



Published in final edited form as:

Neuroimage. 2011 April 15; 55(4): 1878–1888. doi:10.1016/j.neuroimage.2011.01.041.

Neurobehavioral mechanisms of human fear generalization

Joseph E. Dunsmoor, Steven E. Prince, Vishnu P. Murty, Philip A. Kragel, and Kevin S. LaBar

Center for Cognitive Neuroscience, Duke University, Durham, NC USA

Abstract

While much research has elucidated the neurobiology of fear learning, the neural systems supporting the generalization of learned fear are unknown. Using functional magnetic resonance imaging (fMRI), we show that regions involved in the acquisition of fear support the generalization of fear to stimuli that are similar to a learned threat, but vary in fear intensity value. Behaviorally, subjects retrospectively misidentified a learned threat as a more intense stimulus and expressed greater skin conductance responses (SCR) to generalized stimuli of high intensity. Brain activity related to intensity-based fear generalization was observed in the striatum, insula, thalamus/periacqueductal gray, and subgenual cingulate cortex. The psychophysiological expression of generalized fear correlated with amygdala activity, and connectivity between the amygdala and extrastriate visual cortex was correlated with individual differences in trait anxiety. These findings reveal the brain regions and functional networks involved in flexibly responding to stimuli that resemble a learned threat. These regions may comprise an intensity-based fear generalization circuit that underlies retrospective biases in threat value estimation and overgeneralization of fear in anxiety disorders.

INTRODUCTION

Fear learning involves acquiring defensive behaviors to aid survival in response to environmental threats. To be adaptive, it is important that this learning is flexible such that stimuli highly similar to a learned threat are treated as potentially harmful as well. Extensive neurophysiological and brain imaging research has established several key regions involved in fear learning processes, including the amygdala, insula, cingulate gyrus, striatum, sensory cortex, and prefrontal cortex (Phelps and LeDoux, 2005). The goal of the present study is to examine neural systems contributing to the *generalization* of fear learning, and to link generalization-related activity to individual differences in fear expression, functional connectivity, and trait anxiety.

Generalization of fear learning serves a functional purpose because it allows an organism to treat novel stimuli appropriately based on experience with related stimuli. For example, an animal knows to avoid a potential predator if it resembles one that has been encountered in the past. However, in some cases, the transfer of fear following a learning episode has maladaptive consequences. For instance, an animal that widely casts defensive behaviors towards a broad range of stimuli is at risk of wasting energy resources. This overgeneralization of fear to harmless stimuli is often a symptom of clinical anxiety, as

Corresponding author: Kevin S. LaBar, PhD, Center for Cognitive Neuroscience, Box 90999, Duke University, Durham, NC 27708-0999, tel: (919)681-0664, fax: (919)681-0815, klabar@duke.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

exemplified by posttraumatic stress disorder. Thus, to be well adapted to the environment, an organism must balance expressing fear behaviors towards novel threats on the one hand and withholding fear responses to non-threats on the other hand.

The laboratory study of fear generalization has frequently used Pavlovian conditioning procedures, wherein a neutral conditioned stimulus (CS; e.g., a tone) predicts an intrinsically threatening unconditioned stimulus (US; e.g., an electric shock). After a CS-US association is formed, the CS evokes a conditioned fear response (CR), such as a change in heart rate, respiration rate, or sweating. Generalization occurs when these behaviors are evoked by stimuli that are similar to a learned threat, but have never directly predicted the US. Several factors contribute to fear generalization. For instance, in his seminal studies of non-human animals, Pavlov (1927) observed that the CR generalized to graded stimuli that closely resembled the original CS, and diminished as perceptual similarity decreased. Gradients that track perceptual similarity have been consistently observed in animal conditioning experiments (Honig and Urcuioli, 1981). The ability to discriminate a CS from a non-CS additionally affects the breadth of fear generalization; animals trained with a single CS show more widespread generalization than animals trained to discriminate between different stimuli along the same sensory dimension (Jenkins and Harrison, 1960). Studies have also shown generalization to increase as a function of intensity -- for instance from a medium volume sound that predicts the US to a loud volume sound that has never predicted the US (Ghirlanda and Enquist, 2003). Intensity generalization often involves a shift in peak responding, such that a non-CS of greater intensity than the CS evokes a greater response than the CS itself (Ghirlanda and Enquist, 2003). Models of stimulus-intensity generalization may be particularly well suited to describe fear behaviors, since following an aversive experience (e.g., encounter with a vicious dog) an intrinsically salient stimulus (e.g., forbidding dog) may preferentially evoke a greater fear response than a similar but less intense stimulus (e.g., harmless dog). Recent human behavioral studies have shown that generalized fear responses are impacted by manipulations in discriminatory fear learning (Dunsmoor et al., 2009), and have reported generalization along perceptual (Lissek et al., 2008) and emotional intensity dimensions (Dunsmoor et al., 2009).

Neurophysiological research of fear generalization in non-human animals has focused primarily on the amygdala (e.g. Armony et al., 1997), as this region serves a pivotal role in forming the CS-US association and producing conditioned fear behaviors (Davis, 1992). For example, a loss of GABA release in the lateral nucleus of the amygdala has been shown to increase fear generalization (Bergado-Acosta et al., 2008; Shaban et al., 2006). Duvarci and colleagues (2009) conducted a fear conditioning task in rats using two cues -- a CS paired with an electric shock (CS+) and a control CS unpaired with the US (CS-) -- and found large individual differences in the extent to which animals generalized fear from the CS+ to the CS-. They suggested that differences in generalization were determined in part by the bed nucleus of the stria terminalis, a region closely linked with the central nucleus of the amygdala, as rats with excitotoxic lesions to this region showed a small amount of generalization to the CS-. However, the precise role of the amygdala (and other regions) in fear generalization is not clear, as generalization is often conceptualized as heightened responses to the CS-, and formal behavioral tests using graded stimuli have rarely been conducted in neuroscientific models of fear generalization. Human neuroimaging studies have shown that the amygdala, striatum, and medial prefrontal cortex flexibly respond to CSs that change in their association with an aversive US (Schiller and Delgado, 2010), but it is not clear whether these regions show graded responses to stimuli that vary in similarity from the CS+ along some featural dimension.

The present study provides a novel, systematic examination of human fear generalization using event-related functional magnetic resonance imaging (fMRI) with concurrent

measures of psychophysiological arousal (i.e., skin conductance response, SCR). Predictions on how the human brain mediates fear generalization were based on knowledge of the brain systems involved in acquiring and expressing learned fear. First, generalization could be mediated by regions involved in differential fear learning -- that is, areas that show greater activity to a CS+ compared to a CS-. We expected that learning-related regions would show enhanced activity to generalized stimuli of high emotional intensity, consistent with our prior behavioral work showing that generalization increases as a function of the emotional intensity value of non-conditioned cues (Dunsmoor et al., 2009). However, it is possible that regions that show broad enhancement of activity to generalized stimuli are not directly related to the production of fear behaviors, and thus poorly reflect variability in fear learning and generalization. Therefore, to constrain interpretations of neural activity, we quantified behavioral measures of generalization by the change in SCR magnitude from pre-to-post fear conditioning, and used these residualized scores to examine brain-behavior correlates during a test of fear generalization. These individual behavioral profiles take advantage of the fact that psychophysiological responses entail considerable individual variability that may be mediated by components of the fear learning circuitry, such as the amygdala (Cheng et al., 2006; Knight et al., 2005). Thus, we predicted that the change in fear expression following fear conditioning would be correlated with activity in limbic/paralimbic regions involved in sympathetic activation and core affective processes, such as the amygdala and insula. The amygdala is also important for enhancing the sensory representation of feared stimuli (Armony and Dolan, 2002) through reciprocal connections with sensory processing regions like the extrastriate visual cortex (Amaral et al., 2003). We predicted enhanced connectivity between the amygdala and visual processing regions coding for the domain specific properties of the CS and related stimuli following fear conditioning. Finally, to provide a link to clinical anxiety disorders, we investigated whether amygdala connectivity related to fear generalization is correlated with individual differences in trait anxiety, in line with previous findings showing enhanced amygdala-extrastriate connectivity in phobic individuals viewing phobia-relevant images (Ahs et al., 2009) and serotonin transporter (5-HTT) short allele homozygotes viewing fearful faces (Surguladze et al., 2008).

We adapted a Pavlovian conditioning paradigm that allows for simultaneous examination of fear generalization as a function of perceptual similarity and emotional intensity (Dunsmoor et al., 2009) for use with fMRI. During fear conditioning, participants received pairings of a moderately fearful face (CS+) and a shock US, as well as an unreinforced neutral face stimulus (CS-) (see Fig. 1). The generalized stimuli were gradations of the same individual's facial expression morphed incrementally between neutral and fearful endpoints, presented prior to conditioning to measure baseline responses and after conditioning in a steady-state generalization test (Blough, 1975). By detailing brain activation, fear expression, and functional connectivity before and after learning has occurred, the present study can help elucidate how humans generalize from a fear learning episode. These insights, in turn, can inform neural models of anxiety disorders characterized by overgeneralized expression of acquired fear.

Materials and methods

Participants

Twenty-five right-handed healthy young adults provided written informed consent to participate in the study. Two participants were removed from the analysis due to excessive head movement (> 3 mm in any direction), and 9 participants were not included due to a lack of SCR data, which precludes an examination of fear learning and generalization (5 participants lacked SCR data due to technical issues and 4 participants were classified as non-responders as described below). The behavioral and fMRI analysis included 14 participants (7 females; age range = 19 to 30; median age = 22 yrs). Subjects completed the

State-Trait Anxiety Inventory (STAI) (Spielberger, 1983) prior to the start of the experiment. The study was approved by the Duke University Institutional Review Board.

Stimulus set

Stimuli consisted of a male face morphed along a gradient from neutral-to-fearful taken from the Ekman pictures of facial affect (Ekman and Friesen, 1976). The morphs were positioned in a full-frontal orientation and cropped to remove hair, ears, and neckline. Five morphs were created along the continuum using Morph-Man 2000 software (STOIK): 11.11% fear/88.88% neutral, 33.33% fear/66.66% neutral, 55.55% fear/44.44% neutral, 77.77% fear/22.22% neutral, and 100% fear. For clarity, these stimuli are labeled as S1, S2, S3, S4, and S5, respectively. These face morph values were chosen based on our prior published psychometric studies on categorical perception using the same stimuli (Graham et al., 2007; Thomas et al., 2007). These normative studies showed that the S3 stimulus chosen as the CS+ in the present study is as close as possible to the point of subjective equality in categorical perception, such that half of adult participants view the face as expressing fear whereas the other half view the face as neutral. Moreover, d' estimates from these studies showed that individuals can clearly discriminate these specific morph increments (the values are not perceptually confused), each successive value from S2 to S4 adds equally to the cumulative d' function so that perceptual discrimination is linear across this segment of the categorical boundary, and individuals are perceptually sensitive to morphed featural changes that are even more subtle than those chosen for the present study.. The US consisted of a 6 ms electrical stimulation applied to the right wrist, calibrated for each participant to a level deemed “highly annoying, but not painful.”

Procedure

The experimental paradigm was based on Dunsmoor et al. (2009), and is illustrated in Fig. 1. The experiment began with a short habituation phase that included 1 presentation of each of the 5 morph increments, which allowed participants to get accustomed to the experiment and reduced orienting responses. Habituation data are not reported. The scanning session consisted of three consecutive phases that occurred in the same order for each participant: preconditioning/baseline (3 runs), fear conditioning (2 runs), and the generalization test (3 runs). A short 5 minute break followed preconditioning and fear conditioning, during which time participants passively viewed a silent video clip of a train traveling through British Columbia (Highball Productions). Each trial was 4 s in duration, during which time participants were asked to rate whether or not the face was expressing fear (forced choice: yes/no) as quickly and accurately as possible by pressing one of two buttons. The order of button presses was counterbalanced across subjects. The intertrial-interval (ITI) consisted of a white fixation cross on a black background that followed the offset of each trial. The lengths of the ITI were jittered according to an exponential distribution function. Preconditioning contained a total of 9 trials of each of the 5 morph increments (45 total trials) with an average ITI of 5 s (minimum 4 s). Fear conditioning contained a total of 16 S3 (CS+) and 16 S1 (CS-) trials (32 total) with an average ITI of 11 s (minimum of 9 s). The CS+ co-terminated with the US on 10/16 trials, whereas the CS- was never paired with the US (partial reinforcement delay conditioning procedure). The generalization test contained a total of 9 trials of each of the 5 morph increments (45 total) with an average ITI of 9 s (minimum of 5 s). The S3 was intermittently paired with the US on 6/9 trials (“steady-state” generalization test) to offset the effects of extinction over the course of an extended testing session, as routinely implemented in animal models of generalization (Blough, 1975; Honig and Urcuioli, 1981). In all phases, stimulus presentation was counterbalanced and pseudorandomized such that no more than two of the same morph increment occurred in a row. Subjects were not informed of the CS-US contingencies. Following the conclusion of the generalization test, a functional localizer task was performed to isolate cortical regions

selective for processing images of faces. The localizer, based on a previously published design (Morris et al., 2008), included pseudo-randomized blocks of black and white images of faces and flowers that each contained 24 images presented for 500 milliseconds each. Two localizer blocks were run and separated by 12 s of fixation. Data from the functional localizer were entered into a general linear model with 2 conditions of interest: faces and flowers (see below).

Psychophysiological methods and analysis

All psychophysiological recording and shock administration was controlled with the MP-150 BIOPAC system (BIOPAC systems, Goleta, CA). MRI-compatible Ag/AgCl SCR electrodes were placed on the middle phalanx of the second and third digits of the non-dominant hand. The electrical stimulation, applied to the right wrist, was controlled using the STM-100 and STM-200 modules connected to the MP-150 system. All psychophysiological equipment was grounded through the RF filter panel and shielded from magnetic interference. SCR analysis was carried out using AcqKnowledge software (BIOPAC systems) using procedures previously described (Dunsmoor et al., 2009). An SCR was scored as a response if the trough-to-peak response occurred 1–4 s following stimulus onset, lasted between 0.5 and 5.0 s, and was greater than 0.02 microsiemens. A trial that did not meet these criteria was scored as a zero.

A long-standing issue in the study of stimulus generalization concerns measurement of the generalization gradient (Hull, 1943; Pavlov, 1927). The present analysis developed an approach adapted from the animal literature to characterize individual fear generalization profiles by normalizing SCRs to each morph increment as proportion of total response output (Honig and Urcuioli, 1981). This approach is particularly suitable for the present study, as it accounts for individual differences in response profiles during the preconditioning and generalization test phases. To derive this metric, preconditioning data were analyzed by dividing the sum of SCRs for each morph increment (S1–S5) by the total sum of SCRs to all morph increments during the preconditioning phase. Fear conditioning data were also normalized on the basis of response output to the CS+ and CS-. Finally, SCRs obtained during the generalization test were analyzed for each test run, as fluctuations in response patterns over time have been observed in prior behavioral stimulus generalization experiments (Dunsmoor et al., 2009). To normalize generalization test data, the sum of responses to each stimulus type was divided by the sum of responses to all trials within each of the three generalization test runs. This yielded three generalization gradients, one for each run of the generalization test. The behavioral correlates of fear generalization were operationally defined as the difference in normalized SCRs during each generalization test run from the corresponding stimulus during preconditioning. In this way, preconditioning served as a baseline measure for each face stimulus for each participant, and allowed for an analysis of the proportional change in response patterns following fear learning. SCR data were analyzed by ANOVA and polynomial trend analyses, with an α value of 0.05 (SPSS 15.0, Chicago, IL).

Functional image acquisition, preprocessing, and GLM analysis

Scanning was performed on a General Electrical Signa EXCITE HD 3.0 Tesla MRI. Subjects wore ear plugs to reduce scanner noise, and head motion was minimized by using foam pads. Blood oxygenation level-dependent functional images were acquired parallel to the AC-PC line using a SENSE™ spiral in sequence: acquisition matrix, 64 × 64; field of view, 256 × 256; flip-angle, 60°; 34 slices with interleaved acquisition; slice thickness, 3.8 mm with no gaps between slices; repetition time, 2 s; echo time, 27 ms. Functional data were preprocessed using SPM 8 software (Wellcome Department of Cognitive Neurology, University College London, www.fil.ion.ucl.ac.uk) implemented in Matlab (The Mathworks

Inc, Natick, MA). The first 4 functional images from each scanning run were discarded to account for magnetic equilibration effects, and remaining images were corrected for head motion using a threshold of 3 mm in any direction. Preprocessing included realignment, spatial normalization to the Montreal Neurological Institute (MNI) template using a fourth degree B-spline interpolation, and smoothing using an isotropic 8-mm³ Gaussian full width half maximum kernel. To remove low-frequency drifts, high pass temporal filtering was applied using a 128-s cutoff. Individual subject trial-related analysis was conducted using the GLM. Covariates of interest included the 5 morph increments for preconditioning (S1–S5), two morph increments for fear conditioning (CS+ and CS–), and 5 morph increments for generalization test (S1–S5). The hemodynamic response was modeled for each covariate of interest using a variable duration design that incorporated reaction time for each trial (Grinband et al., 2008). The US was modeled as an impulse (dirac) function and, along with the 6 head motion parameters, were included as a covariate of no interest. Second-level random effects analyses were performed using one-sample t-tests in SPM 8 to investigate the main effect of differential fear learning between the CS+ (S3) and CS– (S1). These contrasts included CS+ and CS– trials from the fear conditioning phase and the steady-state generalization test. A threshold value was initially set to $p < 0.001$, uncorrected, with an extent threshold of 5 contiguous voxels, to identify whole brain activity related to differential fear learning. For each *a priori* region of interest (ROI) identified from whole brain analysis, the search space for multiple comparisons was restricted to bilateral anatomical masks from the Wake Forest PickAtlas toolbox (Maldjian et al., 2003) using a family-wise error (FWE) correction of $p < 0.05$. These included separate masks for the caudate, insula, thalamus, and anterior cingulate cortex. All reported regions survived this correction. Contrast images were then created using each generalized stimulus (S2, S4, and S5) versus the CS– (S1), and each generalized stimulus versus the CS+ (S3). For these contrasts images, the mean beta-parameters were extracted from the preconditioning and generalization test phase using an 8 mm radius sphere surrounding the peak voxel identified from independent functional contrast comparing the CS+ versus CS– (coordinates reported in Table 1). For activation in the thalamus extending into the periaqueductal gray (PAG), the peak coordinates used for the ROI analysis were selected from a meta-analytic review by Kober et al. (2008). Statistical tests conducted on the extracted parameter estimates were assessed by one-sample t-tests for each contrast and two-sample t-tests for comparing the same contrasts across the preconditioning and the generalization test phases, with an α value of 0.05 (SPSS 15.0, Chicago, IL). Data from the functional localizer were analyzed by first-level contrasts of faces versus flowers. These contrast maps were analyzed in a second level random-effects analysis, yielding regions preferentially engaged by images of faces versus flowers. Search space was restricted to an anatomical mask of the fusiform gyrus (Maldjian et al., 2003) using FWE correction of $p < 0.05$.

fMRI regression analysis

Regression analyses were conducted using a second-level random-effects regression model in SPM 8 to investigate brain-behavior correlations in human fear conditioning and generalization. Regression analysis of fear conditioning was conducted for the two runs of the fear conditioning phase using the difference in SCRs between the CS+ (S3) and the CS– (S1) as covariates, and brain imaging contrasts of CS+ versus CS– as the dependent variable (excluding one subject from the second run for technical problems with the SCR equipment). This analysis revealed regions positively tracking the difference in SCRs between the CS+ and CS–. Next, we conducted a regression analysis of the generalization test using the change in normalized SCRs from pre-to-post fear conditioning to each stimulus as a covariate. Individual contrasts for S1–S5 (relative to baseline), for the three generalization test runs, were entered as the dependent variable. For clarity, responses to the S3 do not reflect generalization, as the S3 continued to serve as the CS+ throughout the

steady-state generalization test. Mean parameter estimates from the functional ROIs identified from regression analysis were extracted and brain-behavior correlations were plotted for illustrative purposes. No outliers (defined as data points 3 standard deviations from the mean) were detected. Regions from brain-behavior analyses were initially identified for the whole brain at $p < 0.001$ uncorrected and an extend threshold of 5 contiguous voxels. All regions were then subject to FWE correction of $p < 0.05$ applied to the appropriate bilateral anatomical mask from the Wake Forest PickAtlas toolbox.

Amygdala connectivity analysis

The goal of this analysis was to examine functional connectivity between the amygdala and face-selective cortex as a function of phase (preconditioning, generalization test) and stimulus value (S1–S5). The seed region for this analysis was the left amygdala identified from the independent fear conditioning regression analysis. The face-selective region was the right fusiform gyrus identified from an independent functional localizer task. A GLM was created that modeled each individual trial as a separate covariate (Rissman et al., 2004). Correlations were computed for each subject by calculating the Pearson correlation coefficients between the mean parameter estimates from the seed region (amygdala) and fusiform gyrus. These values were converted from correlation coefficients to Z scores using the Fisher transform in order to make group inference. Inputs from the single subject level were input into a full factorial model as implemented in SPM 8, with phase (preconditioning, generalization test) and stimulus (S1–S5) as factors.

Results

Behavioral Results

Analysis of SCRs during the preconditioning phase, using ANOVA with S1–S5 as a repeated measure, revealed no difference in SCRs as a function of stimulus intensity value, $F_{4, 52} = 0.83$, $p = 0.51$ (Fig. 2a). These results demonstrate that SCRs were not sensitive to stimulus intensity value prior to fear learning, consistent with research demonstrating that static facial expressions are not inherently highly arousing (Anderson et al., 2006). Next, assessment of SCRs during the fear conditioning phase revealed that differential fear learning took place -- subjects expressed greater normalized SCRs to the CS+ (mean \pm SEM: 0.72 ± 0.02) than to the CS- (0.27 ± 0.02), $t_{(13)} = 9.20$, $p < 0.001$. Analysis of SCRs during the generalization test phase showed a main effect of stimulus intensity, $F_{4, 52} = 9.003$, $p < 0.001$, with a positive linear trend across the five face exemplars ($p < 0.001$) (Fig. 2a). This asymmetric gradient is in line with generalization based on emotional intensity because the distribution of SCRs show an increasing monotonic function that saturates but does not fall below the CS+ value (Dunsmoor et al., 2009; Ghirlanda and Enquist, 2003). The pre-to-post difference in SCR response profiles (Fig. 2b) provided the primary behavioral index of fear generalization. ANOVA of these fear generalization scores showed a main effect of stimulus intensity, $F_{4, 52} = 3.76$, $p = 0.009$, with a significant positive linear trend across the five face exemplars, $p = 0.01$. Whereas stimuli of less intensity than the CS+ showed an average pre-to-post decrease in SCRs, the CS+ and stimuli of high intensity showed an average increase following fear conditioning. However, behavioral responses across the emotional intensity gradient were variable overall, and these individual variations in behavior provide the basis for the neuroimaging regression analysis. See also Supplemental Table 1 for square root transformed and range corrected SCR analysis.

On each trial, subjects rated whether or not the morphed stimulus was expressing fear. Analysis of reaction time (RT) data revealed a main effect of morph increment during preconditioning, $F_{4,52} = 6.667$, $p < 0.001$, with both linear ($p = 0.01$) and quadratic ($p = 0.02$) trends (Fig. 2c). A main effect of stimulus type on RT was also observed during the

generalization test, $F_{4, 52} = 3.596, p = 0.01$. Post-hoc Bonferroni-corrected t-tests revealed that RTs were faster for the S4 versus the S3 both before and after fear conditioning, $p < .05$. See also Supplemental Fig. 1 for subjective ratings of fear expression. At the conclusion of the experimental session, subjects were asked to identify which morphed stimulus had been paired with the US. A chi-square test revealed that a significant percentage of subjects (71%, 10 out of 14) identified the S4, $\chi^2(4) = 23.86, p < 0.001$, from the array of stimuli (Fig. 2d). This false retrospective identification for the CS+ as a more emotionally intense stimulus confirms prior findings (Dunsmoor et al., 2009) of a co-variation bias (Öhman and Mineka, 2001). It is important to emphasize that this bias was revealed after the generalization test, so it is unknown whether the misidentification would be similar if probed during the training itself.

Imaging Results

We did not observe any brain regions that increased linearly as a function of fear intensity prior to fear conditioning. The neural correlates of differential fear learning were identified by comparing activity to the CS+ (S3) versus the CS- (S1), which revealed enhanced neural activity in the caudate, insula, and thalamus, extending into the PAG (Table 1). The reverse contrast (CS- versus CS+) revealed enhanced activity in the rostral anterior cingulate cortex (ACC) and subgenual ACC. These areas are commonly identified in human fMRI investigations of fear conditioning (Delgado et al., 2008; LaBar and Cabeza, 2006; Phelps and LeDoux, 2005). Activity related to the generalized stimuli (S2, S4, and S5) was examined by probing *a priori* functional ROIs identified from fear learning. First, to examine whether differential learning activity was specific to the CS+, contrasts were created that compared activity to each generalized stimulus against the CS-. In this way, the CS- served as a reference condition (Lissek et al., 2008; Lissek et al., 2010) for both learning (CS+ versus CS-) and generalization (S2 versus CS-; S4 versus CS-; S5 versus CS-). Activity related to these contrasts was extracted from preconditioning and the generalization test. Prior to fear conditioning, no significant effects were observed. Following fear conditioning, enhanced activity relative to the CS- was revealed for the S4 and S5 within the caudate and insula (Fig. 3a), suggesting that generalization-related activity in these regions was driven primarily by the intensity value of non-conditioned stimuli. Enhanced activity was also observed in the thalamus extending into the PAG for the S5 (see Supplemental Fig. 2). That these effects only emerged during the generalization test, and were not present at preconditioning, supports the idea that enhanced activity emerged as a consequence of fear learning. A similar analysis was conducted for the reverse contrast (regions that preferentially signaled the CS- relative to the CS+ during learning). No significant differences in activity to the generalized stimuli were observed in these regions during preconditioning, but enhanced activity was observed to the S2 versus the CS+ within the rostral and subgenual ACC during the generalization test, indicating that these regions exhibit a reversed intensity-based generalization gradient perhaps constituting a 'safety' signal (Fig. 3b).

Brain-behavior correlation results

Analysis of brain-behavior correlations focused on the relationship between brain activity and SCRs during the fear conditioning and generalization test phases. In line with prior findings (Cheng et al., 2006; Knight et al., 2005), activity in the left amygdala ($x = -30, y = 4, z = -26; 256 \text{ mm}^3$) was positively correlated with behavioral measures of differential learning during fear acquisition, $r_{(26)} = 0.52, p = 0.005$ (see Supplemental Fig. 3). This result confirms that amygdala activity was related to measures of autonomic arousal that reflect the acquisition of conditioned fear. Based on findings from our previous behavioral study showing that the S4 in particular evoked considerable fear generalization and induced a false memory as the CS+ (Dunsmoor et al., 2009), we predicted that brain-behavior

correlations in fear generalization would be revealed for this stimulus. Consistent with predictions, positive correlations were found in the left amygdala ($x = -26, y = 4, z = -18; 1280 \text{ mm}^3$) and both right ($x = 42, y = 8, z = -6; 6400 \text{ mm}^3$) and left ($x = -42, y = -12, z = -2; 832 \text{ mm}^3$) insula (Fig. 4) for S4 presentations. These results show that the change in fear expression to the S4 is reflected by increases in activity within regions commonly implicated in conditioned fear learning. For comparison, no regions showed positive brain-behavior correlations for the S2, a stimulus that shared considerable perceptual overlap with the CS+ but contained less emotional intensity value. These results support the idea that emotional intensity, and not perceptual overlap, drove fear generalization. Importantly, we determined that neural responses in these regions were linked to the physiological expression of fear -- when a direct S4 vs. S2 contrast was conducted irrespective of SCR levels, there were no significant results. Thus, activity in the amygdala and insula increased as a function of psychophysiological arousal to a generalized stimulus following fear conditioning. To assess the extent to which correlations identified for the S4 were selective to this stimulus, we conducted a contrast of brain-behavior correlations between the S4 and all other stimulus values, including the CS+. This analysis revealed activity in the left dorsal amygdala ($x = -26, y = 4, z = -18; 192 \text{ mm}^3$), indicating that the amygdala showed the strongest brain-behavior effects for the S4, which was falsely identified as the CS+ by a majority of participants after the generalization test.

Amygdala connectivity results

Finally, because the amygdala is important for modulating activity in sensory processing regions to more effectively detect and evaluate affective stimuli (Armony and Dolan, 2002; Vuilleumier, 2005), we predicted that connectivity between the amygdala and extrastriate visual cortex would be important for supporting generalization from a learned threat to related stimuli. To examine amygdala connectivity, we used an independent functional ROI of the left amygdala identified from the fear conditioning regression analysis ($x = -30, y = 4, z = -26$) as a seed region for a trial-by-trial functional connectivity analysis of the preconditioning and generalization test phases (Rissman et al., 2004). We focused on connectivity between the amygdala and a face-selective region in the fusiform gyrus (FFG) identified from the independent functional localizer task ($x = 38, y = 64, z = 22$). During preconditioning, there was no difference in functional connectivity as a function of stimulus intensity value ($F_{4, 52} = 2.02, p = 0.11$). During the generalization test, there was a main effect of stimulus intensity on amygdala-FFG connectivity ($F_{4, 52} = 3.856, p = 0.008$) with a linear ($p = 0.006$) and quadratic ($p = 0.036$) trend (Fig. 5). A significant phase (preconditioning, generalization test) by stimulus interaction was observed for the CS+ versus the CS- ($F_{1, 13} = 7.449, p = 0.017$), indicating a learning-related change in amygdala connectivity following fear conditioning. A significant phase by stimulus interaction was also observed for the S4 versus the CS- ($F_{1, 13} = 7.454, p = 0.017$), but not for the other generalized stimuli ($p > 0.1$). Prior research has also shown enhanced amygdala-fusiform connectivity in anxious individuals while viewing phobia-relevant stimuli (Ahs et al., 2009), and in 5-HTT short allele homozygotes while viewing fearful faces (Surguladze et al., 2008). To determine whether individual differences in trait anxiety would correlate with increases in amygdala connectivity following fear learning, we correlated the change in connectivity for the CS+ and S4 from pre- to post-conditioning with each participant's trait anxiety level. This analysis was limited to the CS+ and S4, as these were the only two stimuli to show an increase in amygdala-FFG connectivity following fear conditioning. We found that individual differences in trait anxiety were positively related to the pre-to-post fear learning increase in amygdala connectivity for the S4 ($r_{(13)} = 0.62, p = 0.018$), but not the CS+ ($r_{(13)} = -0.12, p = 0.69$) (Fig. 5). Individual differences in trait anxiety were also negatively related to amygdala activity indexing differential learning (CS+ versus CS-)

during early acquisition training due to a generalized response to the CS⁻ (see Supplemental Fig. 3). No other fMRI results were moderated by trait anxiety.

DISCUSSION

Fear generalization is a common occurrence following a highly aversive experience that is exaggerated in some individuals. Here, subjects who underwent classical fear conditioning to a moderately intense emotional face tended to express fear to stimuli that resembled a learned threat but contained greater emotional intensity. Generalized fear expression was mirrored by a false memory of the CS⁺ as a more intense stimulus than it actually was, indicating a retrospective bias in estimating threat value. fMRI results confirmed several key hypotheses. First, regions involved in the acquisition of differential fear learning also showed responses to generalized stimuli as a function of emotional intensity during the generalization test. Second, activity in the amygdala and insula correlated with individual variability in physiological arousal measures related to the expression of fear generalization. Finally, functional connectivity between the amygdala and the face-selective region of the fusiform gyrus was enhanced during the generalization test to a non-conditioned stimulus of high intensity, and this increase in connectivity from pre-to-post fear learning was correlated with trait anxiety. Combined, these results provide new evidence that regions with an established role in fear learning are important for generalizing fear as well, and indicate that the neural substrates of fear generalization may be best understood on the basis of individual differences in behavior and anxiety levels.

Present findings for the striatum, insula, and thalamus/PAG extend previous knowledge concerning the role of these regions in aversive learning. For instance, prior research has shown that the striatum serves a role in learning to predict and fear an aversive stimulus (Delgado et al., 2008). The striatum has also been shown to flexibly adapt when contingencies surrounding CS-US pairing are reversed (Schiller and Delgado, 2010). The thalamus is a key region involved in aversive learning that provides sensory information to the amygdala directly and indirectly (LeDoux, 1996), and this region frequently shows enhanced responses to a CS⁺ in human imaging studies of fear conditioning (LaBar and Cabeza, 2006). The periaqueductal gray supports physiological responses to aversively conditioned stimuli (LeDoux et al., 1988) and is implicated as part of a core functional group of limbic regions involved in affective processes (Kober et al., 2008). Likewise, the insula serves a role in physiological responses to affectively significant stimuli, and prior neuroimaging research has shown that uncertainty and anticipation for receiving an aversive stimulus enhances insula activity (Berns et al., 2006; Dunsmoor et al., 2007). In all, generalized patterns of activity in the striatum, insula, and thalamus/PAG imply that areas involved in fear learning are not specific to an aversively conditioned stimulus. These regions may be important for responding to stimuli that share properties with a learned threat in order to adaptively react to potential threats from the environment. During the generalization test, a non-conditioned stimulus of lower emotional intensity, which most closely resembled the 'safe' CS⁻, showed enhanced activation in the rostral and subgenual ACC when compared to the CS⁺. Prior research has shown that these regions are important for regulating emotional responses (Phelps and LeDoux, 2005; Sotres-Bayon et al., 2004). For instance, the vmPFC is involved during the recall of learned extinction (Milad et al., 2007; Phelps et al., 2004), and may mediate extinction by inhibiting activity in the amygdala (Maren and Quirk, 2004). Human neuroimaging of fear conditioning has shown that during fear acquisition responses to the CS⁺ often decrease in the vmPFC, whereas responses to the CS⁻ increase (Schiller and Delgado, 2010; Schiller et al., 2008). This pattern emerged in the present study, such that responses to the CS⁻ were significantly enhanced relative to the CS⁺, suggesting that the CS⁻ evoked activity related to regulatory processes. Given that the CS⁻ and S2 were both associated with relative decreases in arousal following fear

conditioning, it is noteworthy that the S2 was the only other stimulus to evoke enhanced activity within these regions. This finding suggests that the regulation of fear can generalize beyond a learned safety signal to other stimuli.

The regression analysis of brain-behavior interactions yielded crucial findings on the relationship between the change in arousal following fear learning and neural activity in the amygdala and insula. Animal models of fear conditioning have consistently implicated the amygdala as the final common pathway involved in fear learning and fear expression (LeDoux, 2000). The precise role for the amygdala in fear generalization is not known, but the amygdala may initiate rapid generalized fear responses by way of direct connections with the sensory thalamus (Han et al., 2008) as neurons in the sensory thalamus are broadly tuned (Bordi and LeDoux, 1994). Prior neuroimaging research has shown that amygdala activity is related to the production of conditioned SCRs, and does not merely track the presence of a stimulus with threat value (Cheng et al., 2006; Knight et al., 2005). Consistent with these previous findings, amygdala activity was related with differential SCRs (CS+ versus CS-) during the fear conditioning phase, and selectively tracked behavioral measures related to increases in arousal evoked by the S4 during the generalization test. This result extends previous fMRI findings that have demonstrated correlations in fear expression and amygdala activity to the conditioned stimulus (Cheng et al., 2006; Knight et al., 2005). Activity in the insula was also correlated with generalized fear expression. Previous fMRI studies have linked the insula to visceral and emotional processing in general (Phan et al., 2002). Theories have emerged proposing that the role of the insula is to integrate physiological information concerning internal bodily states to inform psychological awareness and decision making (Craig, 2009) which, in the context of the present study, may relate to assessing the predictive or affective value of each face exemplar. It is important to note that the behavioral metric derived to assess fear generalization captured the *change* in the proportional response to each stimulus following fear conditioning, and did not reflect pure SCR magnitude to each stimulus during the generalization test (c.f. Schiller and Delgado, 2010). This distinction is critical, as this measure provides a novel way to assess how subjects change in their psychophysiological response profile from pre-to-post fear learning. This approach is in contrast to several neuroimaging investigations that have shown a relationship between the direct production of SCRs and brain activity (e.g. Critchley et al., 2000). The present method helps to ensure that changes in behavior are due to the intervening fear learning phase (Weinberger, 2007), either through associative learning or non-associative sensitization processes [see Dunsmoor et al. (2009)]. Therefore, these findings provide a key insight into the relationship between brain activity and behavioral responses to generalized threats following an episode of fear learning.

Functional connectivity analysis revealed increased amygdala-FFG coupling during the generalization test for the CS+ and S4. The amygdala may serve a role in modulating activity in cortical regions to stimuli that have acquired affective significance, for instance through fear conditioning (Armony and Dolan, 2002). Prior research has shown that affectively salient faces preferentially engage the amygdala and FFG (Vuilleumier and Pourtois, 2007). In the present study, functional connectivity between the amygdala and FFG was undifferentiated prior to fear conditioning. The lack of differential amygdala effects during preconditioning for high versus low value fearful faces is in line with prior findings showing that static images of fearful (and angry) faces do not evoke larger responses in the amygdala than emotionally neutral faces (Fitzgerald et al., 2006; LaBar et al., 2003) Following fear conditioning, amygdala-FFG connectivity was characterized by learning-related and generalization-related effects -- connectivity was enhanced to both the CS+ and the S4 following the fear conditioning phase. These connections may facilitate fear responses by enhancing the sensory representation of stimuli related to a learned threat. The finding that correlations between trait anxiety and increases in amygdala connectivity for the

S4 is informative for the way in which amygdala connectivity contributes to stimulus processing in high anxious individuals. The association between anxiety levels and brain activity has been explored across a number of emotional processing tasks (Etkin and Wager, 2007), and high anxiety levels are frequently associated with amygdala activity evoked by negative stimuli (Bishop, 2007). The present finding is consistent with a study reporting amygdala-FFG connectivity in phobic patients viewing phobic stimuli (Ahs et al., 2009). If amygdala connectivity serves to facilitate responses to stimuli that share properties with a feared stimulus, then this pathway might be related to overgeneralization of fears in anxiety disorders. It is interesting that connectivity was not enhanced for the S5 following fear conditioning, considering that the S5 contained the greatest degree of fear expression. This may indicate that functional connectivity is related predominately to the behavioral measures of fear conditioning (i.e., SCRs and retrospective CS+ identification), which show a bias in favor of the S4 above all other stimulus values.

These findings may be interpreted from a number of theoretical perspectives. First, a gradient-interaction theory of stimulus generalization (Spence, 1937) argues that gradients of excitation and inhibition form around the CS+ and CS-, respectively. The summation of these gradients leads to a shift in responses to a value further from the CS-, which could explain the peak-shift in behavioral and neural responses to the S4 and S5 but not the S2. However, gradient-interaction theory has not received strong support from empirical studies of stimulus generalization, as it has been repeatedly shown that inhibition and excitation do not generalize in the same manner (Ghirlanda, 2002; Lissek et al., 2008; Rescorla, 2006). Furthermore, our previous behavioral findings using this design argues against gradient-interaction theory as the root cause of emotion-based intensity generalization (Dunsmoor et al., 2009). In this study, discriminatory fear learning using the most intense stimulus (S5) as the CS- resulted in a sharper generalization gradient around the CS+ (S3), but did not cause a reverse gradient (i.e., greater responses to the S1 and S2 versus the S4 and S5). We also note that the BOLD signal in fMRI does not permit a direct test of inhibitory vs. excitatory gradient-interaction effects. Nonetheless, our fMRI analysis in the present study revealed that brain regions that respond selectively to the CS+ or CS- during the initial learning also generalized their response accordingly, so brain regions that support discrimination learning do make some contribution to the generalization gradient. Interestingly, the results show that effects reported for the amygdala do not peak to the CS+ value used during initial learning but instead peaks to the S4, so the amygdala's role in fear conditioning may be underestimated by using its response to the CS+ as the sole index of learning.

Another interpretation comes from an elemental associative learning model of stimulus generalization (McLaren and Mackintosh, 2002). In this view, elements that predict the US accrue associative value while elements that do not predict the US lose associative value over the course of conditioning (Rescorla and Wagner, 1972). During the generalization test, similarity to the CS+ is measured by those shared elements that have gained associative value (in this case, features related to fear expression) while other perceptual elements (in this case, features related to identity) are given less weight (McLaren and Mackintosh, 2002). A peak shift can occur if a non-CS contains more of those associative elements than the CS+ itself. However, according to this model the S5 would have evoked more generalization than the CS+ and S4, as it contained the most amount of emotional intensity. Moreover, an elemental associative model does not fully accord to our previous behavioral findings which failed to show a reverse intensity based gradient (Dunsmoor et al., 2009). We propose that a stimulus intensity model (Ghirlanda, 2002; Ghirlanda and Enquist, 2003) may best explain the present results. Intensity effects are marked by a response bias (i.e., peak shift) to generalize learned behaviors towards novel stimuli that are somewhat more intense than the CS+ (Ghirlanda and Enquist, 2003). Thus, an intrinsically intense non-CS may be more likely to evoke a heightened fear response after an episode of fear learning than a

stimulus that is similar to but less intense than the CS+. Overall, further brain imaging research will be needed to establish any model of generalization as neurobiologically plausible.

Interestingly, we observed a co-variation bias (or “illusory correlation”) along an emotional intensity dimension, such that the majority of subjects falsely identified a more intense face as the CS+ in a post-experimental test of awareness. Previous paradigms using fear-relevant (e.g., snakes and spiders) and fear-irrelevant (e.g., flowers and mushrooms) cues have shown that subjects often mistakenly conclude that fear-relevant cues were paired with an aversive US at a higher rate, when in fact both classes of stimuli were equally paired with the US (Öhman and Mineka, 2001). This bias is more pronounced in high anxious individuals (Tomarken et al., 1989). The present result is in keeping with these previous findings, and suggests that even healthy adults are biased towards remembering details of a fear learning experience as more emotionally intense than they actually were. Alternatively, this retrospective bias in threat estimation could indicate that subjects were unable to discriminate between the CS+ and S4. However, our prior psychometric studies using these neutral-to-morphs have shown that healthy subjects can readily discriminate between morph values that are even more subtle than those chosen for the present study (Graham et al., 2007; Thomas et al., 2007), and RT data from the present study reveal a clear difference in the time to classify the S3 and S4. Therefore we do not believe that the retrospective bias was indicative of a perceptual confusion during learning itself but future studies could explicitly test awareness intermittent with learning. The relationship between generalization and discrimination has been the matter of historical debate in the conditioned learning literature (see for example Hull, 1943; Lashley and Wade, 1946). More contemporary views on the relationship between discrimination and generalization (e.g. Shepard, 1987) are in part informed by prevailing evidence that generalization can occur despite the capacity to recognize the difference between exemplars (Guttman and Kalish, 1956). That is, generalization along a dimension often follows an orderly gradient to stimuli that are both confusable and discriminable (Pavlov, 1927). Notably, empirical research on stimulus generalization comes predominately from studies of appetitive instrumental learning. Future studies that focus on classically conditioned fear behaviors will be needed to fully address the role of discriminatory processes in aversive learning, and to examine whether the present findings extend to affectively neutral non-intensity dimensions.

A limitation of the present study is the use of a single stimulus dimension (fear intensity), whereas generalization can occur along any dimension. In the animal literature, the use of a single dimension is commonly used when exploring effects of intra-dimensional discrimination training (Honig and Urcuioli, 1981). It is for this reason that we employed only a single dimension in the present study, thus ensuring that intra-dimensional changes in behavioral and neural responses could be attributed to the degree of emotional expression, and not other factors related to identity. Our prior fMRI work using these face morphs has further shown that dynamic changes in identity and emotional expression are partly dissociable in the brain (LaBar et al., 2003). Thus, additional research is warranted to determine how generalization is mediated along other featural dimensions.

In conclusion, the neural and behavioral systems involved in processing and reacting to feared stimuli are becoming increasingly well delineated across species, driven in large part by a desire to better inform models of clinical anxiety disorders. The laboratory study of fear learning has typically involved a systematic examination of the processes involved in acquiring, expressing, and extinguishing fears to a specific stimulus. To understand anxiety disorders marked by heightened fear responses, however, it is necessary to explore the processes involved in the generalization of fear to a wider range of stimuli, especially given that a feared stimulus can be encountered in multiple forms (Shepard, 1987). Results from

the present investigation demonstrate that regions involved in fear learning are not always specific to a learned threat. Moreover, individual differences in intensity-based generalization suggest a dynamic interplay between corticolimbic-autonomic coupling during fear generalization, and underscores the importance for interpreting brain activity by behavioral measures. Lastly, analysis of functional connectivity between the amygdala and FFG suggest that the amygdala may be important for modulating the sensory representation of stimuli that approximate a learned threat, and that heightened amygdala connectivity may be associated with overgeneralization of fears for individuals with heightened anxiety. Collectively, these results provide novel methodological approaches and insights into the neural basis of fear learning and generalization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Matthew Fecteau for assistance with the psychophysiological equipment. This project was supported by NSF grant 0745919 and NIH grant 2 P01 NS041328.

References

- Ahs F, Pissioti A, Michelgard A, Frans O, Furmark T, Appel L, Fredrikson M. Disentangling the web of fear: Amygdala reactivity and functional connectivity in spider and snake phobia. *Psychiatry Research-Neuroimaging*. 2009; 172:103–108.
- Amaral DG, Behniea H, Kelly JL. Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. *Neuroscience*. 2003; 118:1099–1120. [PubMed: 12732254]
- Anderson AK, Yamaguchi Y, Grabski W, Lacka D. Emotional memories are not all created equal: Evidence for selective memory enhancement. *Learning & Memory*. 2006; 13:711–718. [PubMed: 17101871]
- Armory JL, Dolan RJ. Modulation of spatial attention by fear-conditioned stimuli: an event-related fMRI study. *Neuropsychologia*. 2002; 40:817–826. [PubMed: 11900732]
- Armory JL, Servan-Schreiber D, Romanski LM, Cohen JD, LeDoux JE. Stimulus generalization of fear responses: Effects of auditory cortex lesions in a computational model and in rats. *Cerebral Cortex*. 1997; 7:157–165. [PubMed: 9087823]
- Bergado-Acosta JR, Sangha S, Narayanan RT, Obata K, Pape HC, Stork O. Critical role of the 65-kDa isoform of glutamic acid decarboxylase in consolidation and generalization of Pavlovian fear memory. *Learning & Memory*. 2008; 15:163–171. [PubMed: 18323571]
- Berns GS, Chappelow J, Cekic M, Zink CF, Pagnoni G, Martin-Skurski ME. Neurobiological substrates of dread. *Science*. 2006; 312:754–758. [PubMed: 16675703]
- Bishop SJ. Neurocognitive mechanisms of anxiety: an integrative account. *Trends in Cognitive Sciences*. 2007; 11:307–316. [PubMed: 17553730]
- Blough DS. Steady-state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology*. 1975; 104:3–21. [PubMed: 1170274]
- Bordi F, LeDoux JE. Response properties of single units in areas of rat auditory thalamus that project to the amygdala. 1. acoustic discharge patterns and frequency receptive-fields. *Experimental Brain Research*. 1994; 98:261–274.
- Cheng DT, Knight DC, Smith CN, Helmstetter FJ. Human amygdala activity during the expression of fear responses. *Behavioral Neuroscience*. 2006; 120:1187–1195. [PubMed: 17201461]
- Craig AD. How do you feel - now? The anterior insula and human awareness. *Nature Reviews Neuroscience*. 2009; 10:59–70.
- Critchley HD, Elliott R, Mathias CJ, Dolan RJ. Neural activity relating to generation and representation of galvanic skin conductance responses: A functional magnetic resonance imaging study. *Journal of Neuroscience*. 2000; 20:3033–3040. [PubMed: 10751455]

- Davis M. The role of the amygdala in fear and anxiety. *Annual Review of Neuroscience*. 1992; 15:353–375.
- Delgado MR, Li J, Schiller D, Phelps EA. The role of the striatum in aversive learning and aversive prediction errors. *Philosophical Transactions of the Royal Society B-Biological Sciences*. 2008; 363:3787–3800.
- Dunsmoor JE, Bandettini PA, Knight DC. Impact of continuous versus intermittent CS-UCS pairing on human brain activation during Pavlovian fear conditioning. *Behavioral Neuroscience*. 2007; 121:635–642. [PubMed: 17663589]
- Dunsmoor JE, Mitroff SR, LaBar KS. Generalization of conditioned fear along a dimension of increasing fear intensity. *Learning & Memory*. 2009; 16:460–469. [PubMed: 19553384]
- Duvarci S, Bauer EP, Pare D. The Bed Nucleus of the Stria Terminalis Mediates Inter-individual Variations in Anxiety and Fear. *Journal of Neuroscience*. 2009; 29:10357–10361. [PubMed: 19692610]
- Ekman P, Friesen WV. Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*. 1976; 1:56–75.
- Etkin A, Wager TD. Functional neuroimaging of anxiety: A meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *American Journal of Psychiatry*. 2007; 164:1476–1488. [PubMed: 17898336]
- Fitzgerald DA, Angstadt M, Jelsone LM, Nathan PJ, Phan KL. Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *Neuroimage*. 2006; 30:1441–1448. [PubMed: 16368249]
- Ghirlanda S. Intensity generalization: Physiology and modelling of a neglected topic. *Journal of Theoretical Biology*. 2002; 214:389–404. [PubMed: 11846597]
- Ghirlanda S, Enquist M. A century of generalization. *Animal Behaviour*. 2003; 66:15–36.
- Graham R, Devinsky O, LaBar KS. Quantifying deficits in the perception of fear and anger in morphed facial expressions after bilateral amygdala damage. *Neuropsychologia*. 2007; 45:42–54. [PubMed: 16806315]
- Grinband J, Wager TD, Lindquist M, Ferrera VP, Hirsch J. Detection of time-varying signals in event-related fMRI designs. *Neuroimage*. 2008; 43:509–520. [PubMed: 18775784]
- Guttman N, Kalish HI. Discriminability and stimulus-generalization. *Journal of Experimental Psychology*. 1956; 51:79–88. [PubMed: 13286444]
- Han JH, Yiu AP, Cole CJ, Hsiang HL, Neve RL, Josselyn SA. Increasing CREB in the auditory thalamus enhances memory and generalization of auditory conditioned fear. *Learning & Memory*. 2008; 15:443–453. [PubMed: 18519545]
- Honig WK, Urcuioli PJ. The legacy of Guttman and Kalish (1956) - 25 years of research on stimulus-generalization. *Journal of the Experimental Analysis of Behavior*. 1981; 36:405–445. [PubMed: 16812256]
- Hull, CL. *Principles of Behavior*. Appleton-Century-Crofts; New York: 1943.
- Jenkins HM, Harrison RH. Effect of discrimination-training on auditory generalization. *Journal of Experimental Psychology*. 1960; 59:246–253. [PubMed: 14407066]
- Knight DC, Nguyen HT, Bandettini PA. The role of the human amygdala in the production of conditioned fear responses. *Neuroimage*. 2005; 26:1193–1200. [PubMed: 15961053]
- Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD. Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage*. 2008; 42:998–1031. [PubMed: 18579414]
- LaBar KS, Cabeza R. Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*. 2006; 7:54–64.
- LaBar KS, Crupain MJ, Voyvodic JT, McCarthy G. Dynamic perception of facial affect and identity in the human brain. *Cerebral Cortex*. 2003; 13:1023–1033. [PubMed: 12967919]
- Lashley KS, Wade M. The Pavlovian theory of generalization. *Psychological Review*. 1946; 53:72–87. [PubMed: 21023320]
- LeDoux, JE. *The Emotional Brain*. Simon and Schuster; New York: 1996.
- LeDoux JE. Emotion circuits in the brain. *Annual Review of Neuroscience*. 2000; 23:155–184.

- LeDoux JE, Iwata J, Cicchetti P, Reis DJ. Different projections of the central amygdaloid nucleus mediate autonomic and behavioral-correlates of conditioned fear. *Journal of Neuroscience*. 1988; 8:2517–2529. [PubMed: 2854842]
- Lissek S, Biggs AL, Rabin SJ, Cornwell BR, Alvarez RP, Pine DS, Grillon C. Generalization of conditioned fear-potentiated startle in humans: Experimental validation and clinical relevance. *Behaviour Research and Therapy*. 2008; 46:678–687. [PubMed: 18394587]
- Lissek S, Rabin S, Heller RE, Lukenbaugh D, Geraci M, Pine DS, Grillon C. Overgeneralization of Conditioned Fear as a Pathogenic Marker of Panic Disorder. *American Journal of Psychiatry*. 2010; 167:47–55. [PubMed: 19917595]
- Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*. 2003; 19:1233–1239. [PubMed: 12880848]
- Maren S, Quirk GJ. Neuronal signalling of fear memory. *Nature Reviews Neuroscience*. 2004; 5:844–852.
- McLaren IPL, Mackintosh NJ. Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior*. 2002; 30:177–200. [PubMed: 12391785]
- Milad MR, Wright CI, Orr SP, Pitman RK, Quirk GJ, Rauch SL. Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biological Psychiatry*. 2007; 62:446–454. [PubMed: 17217927]
- Morris JP, Green SR, Marion B, McCarthy G. Guided saccades modulate face- and body-sensitive activation in the occipitotemporal cortex during social perception. *Brain and Cognition*. 2008; 67:254–263. [PubMed: 18346831]
- Öhman A, Mineka S. Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*. 2001; 108:483–522. [PubMed: 11488376]
- Pavlov, IP. *Conditioned Reflexes*. Oxford University Press; London: 1927.
- Phan KL, Wager T, Taylor SF, Liberzon I. Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*. 2002; 16:331–348. [PubMed: 12030820]
- Phelps EA, Delgado MR, Nearing KI, LeDoux JE. Extinction learning in humans: Role of the amygdala and vmPFC. *Neuron*. 2004; 43:897–905. [PubMed: 15363399]
- Phelps EA, LeDoux JE. Contributions of the amygdala to emotion processing: From animal models to human behavior. *Neuron*. 2005; 48:175–187. [PubMed: 16242399]
- Rescorla RA. Stimulus generalization of excitation and inhibition. *Quarterly Journal of Experimental Psychology*. 2006; 59:53–67.
- Rescorla, RA.; Wagner, AR. *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement*. Appleton-Century-Crofts; 1972.
- Rissman J, Gazzaley A, D’Esposito M. Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage*. 2004; 23:752–763. [PubMed: 15488425]
- Schiller D, Delgado MR. Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends in Cognitive Sciences*. 2010; 14:268–276. [PubMed: 20493762]
- Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA. From Fear to Safety and Back: Reversal of Fear in the Human Brain. *Journal of Neuroscience*. 2008; 28:11517–11525. [PubMed: 18987188]
- Shaban H, Humeau Y, Herry C, Cassasus G, Shigemoto R, Ciochi S, Barbieri S, van der Putten H, Kaupmann K, Bettler B, Luthi A. Generalization of amygdala LTP and conditioned fear in the absence of presynaptic inhibition. *Nature Neuroscience*. 2006; 9:1028–1035.
- Shepard RN. Toward a universal law of generalization for psychological scienc. *Science*. 1987; 237:1317–1323. [PubMed: 3629243]
- Sotres-Bayon F, Bush DEA, LeDoux JE. Emotional perseveration: An update on prefrontal-amygdala interactions in fear extinction. *Learning & Memory*. 2004; 11:525–535. [PubMed: 15466303]
- Spence KW. The differential response in animals to stimuli varying within a single dimension. *Psychological Review*. 1937; 44:430–444.
- Spielberger, CD. *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press; Palo Alto, California: 1983.

- Surguladze SA, Elkin A, Ecker C, Kalidindi S, Corsico A, Giampietro V, Lawrence N, Deeley Q, Murphy DGM, Kucharska-Pietura K, Russell TA, McGuffin P, Murray R, Phillips ML. Genetic variation in the serotonin transporter modulates neural system-wide response to fearful faces. *Genes Brain and Behavior*. 2008; 7:543–551.
- Thomas LA, De Bellis MD, Graham R, LaBar KS. Development of emotional facial recognition in late childhood and adolescence. *Developmental Science*. 2007; 10:547–558. [PubMed: 17683341]
- Tomarken AJ, Cook M, Mineka S. Fear-relevant selective associations and covariation bias. *Journal of Abnormal Psychology*. 1989; 98:381–394. [PubMed: 2592672]
- Vuilleumier P. How brains beware: neural mechanisms of emotional attention. *Trends in Cognitive Sciences*. 2005; 9:585–594. [PubMed: 16289871]
- Vuilleumier P, Pourtois G. Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia*. 2007; 45:174–194. [PubMed: 16854439]
- Weinberger NM. Associative representational plasticity in the auditory cortex: A synthesis of two disciplines. *Learning & Memory*. 2007; 14:1–16. [PubMed: 17202426]

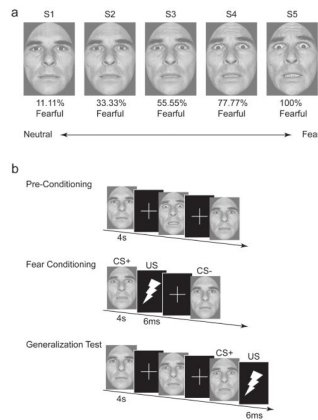


Figure 1. Stimulus set and task design

(a) The stimulus dimension consisted of 5 images, of the same identity, morphed between neutral and fearful endpoints. (b) The task involved rating whether each face was or was not expressing fear by pressing one of two buttons. In the first phase (preconditioning), subjects saw each morph increment in the absence of the US. Fear learning involved repeated pairings of the S3 (CS+) with an electrical shock US, and the S1 (CS-) unreinforced. The generalization test followed fear learning and involved presentation of each of the morph increments. The CS+ was intermittently reinforced during the generalization test. Images are not to scale.

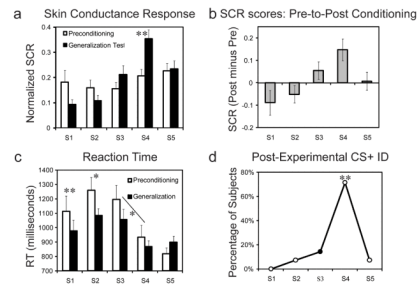


Figure 2. Behavioral results

(a) Mean normalized SCRs from preconditioning and the generalization test show that response output was undifferentiated along the neutral-to-fearful continuum during preconditioning (white bars) and shifted towards stimuli of high emotional intensity during the generalization test (black bars). (b) Difference scores, reflecting the change in response output from preconditioning to generalization test, show an average decrease in psychophysiological responses to the S1 and S2 and an average increase to the S3, S4, and S5. (c) Mean reaction times show that subjects were fastest to categorize the S4 and S5 as fearful. (d) A majority of subjects (71%) mistakenly identified the S4 as the CS+ indicating a strong illusory correlation. Error bars reflect standard error of the mean (SEM), (*) denote significant differences ($p < 0.05$) and (***) at $p < 0.01$.

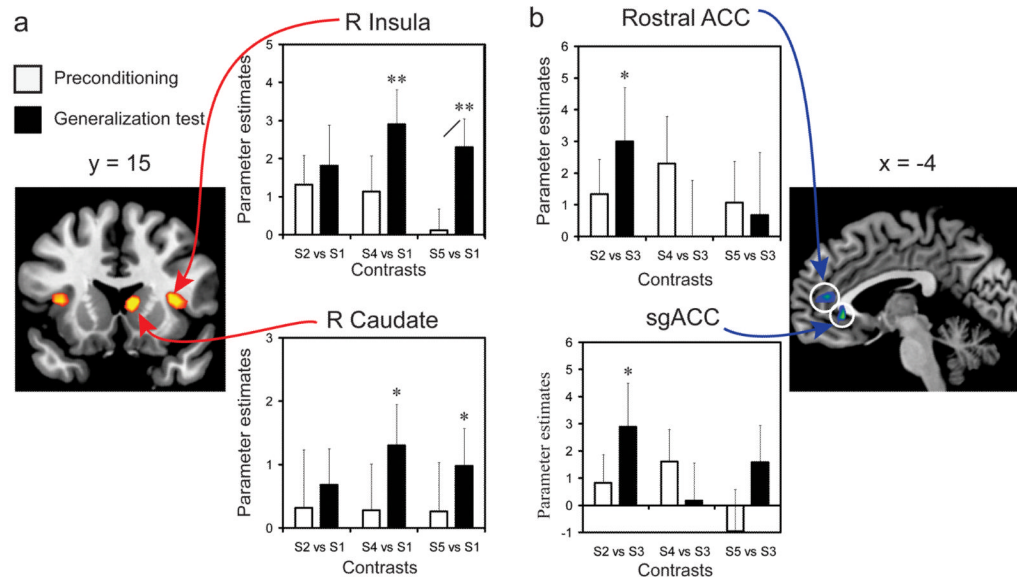


Figure 3. Brain regions involved in differential fear learning show generalized patterns of activity

(a) Regions of interest were identified by contrasts of CS+ versus CS-. Contrasts of each non-conditioned stimulus (S2, S4, and S5) versus the CS- revealed significant differential activity for the S4 and S5 versus the CS- (one-sample t tests) in the right caudate ($x = 10$, $y = 4$, $z = 2$) and right insula ($x = 34$, $y = 12$, $z = 6$) during the generalization test. Activity was significantly enhanced for the S5 contrast from preconditioning to the generalization test (paired-samples t test) in the right caudate. **(b)** Regions of interest were identified by contrasts of CS- versus CS+, and revealed activity in the subgenual ($x = -2$, $y = 32$, $z = -6$) and rostral ($x = -6$, $y = 44$, $z = 10$) ACC. Contrasts comparing non-conditioned stimuli versus the CS+ showed significant differential activity between the S2 and CS+ during the generalization test. All regions were small volume corrected with a family-wise-error $p < 0.05$, shown here at $p < 0.001$ (uncorrected) for visualization purposes. Error bars reflect SEM. For one-sample t tests, (*) denote significant differences ($p < 0.05$) and (**) at $p < 0.01$. For paired-samples t test, (/) denotes significant differences ($p < 0.05$).

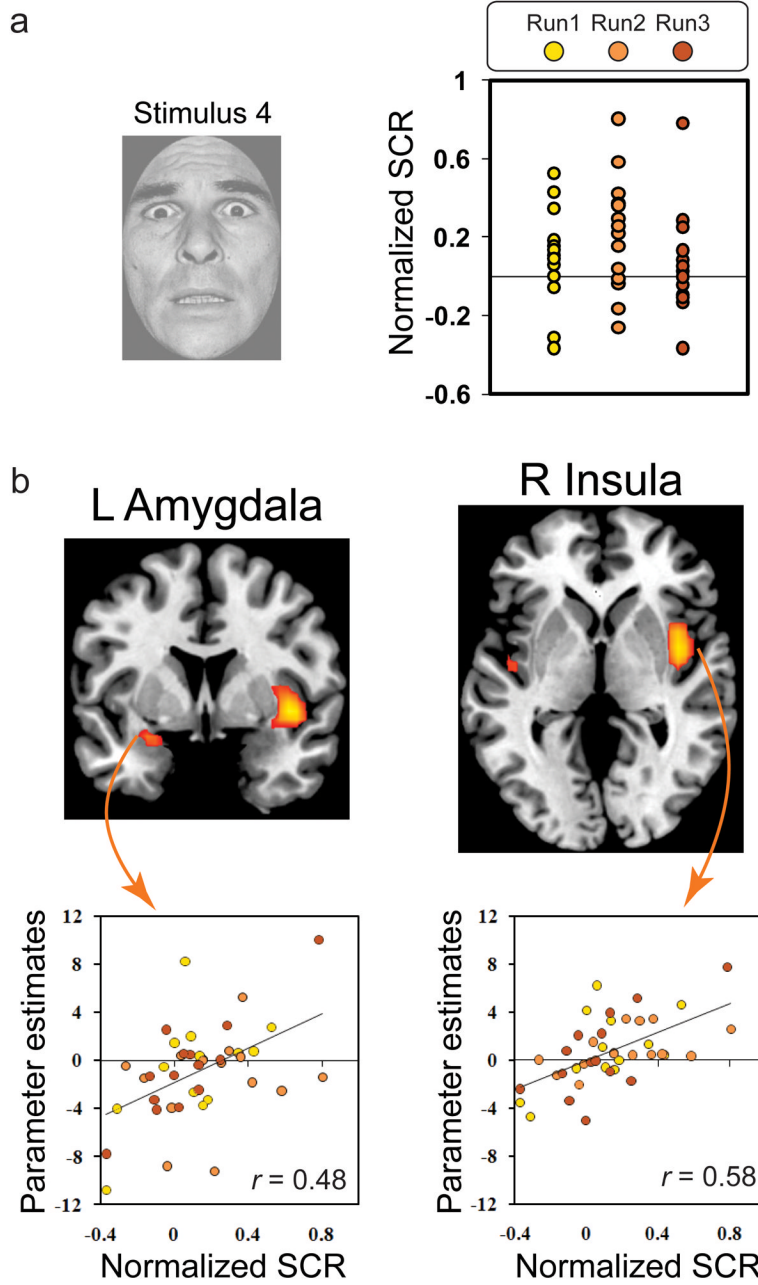


Figure 4. Brain-behavior interactions in fear generalization
(a) Normalized SCR scores for the S4 illustrate the variability in response across subjects and across runs. **(b)** Correlations between SCR scores and brain activity were revealed in the insula and amygdala, such that increases in arousal from pre-to-post fear learning were associated with increases in brain activity in these regions.

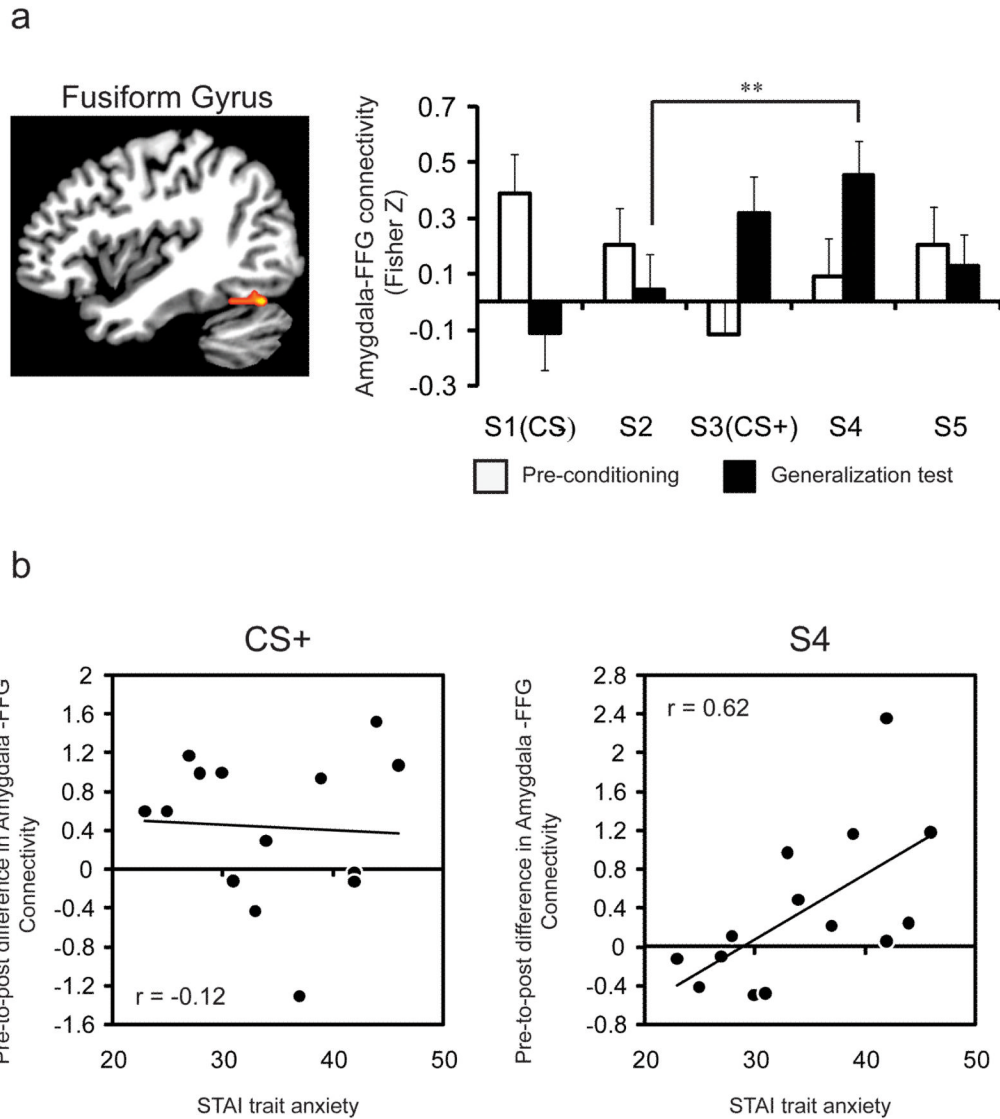


Figure 5. Learning-related and generalization-related increases in amygdala-visual cortex activity its relation to trait anxiety

(a) Single trial connectivity analysis, using the amygdala as a seed region, showed learning-related changes in amygdala connectivity with a face-selective region within the fusiform gyrus that was identified from an independent functional localizer task. Connectivity between these regions was undifferentiated prior to fear learning, and showed a linear increase in connectivity during the generalization test. ANOVA revealed a phase (preconditioning, generalization test) by stimulus interaction for the CS+ versus the CS-, as well as for the S4 versus the S2. (b) Increases in functional connectivity on S4 trials, from preconditioning to the generalization test, were positively related with trait anxiety scores ($r_{(13)} = 0.62, p = 0.018$). Increases in amygdala-FFG connectivity were not related with anxiety scores for the CS+ ($r_{(13)} = -0.12, p = 0.69$).

Table 1

Fear learning-related activity

Region	Hemisphere	MNI coordinates			Volume size (mm ³)	Peak T	Peak Z
		x	y	z			
CS+ versus CS-							
Caudate	Right	10	4	2	2432	6.89	5.13
	Left	-10	0	14	640	4.69	3.96
Brainstem		2	-36	-46	3200	5.77	4.59
Thalamus/PAG		-6	-20	-2	5824	5.35	4.36
Insula	Right	34	12	6	1600	4.61	3.9
	Left	-34	16	6	576	4.02	3.51
Precentral Gyrus	Right	42	-8	54	384	4.27	3.69
Precuneus	Right	10	-48	62	448	4.02	3.51
Parietal Lobe	Left	-30	-32	46	384	3.82	3.37
CS- versus CS+							
Subgenual ACC		-2	32	-6	384	4.11	3.58
Rostral ACC		-6	44	10	960	4.11	3.57