1a**). Participants were instructed to endorse images as 'old' only if they were *exactly* the same as the images seen during the Study phase During the Test phase of the experiment, to examine negative overgeneralization we compared the lure generalization index (LGI: p('Old'|Lure) – p('Old'|Foil)) – the likelihood of participants to false alarm to lures (i.e., call lures 'Old') corrected by their tendency to identify Foils as 'Old' – between negative, neutral, and positive stimuli. The same analyses were performed using the lure discrimination index to assess changes in discrimination (LDI: p('New'|Lure) – p('New'|Target)) – the likelihood of participants to correctly reject lures (i.e., call lures 'New') corrected by their tendency to call Targets 'New.' Generalization was significantly greater for negative compared to both neutral ($t(33) = 5.07$, $P = 1.4\text{x}10^{-5}$) and positive ($\text{t}(33) = 6.44$, $P = 2.6\text{x}10^{-7}$) stimuli, which were

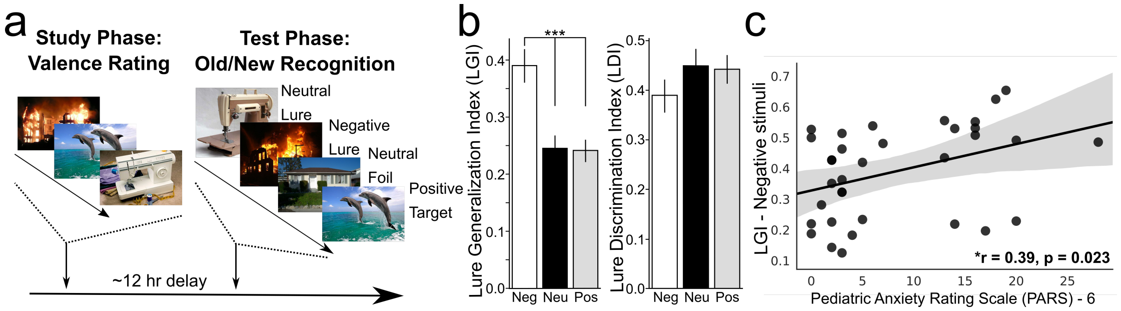


Figure 1 Timeline of experimental procedures and lure behavioral performance. (**a**) The scanning portion of the experiment consisted of two phases: 1) A Study phase during which participants rated the valence of each image. Images were presented for 4 s and were separated by a jittered inter-stimulus-interval (ISI: white fixation cross; 2-6 s). 2) During the Test phase participants performed an old/new recognition memory task. Target (images from the Study phase), foil (new pictures), and lure (new images that were similar but not exactly the same as images from the Study phase) stimuli were presented for 2 s and separated by a jittered ISI (2-6 s). (**b**) The Lure Generalization Index (LGI): p('Old'|Lure) – p('Old'|Foil) – was significantly greater for negative compared to neutral ($t(33) = 5.07$, $P = 1.4 \times 10^{-5}$) and positive ($t(33) = 6.44$, $P = 2.6 \times 10^{-7}$) stimuli. A similar difference was not observed (all $P > .10$) for the Lure Discrimination Index (LDI): p('New'|Lure) – p('New'|Target). (**c**) Negative generalization (Negative LGI) was positively associated with anxiety severity (PARS-6) (r = 0.39, p = 0.023). Neg = Negative; Neu = Neutral; Pos = Positive; *** $P < 0.0001$.

not significantly different from one another ($t(33) = -.16$, $P = .87$) (**Fig. 1b**). No significant differences in discrimination were evident when comparing the different stimulus valences ($\text{LDI}_{\text{Neg v. Neu}}$: $t(33) = -1.56$, $P = .12$; $\text{LDI}_{\text{Neg v. Pos}}$: $t(33) = -1.66$, $P = .106$; $\text{LDI}_{\text{Pos v. Neu}}$: $t(33) = -.19$, $P = .85$). To evaluate the relationship between negative overgeneralization and anxiety we correlated Negative LGI with the Pediatric Anxiety Rating Scale (PARS-6). Participants with more symptoms of anxiety were more likely to generalize negative information ($r = 0.39$, $P = 0.023$) (**Fig. 1c**). These results demonstrate that generalization was elevated only for negative stimuli and was related to anxiety severity.

MRI results

Negative generalization is associated with greater activation in the mPFC and CA1

To investigate the neurobiological mechanisms supporting negative overgeneralization, we next examined differences in activations for negative relative to

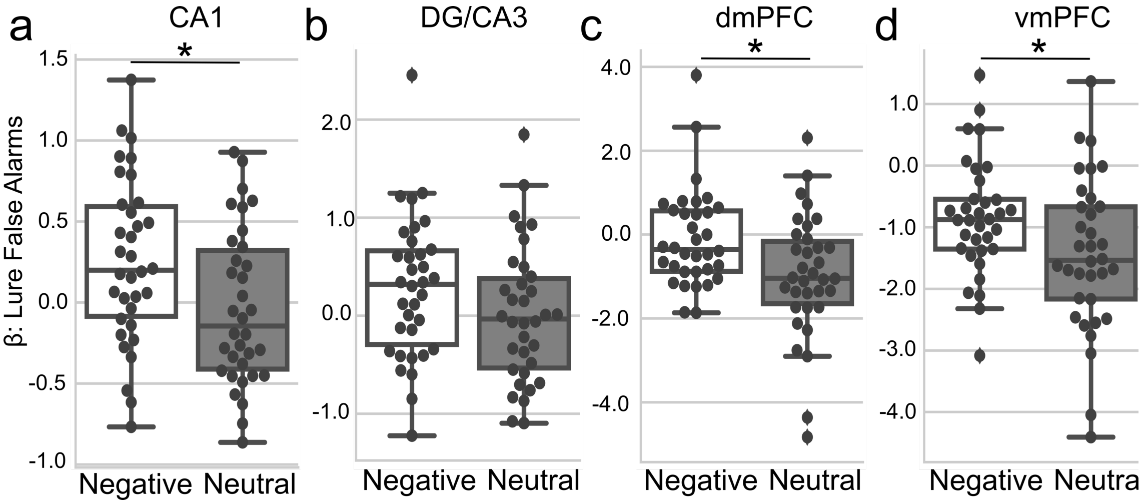


Figure 2 Anatomical region-of-interest analysis showing hippocampal subfield (CA1 and DG/CA3) and both dorsal medial prefrontal cortex (dmPFC) and ventral medial prefrontal cortex (vmPFC) activations. (**a**) The CA1 hippocampal subfield (*P* = 0.001) exhibited greater activations for negative lure false alarms relative to neutral lure false alarms. (**b**) The DG/CA3 subfields (*P* = 0.03) exhibited a trend towards greater activations for negative lure false alarms relative to neutral lure false alarms. (**c,d**) Both the dmPFC (*P* = 0.009) and the vmPFC (*P* = 0.002) showed greater activations for negative compared to neutral lure false alarms. *Significant following Bonferroni correction for multiple comparisons.

neutral lure stimuli that were incorrectly identified as old (false alarmed) in anatomically defined *a priori* regions of interest (ROIs). Our primary hypotheses concerned the contributions of the bilateral CA1 as an index of pattern completion, as well as the dmPFC as an index of potentiation of negative information. We predicted that each would exhibit greater activation for negative relative to neutral false alarms, reflecting negative overgeneralization. Alternatively, negative overgeneralization may result from a failure to pattern separate reflected by a reduction in DG/CA3 activations for negative relative to neutral false alarms or a failure to inhibit negative information associated with a similar reduction in vmPFC activation for negative relative to neutral false alarms. We found greater activation in the CA1 for negative compared to neutral lures that were false alarmed (*t*(33) = 3.51 *P* = 0.001) (**Fig. 2a**). Activations between negative and neutral lure false alarms in the DG/CA3 did not significantly differ following corrections for multiple comparisons but exhibited a trend towards greater activation for negative relative to neutral false alarms (*t*(33) = 2.29, *P* = 0.03), in the opposite direction predicted based on a failure to pattern separate (**Fig. 2b**). Both the dmPFC (*t*(33) = 2.74, *P* = 0.009) and vmPFC (*t*(33) = 3.24, *P* = 0.002) exhibited greater activation for negative when compared to neutral lures that were false alarmed (**Fig. 2c,d**).

To further evaluate whether negative overgeneralization was a result of impaired discrimination resulting from poor pattern separation we compared activations in the same regions for negative and neutral lure stimuli that were correctly rejected (identified as 'New'). We predicted reduced activations to negative lures relative to neutral lures that were correctly rejected in the DG/CA3. Amongst our *a priori* ROIs no regions exhibited a significant difference between negative and neutral lures that were correctly rejected following corrections for multiple comparisons (CA1: *t*(33) = 2.32, *P* = 0.03; DG/CA3: *t*(33) = 2.50, *P* = 0.02; vmPFC: *t*(33) = 0.29, *P* = 0.76; dmPFC: *t*(33) = 0.39, *P* = 0.69) (**Supplementary Fig. 1**). In fact, similar to false alarms in the DG/CA3, there was a trend towards greater activation in the CA1 and DG/CA3 for negative relative to neutral lures that were correctly rejected – opposite the predicted difference based on an assumption of impaired pattern separation. These results support our view that negative overgeneralization in peripubertal youth is in part supported by heightened activation in the CA1, reflecting increased pattern completion. Moreover, both the dmPFC and the vmPFC exhibited signals in line with negative overgeneralization.

A subcortical and cortical network contribute to negative overgeneralization

To evaluate the contribution of other brain areas to negative overgeneralization outside our *a priori* anatomical ROIs, we performed an exploratory whole-brain analysis at the voxel-wise level. We used FSL's *Randomise* threshold free cluster enhancement (tfce) multiple comparisons correction approach to perform one-sample t-tests with a corrected threshold of *P* < 0.01. We compared negative lure false alarms to neutral lure false alarms. Similar to our anatomical region of interest approach, clusters in the hippocampus survived corrections for multiple comparisons. We also observed clusters in the bilateral amygdala and midline thalamus. Clusters in the bilateral mPFC, posterior cingulate cortex extending into the precuneus, inferior parietal lobule, and middle temporal gyrus were observed (**Fig. 3**). When evaluating whether neutral lure false alarms exhibited greater activations compared to negative lure false alarms at the voxel-wise level, no regions survived corrections for multiple comparison. The exploratory whole brain results suggest that negative generalization relies on a broad cortical/subcortical network.

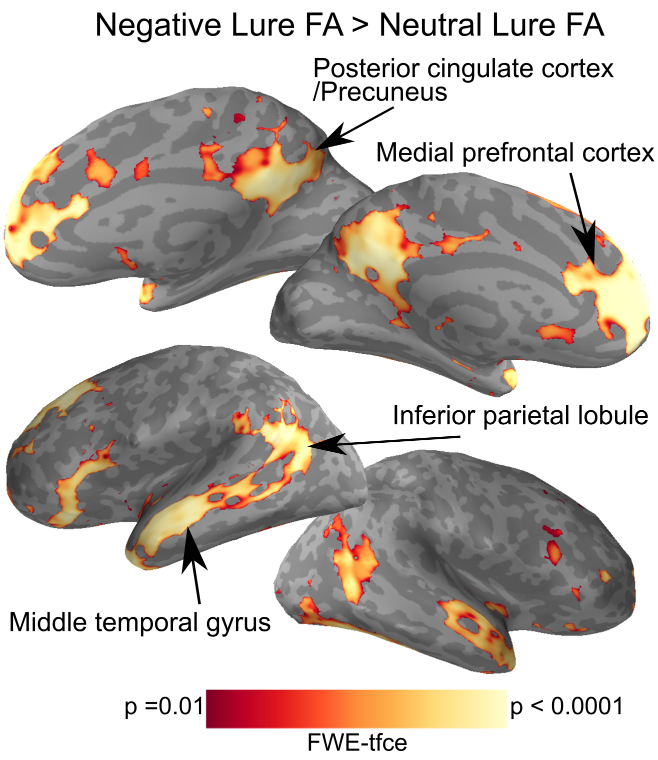


Figure 3 Whole brain exploratory analysis showing regions that exhibited greater activations for negative relative to neutral lure false alarms (FA) in the posterior cingulate cortex extending into the precuneus, medial prefrontal cortex, inferior parietal lobule, and middle temporal gyrus. FWE-tfce corrected *P* < 0.01.

Functional correlations between the amygdala ↔ CA1 and amygdala ↔ dmPFC enhanced during initial encoding of negative stimuli that were subsequently false alarmed

We hypothesized that the amygdala plays an important role in negative overgeneralization during the initial encoding of stimuli. Specifically, we predicted that during the Study phase the amygdala would exhibit enhanced functional coupling with the CA1 and the mPFC for negative stimuli that were subsequently false alarmed. To assess the functional correlations between amygdala ↔ CA1, amygdala ↔ dmPFC, and

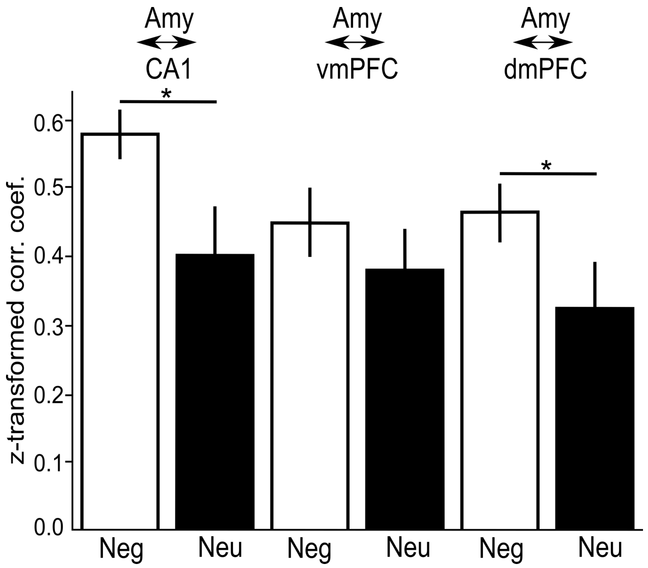


Figure 4 Task based functional connectivity during the Study phase between the amygdala (Amy) and three target regions: CA1, ventral medial prefrontal cortex (vmPFC), and dorsal medial prefrontal cortex (dmPFC) for items that were subsequently replaced by lures and false alarmed. Increased functional connectivity was observed between the Amy ↔ CA1 (*P* = 0.007) and the Amy ↔ dmPFC (*P* = 0.013) for negative (Neg) relative to neutral (Neu) stimuli that were replaced by similar lure stimuli and subsequently false alarmed. *Significant following Bonferroni correction for multiple comparisons.

amygdala ↔ vmPFC we performed a task-based beta series correlation analysis (Rissman et al., 2004). Using anatomical regions of interest, we correlated the beta-series for negative and neutral stimuli that during the Test phase were replaced with a similar lure and subsequently false alarmed. We observed enhanced functional coupling between the amygdala and both the CA1 (amygdala ↔ CA1: *t*(33) = 2.9, *P* = 0.007) and the dmPFC (amygdala ↔ dmPFC: *t*(33) = 2.6, *P* = 0.013) for negative stimuli compared to neutral stimuli that were subsequently false alarmed but only a trend for the amygdala and the vmPFC (amygdala ↔ vmPFC: *t*(33) = 1.6, *P* = 0.13) (**Fig. 4**). No significant difference in functional connectivity between the same regions was observed when comparing negative and neutral stimuli that were subsequently correctly rejected (amygdala ↔ CA1: *t*(33) = 1.03, *P* = 0.32; amygdala ↔ dmPFC: *t*(33) = -0.46, *P* = 0.64; amygdala ↔ vmPFC: *t*(33) = -0.76, *P* = 0.45) (**Supplementary Fig. 2**). These results demonstrate that negative overgeneralization may in part be dependent on modulatory influences of the amygdala on target structures at encoding.

Increased representational similarity in the mPFC and left CA1 associated with enhanced negative overgeneralization

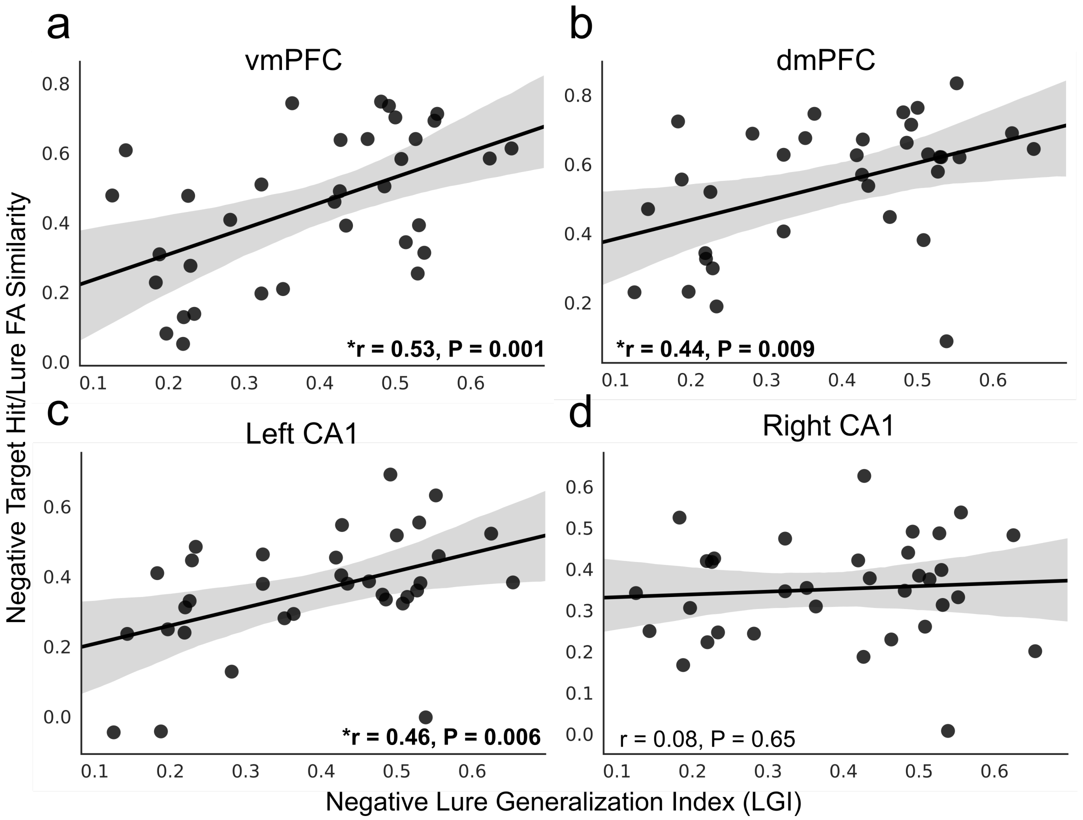


Figure 5. Correlations between negative target hit and lure false alarm (FA) similarity in the ventral medial prefrontal cortex (vmPFC), dorsal medial prefrontal cortex (dmPFC), and bilateral CA1 and LGI behavioral performance. (**a,b**) Both the vmPFC (*P* = 0.001) and the dmPFC (*P* = 0.008) exhibited a significant correlation between negative target hit/lure FA similarity and negative LGI performance. As the similarity between target hit and lure FA activations increased (i.e., became more similar) negative LGI performance increased. (**c,d**) A similar correlation was identified in the left CA1 (*P* = 0.006) but not the right CA1 (*P* = 0.65). *Significant following Bonferroni correction for multiple comparisons.

We also employed representational similarity analyses to elucidate further neurobiological mechanisms of negative overgeneralization in the CA1, dmPFC, and vmPFC. We hypothesized that negative overgeneralization is strongest when the representations of novel stimuli come to resemble the representations of previously encountered items. Accordingly, patterns of activation evoked by lure stimuli that were false alarmed should share similarity with patterns of activation evoked by target hits. The degree to which these patterns overlap should relate to overgeneralization during the task. Thus, to evaluate the relationship between patterns of brain activation and our behavioral indices of generalization (i.e., LGI) we examined the correlation between target hit and lure false alarm similarity and LGI scores. Negative LGI increased as both vmPFC (*r* = 0.53, *P* = 0.001) and dmPFC (*r* = 0.44, *P* = 0.009) similarity between negative target hits and negative lure false alarms increased (i.e., became more similar) (**Fig. 5a,b**). A similar pattern was observed in the left CA1 (*r* = 0.46, *P* = 0.006) (**Fig. 5c**) – negative LGI increased as similarity in the left CA1 increased – but not in the right CA1 (*r* = 0.08, *P* = 0.65) (**Fig. 5d**).

Both lure false alarms and target hits were endorsed with an 'old' response, thus the similarity in activations are confounded by the similar response to the two conditions. To evaluate this potential confound and the specificity of increased pattern similarity as a mechanism of negative overgeneralization we performed the same analyses using both positive and neutral stimuli. No significant relationship between pattern similarity and behavior, despite the same response type, was observed for positive stimuli (left CA1: r = -0.015, p = 0.93; right CA1: r = -0.16, p = 0.35; vmPFC: r = 0.09, p = 0.59; dmPFC: r = 0.19, p = 0.27) (**Supplementary Fig. 3**) and only a nonsignificant trend for neutral stimuli was observed (left CA1: r = 0.40, p = 0.018; right CA1: r = 0.30, p = 0.08; vmPFC: r = 0.29, p = 0.09; dmPFC: r = 0.13, p = 0.46; **Supplementary Fig. 4**). These results suggest that the relation between brain activation and behavior is not likely a result of similar responses to the two events. Rather, participants who exhibited highly dissimilar patterns of activation between target hits and lure false alarms in the dmPFC, vmPFC, and left CA1 for negative stimuli, exhibited low levels of negative overgeneralization. In contrast, participants who exhibited similar patterns of activation for negative target hits and negative lure false alarms, exhibited increased overgeneralization of negative information.

Relationship between anxiety severity and brain measures

Lastly, we wanted to explore the relationship between the identified neurobiological mechanisms of negative overgeneralization and anxiety symptom severity. These results are exploratory and should be interpreted with caution. No significant relationship between severity of anxious symptoms (e.g., PARS-6) and univariate activations amongst our *a priori* ROIs was evident (CA1: r = 0.04, p = 0.84; vmPFC: r = 0.18, p = 0.32; dmPFC: r = -0.03, p = 0.87). Functional connectivity during the Study phase between the amygdala ↔ CA1 and the amygdala ↔ dmPFC for negative items that were subsequently false alarmed exhibited a trend, whereby greater connectivity was associated with more severe symptoms of anxiety ($r_{AMY\leftrightarrow CA1}$ = 0.29, p = 0.09; $r_{AMY\leftrightarrow dmPFC}$ = 0.35, p = 0.04) (**Supplementary Fig. 5a**). A similar trend in our multivariate measure was noted, wherein representational similarity in the left CA1 between negative target hits and negative lure false alarms was elevated for participants with greater levels of anxiety (r = 0.32, p = 0.06) (**Supplementary Fig. 5b**). This relationship was not evident in the vmPFC (r = 0.02, p = 0.85) or the dmPFC (r = 0.44, p = 0.80). These results suggest the neurobiological mechanisms of negative overgeneralization may represent an etiological feature of anxiety.

DISCUSSION

We aimed to elucidate the neurobiology of negative overgeneralization. Behavioral and neurobiological data from an emotional mnemonic similarity task converged to support our theory that negative overgeneralization during peripuberty results from enhanced pattern completion. Specifically, results indicated greater behavioral generalization (e.g., LGI) and greater neural activation in CA1 and mPFC when generalizing negative relative to neutral lures. Neurobiological mechanisms of negative overgeneralization are modulated by interactions between the amygdala with target regions during encoding, as evidenced by greater amygdala ↔ CA1 and amygdala ↔ dmPFC connectivity during the Study phase for negative (relative to neutral) stimuli that were subsequently generalized. Behaviorally, greater negative overgeneralization was associated with higher anxiety severity, however, only a trend with neurobiological mechanisms was identified. Representational similarity analyses also showed that more similar neural activation patterns in left CA1 and mPFC when participants accurately recalled a stimulus versus when they inaccurately generalized a lure, were associated with greater behavioral generalization for negative stimuli. Together, these results outline candidate neurobiological mechanisms underlying negative overgeneralization. Contributions from these neural circuits and increasing symptoms of anxiety during peripuberty highlight the susceptibility and plasticity of these mechanisms during this developmental window providing a substrate from which anxiety and related disorders could emerge or escalate.

Behavioral results showing that a corrected measure of performance evidenced greater generalization for negative stimuli, relative to neutral, were especially compelling in light of tests of alternative hypotheses. Namely, we did not observe evidence supporting impairment in discrimination. Also, similar tests of positively-valenced stimuli did not reveal generalization effects. Further, LGI for negative stimuli was positively associated with anxious symptoms.

Neurobiological findings paralleled these behavioral results by demonstrating activation in regions that were identified *a priori* as being associated with the computational process of pattern completion. Prior studies have consistently identified the hippocampus as an important contributor to negative overgeneralization (Antoniadis and McDonald, 2006; Cullen et al., 2015; Lissek et al., 2014; Onat and Büchel, 2015), however the contributions of hippocampal subfields have remained underspecified. Using high resolution imaging, we observed greater activation in the CA1 for negative versus neutral lure false alarms – consistent with generalization through pattern completion. In contrast, the DG/CA3, a subfield of the hippocampus that contributes to successful discrimination through pattern separation (Yassa and Stark, 2011; Marr, 1971), exhibited a trend towards greater activation for negative compared to neutral lure false alarms and correct rejections – a pattern of directionality that is inconsistent with predictions based on impairments in pattern separation.

We also found support for involvement of mPFC in negative overgeneralization. However, both the dmPFC, as well as the vmPFC exhibited greater activation for negative compared to neutral lure false alarms suggesting that the functional distinction between these regions identified during aversive conditioning paradigms (Baldi and Bucherelli, 2015; Dymond et al., 2015; Giustino and Maren, 2015; Jasnow et al., 2017; Lissek et al., 2014; Onat and Büchel, 2015; Spalding, 2018) may not hold for more complex tasks and/or this developmental period. Similar conclusions have been drawn in rodent models using context fear conditioning paradigms (Antoniadis and McDonald, 2006; Baldi and Bucherelli, 2015; Cullen et al., 2015; Giustino and Maren, 2015; Jasnow et al., 2017). Negative overgeneralization was not limited to the CA1 and mPFC, rather a broad network including regions like the amygdala, posterior cingulate cortex/precuneus, and midline thalamus also exhibited greater activations for negative relative to neutral lure stimuli that were false alarmed lending support to the hypothesis that a broad network governs negative overgeneralization.

Taken together, these behavioral and neurobiological findings stand in contrast to existing literature pointing to failures in discrimination and related impairment in pattern separation as an underlying mechanism of negative overgeneralization. However, these studies have been conducted primarily with rodent models, healthy adults, aged adults, or adults with subclinical depressive symptoms (see Leal and Yassa, 2018 for a Review). Our focus on peripubertal youth and sampled across a full range of anxious symptoms to maximize variability in the construct of interest (i.e. negative overgeneralization), likely account for these differences.

The 'generalization' interpretation of CA1 activation in our study is in line with computational models that have proposed CA1 activity may generalize following partial or incomplete pattern completion in the CA3 (Rolls, 2013). It is important to consider that another line of research has proposed that the CA1 serves as a comparator between retrieved and current representations (Lisman and Grace, 2005). Studies in young adults have observed increased activation for recognition probes that mismatched the original stimuli (Chen et al., 2011; Duncan et al., 2012). In this capacity, CA1 activation may serve as a novelty signal, important for the integration of new information across episodes. These studies reported activation differences for correct rejections while in the current study we did not observe differences in activation for lure stimuli that were correctly rejected rather for stimuli that were false alarmed. Therefore, greater activation in the CA1 for false alarmed stimuli likely reflect a pattern completion signal.

During the Study phase of our experiment we observed increased functional correlations between the amygdala and both the CA1 and dmPFC for negative relative to neutral items that were subsequently replaced by a lure and false alarmed. These regions are important for generalization and fear potentiation respectively (Schafe et al., 2005; Spalding, 2018). Greater functional coupling for similar stimuli that were correctly rejected was not observed, discounting the possibility that the amygdala is more correlated with negative stimuli regardless of behavioral performance. These results support our theory that the amygdala at encoding biases neurobiological mechanisms towards generalization, consistent with prior studies of emotions preferentially enhancing gist rather than detailed memories (Mather and Sutherland, 2011).

As a complementary analytic approach, we used representational similarity analyses to investigate the contributions of the CA1 and mPFC. We found support for our hypothesis that negative overgeneralization is strongest when the representations of novel lure stimuli resemble the representations of previously encountered items.

Consistent with this prediction, negative LGI was greater among individuals with highly similar patterns of activation between negative target hits and lure false alarms. These results are important given that both the hippocampus and mPFC play a role in memory integration and concept formation (Bowman and Zeithamova, 2018; Mack et al., 2018; Schlichting and Preston, 2015; van Kesteren et al., 2012). Thus, the relation between the convergent patterns of brain activation and negative overgeneralization behaviorally may reflect mechanisms of pattern completion and memory integration respectively.

Finally, our exploratory analyses of associations between neurobiological mechanisms of negative overgeneralization and anxiety severity evidenced either weak or absent effects. We believe there could be at least two explanations for the observed results. First, anxiety disorders are highly heterogeneous, and although negative overgeneralization is one common symptom dimension, it is not directly assessed in the anxiety severity measure (PARS). Therefore, the summary score of anxiety severity may be diffuse, contributing to weak associations with neurobiological mechanisms of negative overgeneralization. Additionally, the sample size is likely underpowered to support multiple interactions across response systems. Future work could use measures specifically designed to assess the symptom dimension of negative overgeneralization in a larger sample in order to clarify symptom-mechanism associations.

Despite these limitations, this study also included attributes that lend unique support to theoretical frameworks, and can propel future research. High-resolution fMRI allows for isolation of activation from small regions that have previously been identified as reflecting core computational mechanisms of interest (e.g., pattern completion), while sampling across a full range of anxious symptoms in youth allows for maximal variability in the dimension of interest. Moreover, multiple convergent behavioral and neurobiological analyses probed the mechanisms of negative overgeneralization using a well-validated task, and identified unique targets for further investigation and intervention.

These data support a neurobiological mechanism underlying negative overgeneralization. Interactions between the amygdala and target regions during initial encoding contribute to the formation of gist like representations. At Test, both enhanced univariate as well as convergence of multivariate patterns in activation in both the mPFC and CA1 reflect enhanced pattern completion. In this way interactions between the amygdala and mPFC and CA1 at encoding determine their contribution to negative overgeneralization at retrieval. These findings emphasize the importance of considering developmental stages when defining mechanisms of negative overgeneralization, and can further detail working models of anxiety pathogenesis during the transition from childhood to adolescence. Moreover, the finding that pattern completion underlies negative overgeneralization in peripubertal youth across a range of anxious symptoms provides a defined target for measuring treatment outcomes and for novel treatment development.

ACKNOWLEDGEMENTS

We would like to thank the team at the Center for Imaging Sciences at Florida International University, the EMU project staff, and participant families, for their assistance in collecting these data. We thank Drs. Timothy Allen and Michael Yassa for their useful feedback on the project and manuscript. The research was funded in part by an R01 from the NIMH to DLM and ATM (MH116005).

REFERENCES

Adolphs, R., Tranel, D., & Buchanan, T. W. (2005). Amygdala damage impairs emotional memory for gist but not details of complex stimuli. *Nature Neuroscience*, *8*(4), 512–518. https://doi.org/10.1038/nn1413

Antoniadis, E. A., & McDonald, R. J. (2006). Fornix, medial prefrontal cortex, nucleus accumbens, and mediodorsal thalamic nucleus: Roles in a fear-based context discrimination task. *Neurobiology of Learning and Memory*, *85*(1), 71–85. https://doi.org/10.1016/j.nlm.2005.08.011

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Avants, Brian B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044. https://doi.org/10.1016/j.neuroimage.2010.09.025

Avants, Brian B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., & Gee, J. C. (2010). The optimal template effect in hippocampus studies of diseased populations. *NeuroImage*, *49*(3), 2457–2466. https://doi.org/10.1016/j.neuroimage.2009.09.062

Bacon, S. J., Headlam, A. J., Gabbott, P. L., & Smith, A. D. (1996). Amygdala input to medial prefrontal cortex (mPFC) in the rat: A light and electron microscope study. *Brain Research*, *720*(1–2), 211–219. https://doi.org/10.1016/0006-8993(96)00155-2

Baldi, E., & Bucherelli, C. (2015). Brain sites involved in fear memory reconsolidation and extinction of rodents. *Neuroscience & Biobehavioral Reviews*, *53*, 160–190. https://doi.org/10.1016/j.neubiorev.2015.04.003

Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and Anxiety Disorders in Children and Adolescents: Developmental Issues and Implications for DSM-V. *The Psychiatric Clinics of North America*, *32*(3), 483–524. https://doi.org/10.1016/j.psc.2009.06.002

Bowman, C. R., & Zeithamova, D. (2018). Abstract Memory Representations in the Ventromedial Prefrontal Cortex and Hippocampus Support Concept Generalization. *The Journal of Neuroscience*, *38*(10), 2605–2614. https://doi.org/10.1523/JNEUROSCI.2811-17.2018

Brown, E. S., Kulikova, A., Van Enkevort, E., Nakamura, A., Ivleva, E. I., Tustison, N. J., … Malone, K. (2019). A randomized trial of an NMDA receptor antagonist for reversing corticosteroid effects on the human hippocampus. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*. https://doi.org/10.1038/s41386-019-0430-8

Cha, J., Greenberg, T., Carlson, J. M., DeDora, D. J., Hajcak, G., & Mujica-Parodi, L. R. (2014). Circuit-Wide Structural and Functional Measures Predict Ventromedial Prefrontal Cortex Fear Generalization: Implications for Generalized Anxiety Disorder. *Journal of Neuroscience*, *34*(11), 4043–4053. https://doi.org/10.1523/JNEUROSCI.3372-13.2014

Chen, J., Olsen, R. K., Preston, A. R., Glover, G. H., & Wagner, A. D. (2011). Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. *Learning & Memory*, *18*(8), 523–528. https://doi.org/10.1101/lm.2135211

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*(3), 162–173.

Cullen, P. K., Gilman, T. L., Winiecki, P., Riccio, D. C., & Jasnow, A. M. (2015). Activity of the anterior cingulate cortex and ventral hippocampus underlie increases in contextual fear generalization. *Neurobiology of Learning and Memory*, *124*, 19–27. https://doi.org/10.1016/j.nlm.2015.07.001

Dahl, R. E., & Gunnar, M. R. (2009). Heightened stress responsiveness and emotional reactivity during pubertal maturation: Implications for psychopathology. *Development and Psychopathology*, *21*(1), 1–6. https://doi.org/10.1017/S0954579409000017

Duncan, K., Ketz, N., Inati, S. J., & Davachi, L. (2012). Evidence for area CA1 as a match/mismatch detector: A high-resolution fMRI study of the human hippocampus. *Hippocampus*, *22*(3), 389–398. https://doi.org/10.1002/hipo.20933

Dymond, S., Dunsmoor, J. E., Vervliet, B., Roche, B., & Hermans, D. (2015). Fear Generalization in Humans: Systematic Review and Implications for Anxiety Disorder Research. *Behavior Therapy*, *46*(5), 561–582. https://doi.org/10.1016/j.beth.2014.10.001

Forbes, E. E., Phillips, M. L., Silk, J. S., Ryan, N. D., & Dahl, R. E. (2011). Neural systems of threat processing in adolescents: Role of pubertal maturation and relation to measures of negative affect. *Developmental Neuropsychology*, *36*(4), 429–452. https://doi.org/10.1080/87565641.2010.550178

Giustino, T. F., & Maren, S. (2015). The Role of the Medial Prefrontal Cortex in the Conditioning and Extinction of Fear. *Frontiers in Behavioral Neuroscience*, *9*. https://doi.org/10.3389/fnbeh.2015.00298

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. *Frontiers in Neuroinformatics*, *5*. https://doi.org/10.3389/fninf.2011.00013

Greenberg, T., Carlson, J. M., Cha, J., Hajcak, G., & Mujica-Parodi, L. R. (2013). Neural reactivity tracks fear generalization gradients. *Biological Psychology*, *92*(1), 2–8. https://doi.org/10.1016/j.biopsycho.2011.12.007

Jasnow, A. M., Lynch, J. F., Gilman, T. L., & Riccio, D. C. (2017). Perspectives on fear generalization and its implications for emotional disorders: Fear Generalization and Emotional Disorders. *Journal of Neuroscience Research*, *95*(3), 821–835. https://doi.org/10.1002/jnr.23837

Keresztes, A., Bender, A. R., Bodammer, N. C., Lindenberger, U., Shing, Y. L., & Werkle-Bergner, M. (2017). Hippocampal maturity promotes memory distinctiveness in childhood and adolescence. *Proceedings of the National Academy of Sciences*, *114*(34), 9212–9217. https://doi.org/10.1073/pnas.1710654114

Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. L. (2010). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning & Memory*, *18*(1), 15–18. https://doi.org/10.1101/lm.1971111

Lavenex, P., & Banta Lavenex, P. (2013). Building hippocampal circuits to learn and remember: Insights into the development of human memory. *Behavioural Brain Research*, *254*, 8–21. https://doi.org/10.1016/j.bbr.2013.02.007

Leal, S. L., & Yassa, M. A. (2014). Effects of aging on mnemonic discrimination of emotional information. *Behavioral Neuroscience*, *128*(5), 539–547. https://doi.org/10.1037/bne0000011

Leal, S. L., & Yassa, M. A. (2018). Integrating new findings and examining clinical applications of pattern separation. *Nature Neuroscience*, *21*(2), 163–173. https://doi.org/10.1038/s41593-017-0065-1

Lee, J. K., Ekstrom, A. D., & Ghetti, S. (2014). Volume of hippocampal subfields and episodic memory in childhood and adolescence. *NeuroImage*, *94*, 162–171. https://doi.org/10.1016/j.neuroimage.2014.03.019

Lisman, John E., & Grace, A. A. (2005). The hippocampal-VTA loop: Controlling the entry of information into long-term memory. *Neuron*, *46*(5), 703–713. https://doi.org/10.1016/j.neuron.2005.05.002

Lissek, S., Bradford, D. E., Alvarez, R. P., Burton, P., Espensen-Sturges, T., Reynolds, R. C., & Grillon, C. (2014). Neural substrates of classically conditioned fear-generalization in humans: A parametric fMRI study. *Social Cognitive and Affective Neuroscience*, *9*(8), 1134–1142. https://doi.org/10.1093/scan/nst096

Mack, M. L., Love, B. C., & Preston, A. R. (2018). Building concepts one episode at a time: The hippocampus and concept formation. *Neuroscience Letters*, *680*, 31–38. https://doi.org/10.1016/j.neulet.2017.07.061

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *262*(841), 23–81. https://doi.org/10.1098/rstb.1971.0078

Mather, M., & Sutherland, M. R. (2011). Arousal-Biased Competition in Perception and Memory. *Perspectives on Psychological Science*, *6*(2), 114–133. https://doi.org/10.1177/1745691611400234

McGaugh, J. L. (2013). Making lasting memories: Remembering the significant. *Proceedings of the National Academy of Sciences*, *110*(Supplement 2), 10402–10407. https://doi.org/10.1073/pnas.1301209110

Onat, S., & Büchel, C. (2015). The neuronal basis of fear generalization in humans. *Nature Neuroscience*, *18*(12), 1811–1818. https://doi.org/10.1038/nn.4166

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the Amygdala to Emotion Processing: From Animal Models to Human Behavior. *Neuron*, *48*(2), 175–187. https://doi.org/10.1016/j.neuron.2005.09.025

Pine, D. S., Cohen, P., Gurley, D., Brook, J., & Ma, Y. (1998). The risk for early-adulthood anxiety and depressive disorders in adolescents with anxiety and depressive disorders. *Archives of General Psychiatry*, *55*(1), 56–64. https://doi.org/10.1001/archpsyc.55.1.56

Resnik, J., Sobel, N., & Paz, R. (2011). Auditory aversive learning increases discrimination thresholds. *Nature Neuroscience*, *14*(6), 791–796. https://doi.org/10.1038/nn.2802

Rolls, E. T. (2013). The mechanisms for pattern completion and pattern separation in the hippocampus. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00074

Schafe, G. E., Doyère, V., & LeDoux, J. E. (2005). Tracking the fear engram: The lateral amygdala is an essential locus of fear memory storage. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(43), 10010–10014. https://doi.org/10.1523/JNEUROSCI.3307-05.2005

Schechtman, E., Laufer, O., & Paz, R. (2010). Negative Valence Widens Generalization of Learning. *Journal of Neuroscience*, *30*(31), 10460–10464. https://doi.org/10.1523/JNEUROSCI.2377-10.2010

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8. https://doi.org/10.1016/j.cobeha.2014.07.005

Sinha, N., Berg, C. N., Tustison, N. J., Shaw, A., Hill, D., Yassa, M. A., & Gluck, M. A. (2018). APOE ε4 status in healthy older African Americans is associated with deficits in pattern separation and hippocampal hyperactivation. *Neurobiology of Aging*, *69*, 221–229. https://doi.org/10.1016/j.neurobiolaging.2018.05.023

Smith, S. M., & Brady, J. M. (1997). SUSAN—A New Approach to Low Level Image Processing. *International Journal of Computer Vision*, *23*(1), 45–78. https://doi.org/10.1023/A:1007963824710

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., … Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23 Suppl 1*, S208-219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Spalding, K. N. (2018). The role of the medial prefrontal cortex in the generalization of conditioned fear. *Neuropsychology*, *32*(1), 1–17. https://doi.org/10.1037/neu0000384

Starita, F., Kroes, M. C. W., Davachi, L., Phelps, E. A., & Dunsmoor, J. E. (2019). Threat learning promotes generalization of episodic memory. *Journal of Experimental Psychology. General*, *148*(8), 1426–1434. https://doi.org/10.1037/xge0000551

Tustison, N. J., & Avants, B. B. (2013). Explicit B-spline regularization in diffeomorphic image registration. *Frontiers in Neuroinformatics*, *7*. https://doi.org/10.3389/fninf.2013.00039

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

van Praag, H., Shubert, T., Zhao, C., Gage, F. H. (2005). Exercise enhances learning and hippocampal neurogenesis in aged mice. The Journal of Neuroscience, 25(38), 8680-8685.

Wang, H., & Yushkevich, P. A. (2013). Multi-atlas segmentation with joint label fusion and corrective learning—An open source implementation. *Frontiers in Neuroinformatics*, *7*. https://doi.org/10.3389/fninf.2013.00027

Yassa, M. A., Mattfeld, A. T., Stark, S. M., & Stark, C. E. L. (2011). Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(21), 8873–8878. https://doi.org/10.1073/pnas.1101567108

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yushkevich, P. A., Amaral, R. S. C., Augustinack, J. C., Bender, A. R., Bernstein, J. D., Boccardi, M., … Hippocampal Subfields Group (HSG). (2015). Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards a harmonized segmentation protocol. *NeuroImage*, *111*, 526–541. https://doi.org/10.1016/j.neuroimage.2015.01.004

ONLINE METHODS AND MATERIALS

Participants

Participants between the ages of 9 and 14 years were recruited from anxiety clinic referrals and the broader Miami-Dade community and paid for their participation. A clinic and community recruitment strategy was employed to maximize variability in the construct of interest (negative overgeneralization). All participants were right-handed and had normal or corrected to normal vision. A total of 52 participants were enrolled in the study. All participants were screened for major medical and psychiatric exclusionary comorbidities (e.g., current depressive episode, bipolar disorder, post-traumatic stress disorder, attention-deficit/hyperactivity disorder, conduct disorder, oppositional defiant disorder, psychotic disorders, obsessive compulsive disorder). These diagnostic exclusionary criteria were chosen in order to enhance specificity of the construct and mechanisms of interest. Following the intake session three participants were determined to be ineligible (i.e., left-handed) and another dropped out of the study. Forty-eight participants completed the Study session scan. Two of the 48 Study session participants were not invited to return for the 12 hr delay Test session due to excessive motion during the first scan and errors in the administration of the experimental protocol, while a third failed to show-up to their scheduled appointment leaving 45 volunteers who participated in the Test session. Of the remaining 45 participants, one participant was removed due to excessive motion during their second scan (>20% of trials were flagged for removal following artifact detection), one failed to show up to their appointment, six participants were removed due to poor performance on the task (hit rate for targets was 1.5SD below the average performance), and three participants were removed due to errors in the triggering of the task with the onset of the scanner, leaving N = 34 (mean age ± SD, 11.4 ± 2.0 years, 16 female) for the final sample.

General design procedures

The local institutional review board approved all study procedures and each participant completed informed consent (parents) and assent (youth participants). The study design included baseline measures assessing inclusion/exclusion criteria (see above), followed by randomization of participants to sleep (N = 16) or wake (N = 18) conditions in order to evaluate the role of sleep in emotional memory consolidation. In both conditions, Study scan and Test scan were separated by a 12 hour retention period. The timing for the two conditions differed such that the sleep condition completed Study scan in the evening and Test scan in the morning, while the wake condition completed Study scan in the morning and Test scan in the evening. The effects of sleep versus wake conditions on negative overgeneralization will be detailed in a future report but are outside of the scope of the current manuscript which is instead focused on general mechanisms of negative overgeneralization.

Emotional Memory Task Procedures

The emotional mnemonic similarity task (Leal and Yassa, 2014) consisted of two phases: Study and Test (**Fig. 1a**). During the Study phase, participants viewed (2 s) emotional and non-emotional scenes presented in a randomized order. Participants were instructed to indicate the emotional valence (negative, neutral, or positive) of each image with one of three button responses using an MR-compatible response device. Each scene was followed by a white fixation cross in the middle of a black screen of variable duration (2-6 s) before the onset of the next trial. During the Study phase participants saw a total of 145 stimuli (48 negative, 47 neutral, 50 positive) equally divided between two scanning runs lasting approximately 8 minutes. Participants returned to the scanner after a 12 hr delay for the Test phase of the experiment at which point they were given a surprise memory test. Participants were presented with scenes which were viewed during the Study phase of the experiment (targets), scenes that were similar but not identical to the images presented during the Study phase (lures), and completely new scenes (foils). Each scene was presented in the middle of the screen for 2 s and followed by a variable inter-stimulus-interval (2-6 s) during which a white fixation cross was presented in the middle of a black screen. Participants were instructed to indicate whether items were “old” or “new” by pressing one of two buttons using a MR-compatible response device. The instructions explicitly indicated that for an item to be endorsed as “old” it had to be the *exact same* image that was presented during the Study phase. A total of 284 stimuli were presented during the Test phase 16 negative targets, 16 neutral targets, 16 positive targets, 32 negative lures, 33 neutral lures, 33 positive lures, 42 negative foils, 49 neutral foils, 48 positive foils).

Neuroimaging Data Collection and Preprocessing

Neuroimaging data were collected on a 3T Siemens MAGNETOM Prisma scanner with a 32-channel head coil at the Center for Imaging Science at Florida International University. A T2*-weighted EPI sequence (TR = 993 ms, TE = 30 ms, flip angle = 52º, FOV = 216 mm, 56 axial slices, slice acceleration = 4, voxel size = 2.4 mm isotropic) in addition to a T1-weighted magnetization-prepared rapid gradient echo sequence (MPRAGE: TR = 2500 ms, TE = 2.9 ms, flip angle = 8º, FOV = 256 mm, 176 sagittal slices, voxel size = 1 mm isotropic) were collected. During each run of the emotional memory task acquisition began after the first four volumes were collected to allow for T1-equilibriazation. A total of 448 whole brain volumes were acquired during the Study portion and 454 volumes were collected during the test portion.

The following software packages were utilized for neuroimaging data preprocessing using a custom scripted Neuroimaging in Python (Nipype version 0.12.1; Gorgolewski et al., 2011) pipeline: Analysis of Functional Neuroimages (AFNI version 16.3.18 Cox, 1996), FMRIB Software Library (FSL version 5.0.10; Smith et al., 2004), FreeSurfer (version 6.0.0; Fischl, 2012), and Advanced Normalization Tools (ANTs version 2.1.0, Avants, Epstein, Grossman, & Gee, 2008). Cortical surface reconstruction and cortical/subcortical segmentation was obtained for each T1-weighted MPRAGE. Surfaces were inspected for errors, manually edited (if necessary), and resubmitted for reconstruction. Functional volumes that exceeded three standard deviations of the mean intensity within a run were removed and replaced ('despiked'). Functional data were then 1) motion corrected, aligning all volumes across each run to the middle volume of the first run; 2) co-registered to the structural scan; 3) motion and intensity outliers (>1 mm frame-wise displacement; >3 SD mean intensity) were identified; and 4) spatially filtered with a 4mm kernel using the *SUSAN* algorithm (Smith & Brady, 1997).

Each participant's T1-weighted structural scan was skull-stripped and registered to an MNI-152 template via a rigid body transformation. We used this pass to reduce normalization errors due to large differences in position across participants and to generate a template close to a commonly used reference. The rigid-body transformed structural scans were then used to create a study-specific template using ANTs. Following template generation each participant's skull-stripped brain in FreeSurfer space was normalized to the study template using ANTs non-linear symmetric diffeomorphic mapping. The warps obtained from normalization were used to move contrast parameter estimates following fixed-effects modeling into the study-specific template space for group-level tests.

Behavioral Data Analysis

To assess generalization, the primary behavioral outcome measure for the emotional memory task was the lure generalization index (LGI). The LGI score provided a bias-corrected measure of how likely participants were to false alarm to similar lure items. LGI was defined as p('Old'|Lure) – p('Old'|Foil), correcting for a general tendency to call new stimuli 'Old'. Similar corrections have been used in prior work (Lacy et al., 2010; Yassa et al., 2011). Similarly, to assess valence related changes to discrimination we calculated the lure discrimination index (LDI: p('New'|Lure) – p('New'|Target)). The

LDI score corrected for the propensity of participants to endorse items as new. Pairwise comparisons were performed across the different stimulus valences for both the LGI and LDI scores. A Bonferroni correction was used to account for Type I error inflation following multiple comparisons.

MRI Data Analysis

Anatomical Regions of Interest

Our study was motivated to understand the contributions of *a priori* anatomical ROIs to negative generalization: CA1 subfield of the hippocampus, CA3/Dentate Gyrus combined subfield of the hippocampus, amygdala, and dorsal/ventral medial prefrontal cortex (mPFC). The amygdala was defined by binarizing the FreeSurfer subcortical segmentation obtained from the aparc+aseg.mgz file. The dorsal/ventral medial prefrontal cortex was also defined by binarizing the following FreeSurfer cortical segmentations: dorsal mPFC: superior frontal, caudal anterior cingulate, and rostral anterior cingulate; ventral mPFC: medial orbitofrontal.

Hippocampal subfield segmentation was performed using a consensus labeling approach based on an in-house atlas set of 19 T1 MPRAGE scans (0.75mm isotropic) and their corresponding T2-FSE scans, acquired in an oblique orientation perpendicular to the long axis of the hippocampus (0.47x0.47 mm^2 in-plane, 2.0 mm slice thickness). Expert manual segmentations were applied to this atlas set based on highly reliable and published protocols (Yushkevich et al., 2015) and comprised the following ROIs: left/right DG/CA3, left/right CA1, and left/right subiculum.

To propagate a weighted consensus labeling from the expertly labeled atlas set to one of the unlabeled T1-weighted images of our study cohort, we spatially normalized the atlas set to the unlabeled subject and applied joint label fusion (Wang & Yushkevich, 2013). This consensus-based labeling approach as well the spatial normalization steps employed the ANTs toolkit (Brian B. Avants et al., 2011). First, the intra-subject atlas

T1/T2 rigid transforms were calculated. In order to minimize the total number of deformable registrations, a “pseudo-geodesic” approach was used to align the data (Tustison & Avants, 2013). This required the construction of an optimal T1-weighted template (Brian B. Avants et al., 2010) representing the average shape/intensity information of the T1 component of the atlas set. The deformable transformations between each T1-weighted image of the study cohort and the T1 atlas template were calculated. The transformation between the atlas labels and unlabeled study cohort image was then computed by concatenating the $T1_{atlas}$ /$T2_{atlas}$ rigid transformation, the $T1_{atlas}$ /T1 template deformable transformation, and the T1 template/and $T1_{subject}$ deformable transforms. Once the atlas set was normalized to the unlabeled subject, the regional labeling was determined using weighted averaging where the weighting takes into account the unique intensity information contributed by each atlas member. This approach which combines label information and intensity information has been used in a number of recent publications to segment hippocampal subfields (Brown, Kulikova, et al., 2019; Sinha et al., 2018).

All masks were back-projected to functional space using the inverse of the co-registration transformation for data analyses.

Task Neuroimaging Data Analysis

Functional neuroimaging data were analyzed using a general linear model approach in FSL. Separate models were created for the Study and Test data to evaluate the neurobiological correlates of negative generalization. Both the Study and Test models included the following regressors of no interest: motion (x, y, z translations; pitch, roll, yaw rotation), the first and second derivatives of the motion parameters, normalized motion, first through third order Lagrange polynomials to account for low frequency changes in the signal, as well as a regressor for each outlier time-point that exceeded outlier thresholds. The Test model included 12 regressors of interest: lure false alarms

(FAs), lure correct rejections (CRs), target hits, and target misses. All regressors were separately modeled for the three different valences (negative, neutral and positive). The Study model included a similar combination of regressors, however, these were scenes that would subsequently be target hits and misses or be replaced by similar lures and subsequently correctly rejected or false alarmed. To investigate signals related to negative generalization we focused our analyses on negative compared to neutral lure trials that were (subsequently) false alarmed.

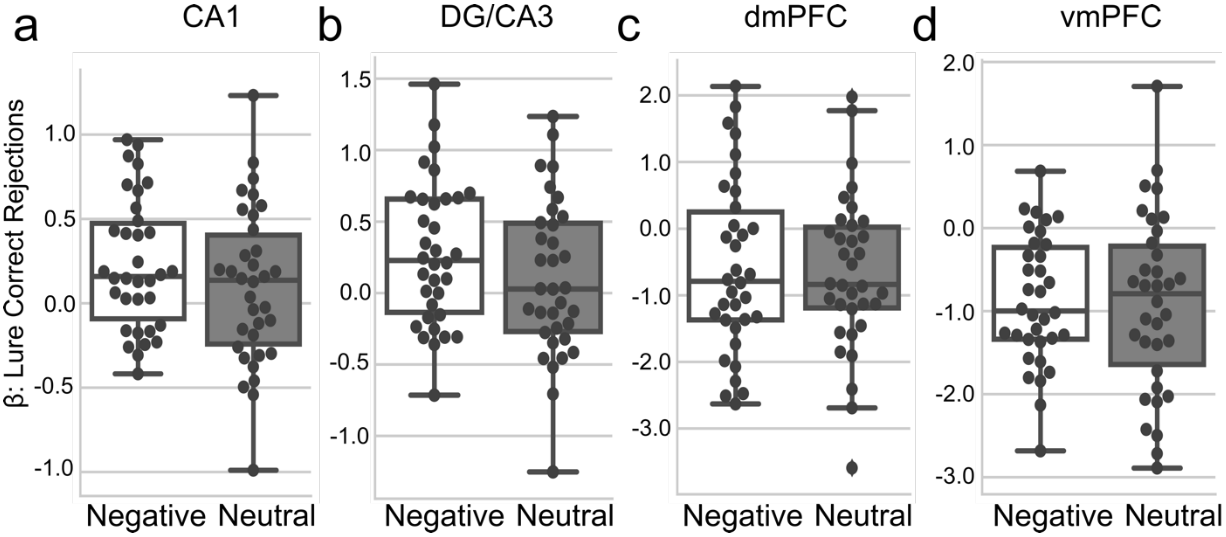
Task-Based Functional Connectivity Analysis

We used a beta-series correlation analysis (Rissman et al., 2004) to assess the functional interactions between the amygdala and the CA1, dorsal mPFC, and ventral mPFC during encoding. To isolate our trials of interest (negative and neutral stimuli that were subsequently replaced by lures and either false alarmed or correctly rejected) we used a least-squre single (LSS) approach (Mumford et al., 2012). A general linear model was run for each trial of interest. Othe regressors in the models included the typical task-based regressors (minus the relevant trial of interest) and regressors of no interest accounting for motion and low frequency noise. The resulting contrast parameter estimates for each isolated trial of interest were constructed into a beta-series and averaged across voxels within each region of interest. Correlations between regions of interest were calculated. The arctangent of the resulting Pearson's correlation coefficients were calculated and pair-wise comparisons were made.

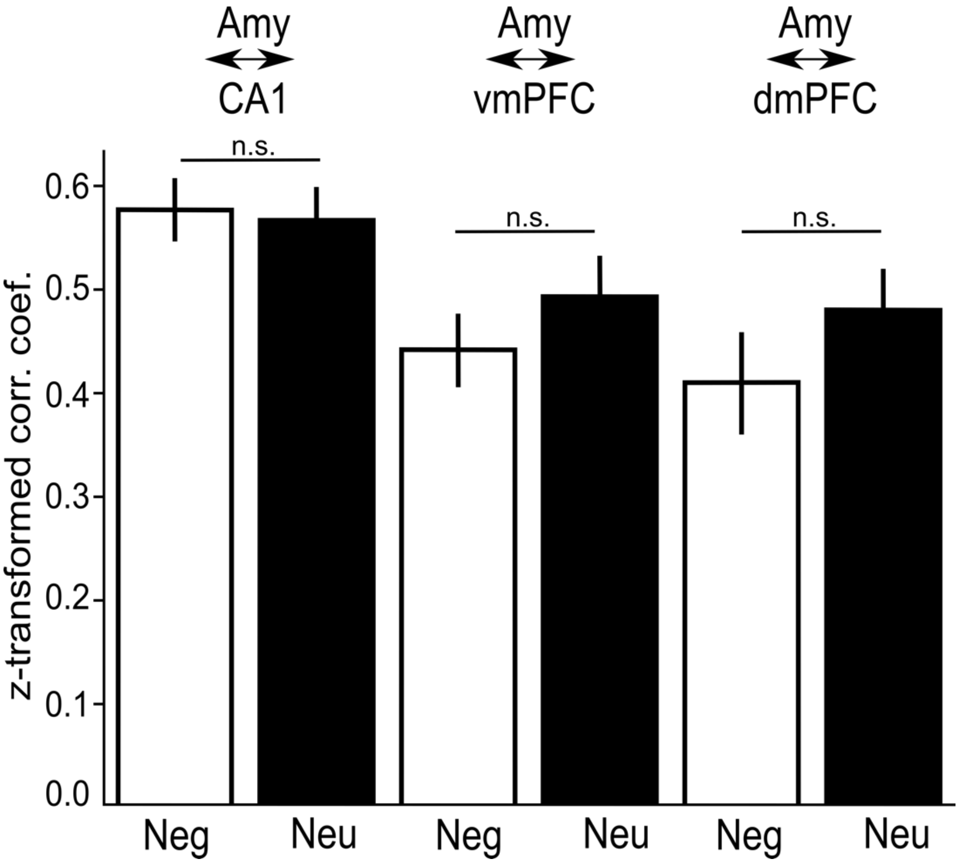
Representational Similarity Analysis

Neurobiological correlates of generalization may arise as similar/dissimilar patterns of activation. To evaluate this possibility, we performed a representational similarity analysis (RSA) using the bilateral CA1 subfield of the hippocampus and dorsal/ventral mPFC as our anatomical regions of interest. The original Test phase model was re-run however this time using un-spatially smoothed functional data – as fine differences in activation patterns were the goal of the RSA analysis. Representational similarity (voxel-wise correlations) was separately calculated in the anatomical regions of interest between target hits and lure false alarms across the different valences (negative, neutral, positive). Large values indicated similar patterns of activation and small values related to dissimilar patterns of activation between target hits and lure false alarms. To evaluate the relation between behavioral generalization and neural measures of pattern dissimilarity separate correlations were performed. Bonferroni corrections were used to correct for inflated Type I error due to multiple comparisons.

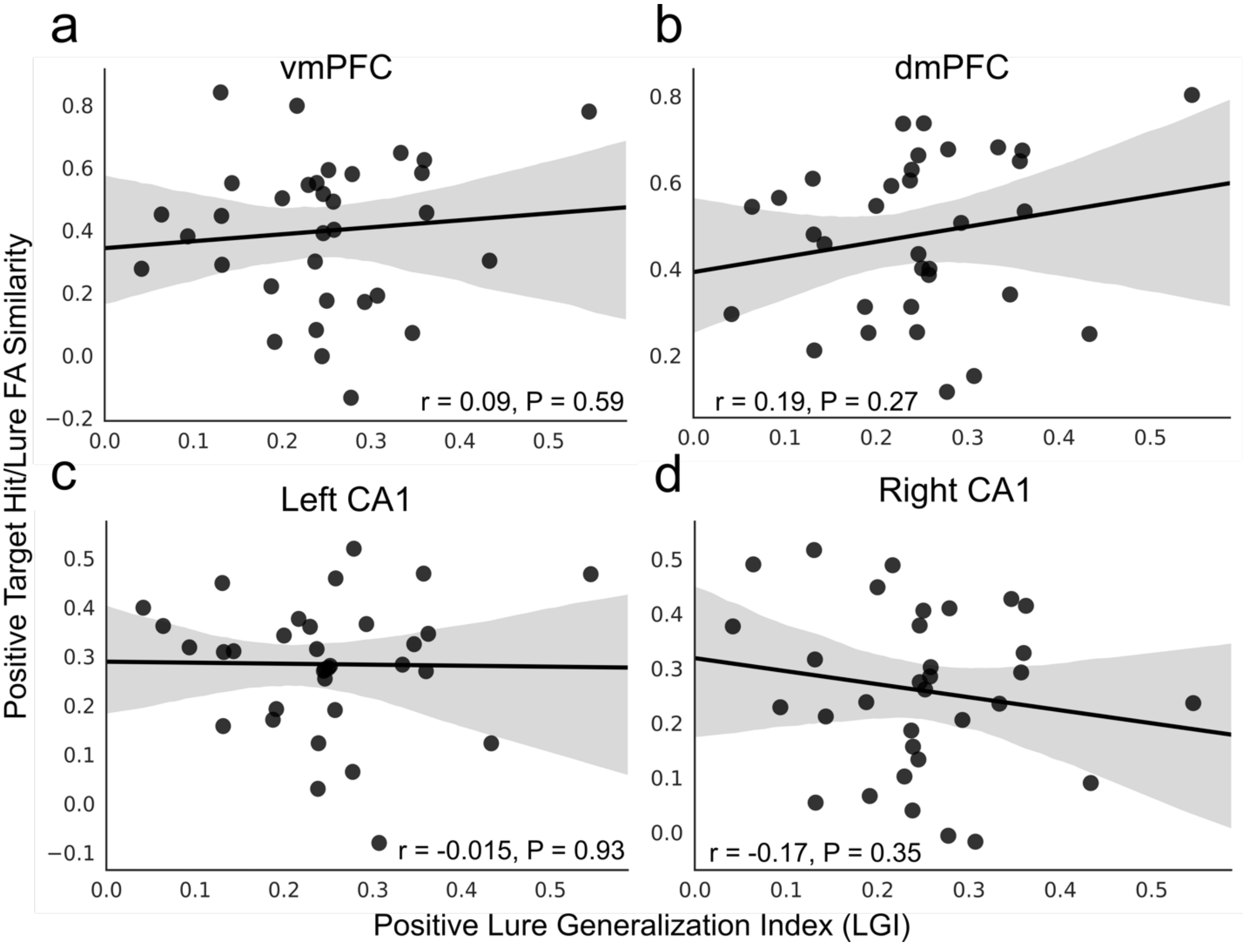
SUPPLEMENTAL MATERIALS



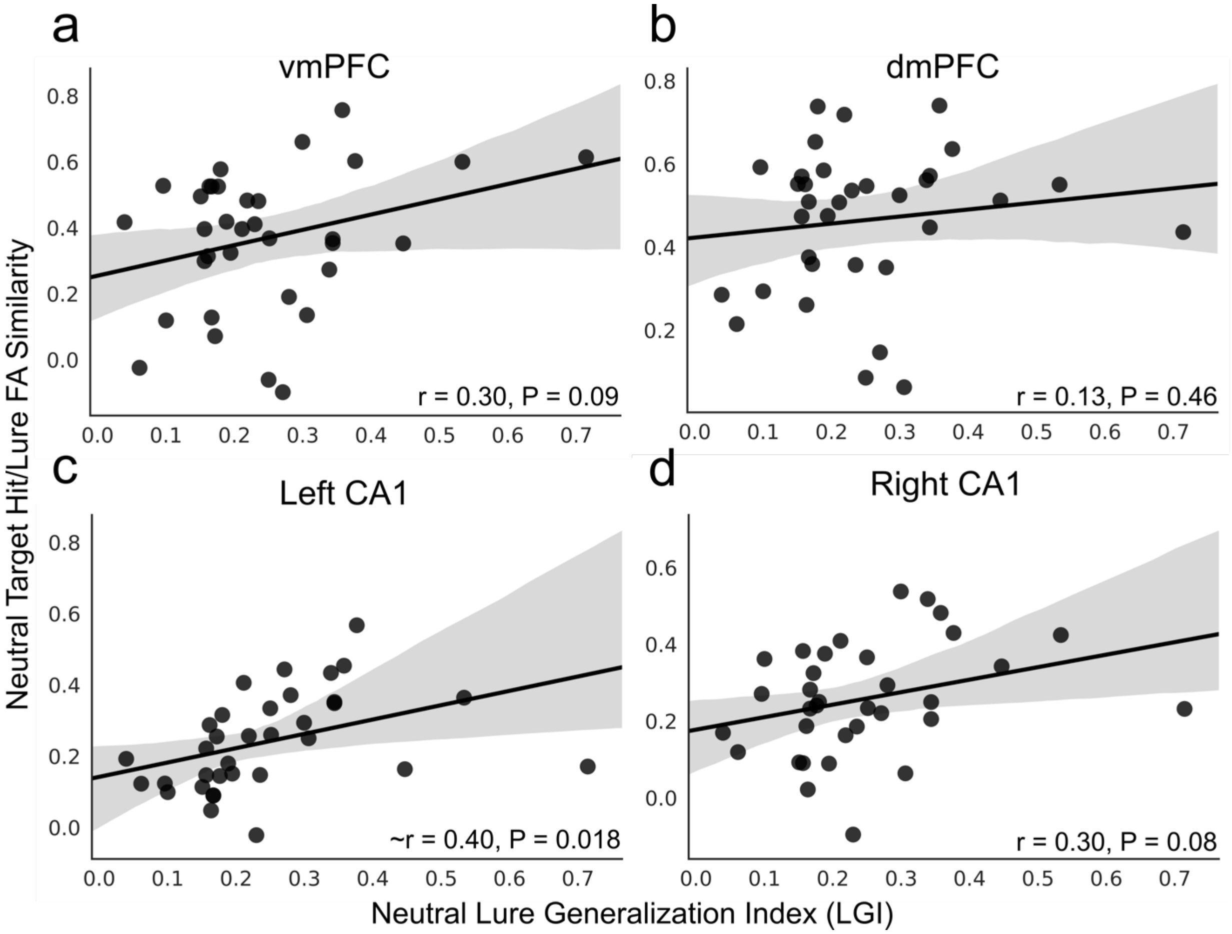
Supplemental Figure 1. Anatomical region-of-interest – hippocampal subfield (CA1 and DG/CA3), dorsal medial prefrontal cortex (dmPFC), or ventral medial prefrontal cortex (vmPFC) – examining differences in activations between negative and neutral lure stimuli when they were correctly rejected. (**a**) The CA1 hippocampal subfield (*P* = 0.03) and (**b**) the DG/CA3 subfield (*P* = 0.03) exhibited a trend towards greater activation for negative relative to neutral correct rejections. (**c**) The dmPFC (*P* = 0.76) and (**d**) the vmPFC (*P* = 0.69) did not exhibit significantly different activations between negative and neutral correct rejections. All comparisons were not significant following Bonferroni correction for multiple comparisons.



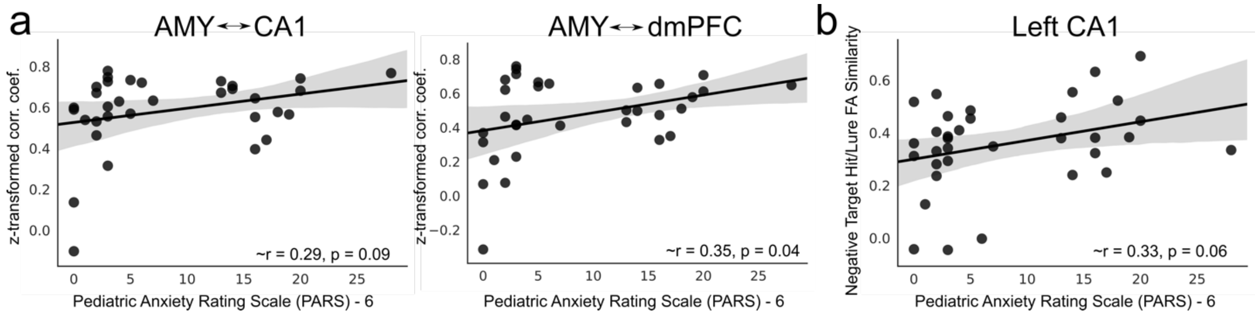
Supplemental Figure 2. Task based functional connectivity during the Study phase for Correctly Rejected negative (Neg) and neutral (Neu) lures between the amygdala (Amy) and three target regions: CA1, ventral medial prefrontal cortex (vmPFC), and dorsal medial prefrontal cortex (dmPFC). Functional coupling for negative and neutral items that were subsequently correctly rejected were not significantly different between any of the a priori anatomically defined regions of interest: Amy ↔ CA1 (P = 0.32); Amy ↔ dmPFC (P = 0.64); Amy ↔ vmPFC (P = 0.45).



Supplemental Figure 3. Correlations between positive target hit and lure false alarm (FA) similarity in the (**a**) ventral medial prefrontal cortex (vmPFC), (**b**) dorsal medial prefrontal cortex (dmPFC) and the (**c**) left and (**d**) right CA1 and LGI behavioral performance. No regions exhibited a significant correlation between positive target hit/lure FA similarity and positive LGI performance.



Supplemental Figure 4. Correlations between neutral target hit and lure false alarm (FA) similarity in the (**a**) ventral medial prefrontal cortex (vmPFC), (**b**) dorsal medial prefrontal cortex (dmPFC) and the (**c**) left and (**d**) right CA1 and LGI behavioral performance. The vmPFC (*P* = .09) and the left (*P* = 0.018) and right (*P* = 0.08) CA1 exhibited a trend in the relation between neutral target hit/lure FA similarity and neutral LGI performance. ~Trending significant relation.



Supplemental Figure 5. Correlations between select brain measures and anxiety severity as measured by the Pediatric Anxiety Rating Scale (PARS) – 6. (**a**) A trend towards increased functional connectivity between the amygdala and CA1 and the amygdala and dorsal medial prefrontal cortex (dmPFC) during the Study phase for items that were subsequently false alarmed and anxiety severity identified (AMY ↔ CA1: r = 0.29, p = 0.09; AMY ↔ dmPFC: r = 0.35, p = 0.04). (**b**) A similar trend was noted in our representational similarity analysis between negative target hits and lure false alarms (FA). More anxious symptoms were related to greater similarity between these trial types in the left CA1 (r = 0.33, p = 0.06). ~Trending significant relation.