Article

Current Biology

Behavioral and Neural Mechanisms of Overgeneralization in Anxiety

Highlights

- In anxiety, learning modifies stimulus representation in primary cortex and amygdala
- Altered representation leads to later reduced discrimination even in a safe context
- These changes in perceptual thresholds contribute to overgeneralization in patients
- A wide affective network mediates both positive and negative stimuli during learning

Authors

Offir Laufer, David Israeli, Rony Paz

Correspondence

rony.paz@weizmann.ac.il

In Brief

Emotional learning in anxiety patients leads to less discrimination of the dangerous stimulus from other similar stimuli. Due to plasticity in the neural representation that persists well after the emotional context ceased, patients cannot distinguish new safe stimuli from the previously dangerous stimulus. Laufer et al. further identify the underlying brain regions.

Laufer et al., 2016, Current Biology *26*, 713–722
March 21, 2016
http://dx.doi.org/10.1016/j.cub.2016.01.023

CellPress

Current Biology
Article

Behavioral and Neural Mechanisms of Overgeneralization in Anxiety

Offir Laufer,[1] David Israeli,[2,3] and Rony Paz[1,3,*]
[1]Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel
[2]The Jerusalem Mental Health Center and the Hebrew University, Jerusalem 91120, Israel
[3]Co-senior author
*Correspondence: rony.paz@weizmann.ac.il
http://dx.doi.org/10.1016/j.cub.2016.01.023

SUMMARY

Overgeneralization of dangerous stimuli is a possible etiological account for anxiety disorders, yet the underlying behavioral and neural origins remain vague. Specifically, it is unclear whether this is a choice behavior in an unsafe environment (“better safe than sorry”) or also a fundamental change in how the stimulus is perceived. We show that anxiety patients have wider generalization for loss-conditioned tone when compared to controls and do so even in a safe context that requires a different behavioral policy. Moreover, patients overgeneralized for gain-conditioned tone as well. Imaging (fMRI) revealed that in anxiety only, activations during conditioning in the dACC and the putamen were correlated with later overgeneralization of loss and gain, respectively, whereas valence distinction in the amygdala and hippocampus during conditioning mediated the difference between loss and gain generalization. During generalization itself, neural discrimination based on multivoxel patterns in auditory cortex and amygdala revealed specific stimulus-related plasticity. Our results suggest that overgeneralization in anxiety has perceptual origins and involves affective modulation of stimulus representations in primary cortices and amygdala.

INTRODUCTION

Stimulus generalization is a fundamental learning mechanism by which an acquired response for a specific stimulus is transferred to other similar, but not identical, stimuli [1, 2]. When the stimulus predicts aversive outcomes, it makes sense to have a wider generalization and respond similarly to stimuli that are even less similar to the original conditioned one; this is because a miss (incorrectly identifying the dangerous stimulus as a safe one) is more costly than a false alarm (incorrectly identifying a safe stimulus as the conditioned one) [3–5]. Here, we call such response bias a “better safe than sorry” approach.

Recent studies have suggested that overgeneralization can underlie anxiety disorders [6–8]. In this account, following formation of association between a stimulus and an aversive outcome, stimuli that are only somewhat similar to the original stimulus would still elicit increased anxiety in affected individuals [8–13]. In an unsafe environment full of complex stimuli where the original stimulus is still dangerous, this is an extreme but rational strategy.

An additional possibility is that anxiety patients overgeneralize not only because of a choice bias [14] but also because they perceive the stimulus differently after it was associated with affective content [15, 16]. In this account, learning with reinforcers induces plasticity that changes early stimulus representation and results in less discrimination and therefore more generalization, even in a later safe context. Supporting evidence comes from findings that negative reinforcers can result in changes in stimulus perception [11, 17, 18]. Here, we hypothesized that such a basic mechanism can contribute to anxiety responses and that patients would have compromised perception that contributes to overgeneralization. We further hypothesized that this would be paralleled by changes in stimulus representations in sensory regions that are modulated by affective regions.

RESULTS

Human subjects underwent a conditioning session in which three well-separated pure tones were assigned as conditioned stimulus (CS) gain, CS loss, and CS neutral (300 Hz, 500 Hz, and 700 Hz, counterbalanced across subjects; 21 trials each, presented in pseudorandom order). The CS loss could be followed by monetary loss and the CS gain by monetary gain (Figure 1A; [17]; CS neutral was never followed by any outcome). Once conditioning was over, a generalization phase was initiated and announced (Figure 1A), in which a tone appeared in each trial and the subjects had to indicate whether this is one of the two reinforced tones they had heard in the previous conditioning phase or a different new tone (a loss, gain, or neither response). A correct classification resulted in monetary gain. There were 15 unique tones overall, the two original CSs and [±20%, ±10%, ±3%] around each CS (generalized stimulus, GS), and the CS neutral. Importantly, notice that in this generalization phase the context is safe (the CS loss no longer entails loss), and, moreover, the two original reinforced tones, CS loss and CS gain, require the same behavioral policy. The optimal behavior is therefore to have an as-narrow-as-possible generalization around both original tones, i.e., to identify all GSs as “neither” and the two CSs as either “loss” or “gain.”

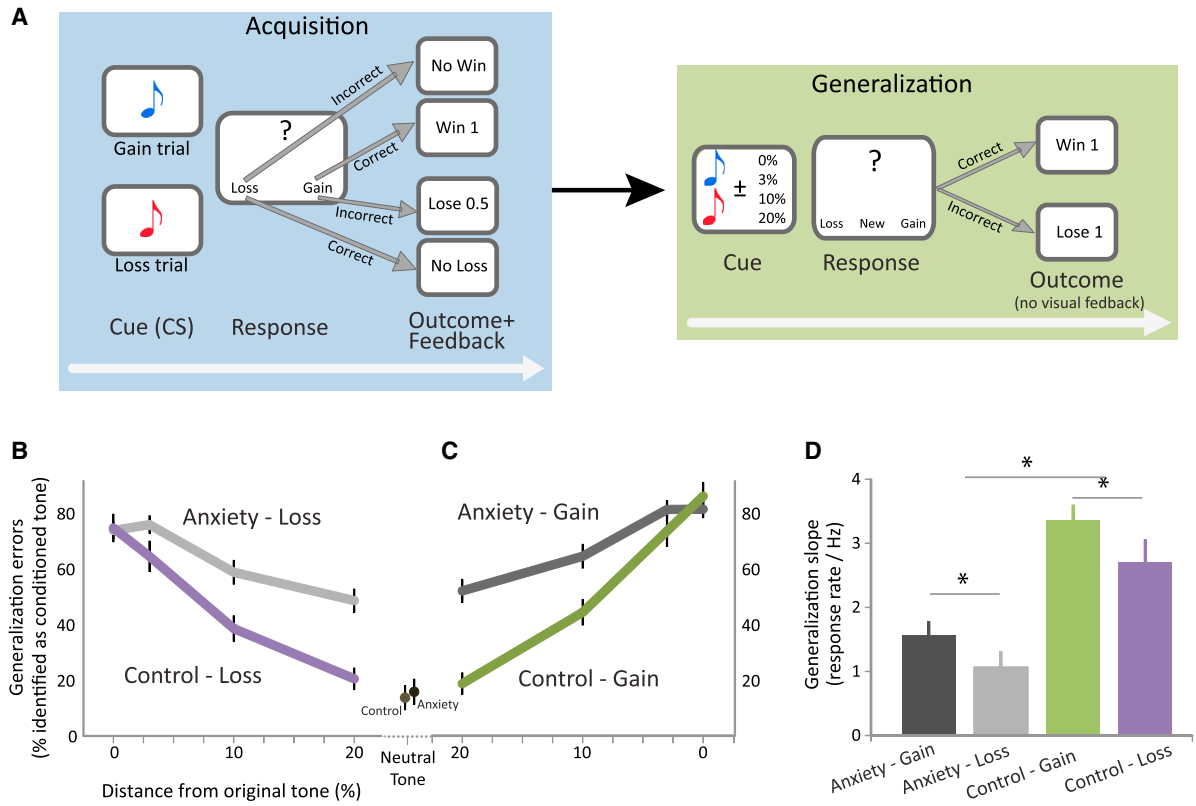

Figure 1. Generalization in Anxiety

(A) Conditioning phase required subjects to learn to associate one tone (CS gain) with one button to obtain monetary reward and another tone (CS loss) with the other button to avoid monetary loss. In the generalization phase that followed (the focus of this study), subjects heard different tones and had to choose whether it was one of the tones that was conditioned in the previous phase (independent if it was the gain- or loss-related tone) or another, new tone.

(B) Proportion of trials identified as the loss-conditioned tone as a function of distance (in % Hz) from the loss-conditioned tone.

(C) Same as (B) for the gain-conditioned tone (the graph is reflected only for presentational reasons, the x axis in B and C represents ±3%, ±10%, ±20%).

(D) Generalization quantified as the slope (averaged over subjects) reveals that there was wider generalization (smaller slopes) for loss-related tone in both patients and controls but that patients had even wider generalization (n = 28 GAD, 16 controls).

See also Figure S1.

We combine here the data acquired previously in healthy subjects [17] with new data acquired from patients with generalized anxiety disorder (GAD, n = 28). The goal is to compare and identify differences between the two populations in behavior and in brain activations.

Wider Generalization in Anxiety

We found that subjects with GAD (n = 28) have wider generalization than controls (n = 16) around the aversive CS loss (Figure 1B; $F = 41.34$, $p < 0.01$, group main effect, two-way ANOVA). Interestingly, patients showed wider generalization around gain-related tone as well (Figure 1C). Still, in both groups there was wider generalization for loss than for gain (Figure 1D; $F = 4.73$, $p < 0.05$, valence main effect, two-way ANOVA) [5, 11, 17]. Moreover, the width of generalization around the loss-related tone corresponded to Hamilton anxiety scores (Figures S1A–S1C; $F = 3.58$, $p < 0.05$, one-way ANOVA and $p < 0.05$, Pearson; but not around gain-related tone, $F = 1.33$, $p > 0.1$; or for the differences between loss and gain slopes, $F = 0.94$, $p > 0.1$).

The wider generalization in anxiety was not due to differences in baseline auditory performance or in the learning (Figures 2A–2C), was not observed around a neutral unconditioned tone (Figures 2D and 2E; Figure 1C), and did not differ across gender (two-way ANOVA, $F = 0.62$, $p = 0.43$ for gender effect; $F = 1.95$, $p = 0.16$ for valence effect; $F = 0.7$, $p = 0.4$ for interaction). We also found seemingly normal generalization curves in a control group of major depressive disorder patients (Figures S1D and S1E; $F = 2.18$, $p > 0.1$, group main effect; n = 11 with no GAD diagnosis but with similar medication profile).

These findings support the notion that anxiety patients overgeneralize due to altered perception of the stimulus and therefore do so even in a safe context when the original stimulus is no longer dangerous and even when the generalization harms them. We further hypothesized that there should be evidence for altered neural representations that are formed during conditioning.

Brain Activity that Differentiates Generalization in Anxiety

To identify neural activations, a subset of patients from the GAD group and all healthy controls underwent functional imaging during task performance (fMRI, n = 16 GAD and n = 16 controls), and

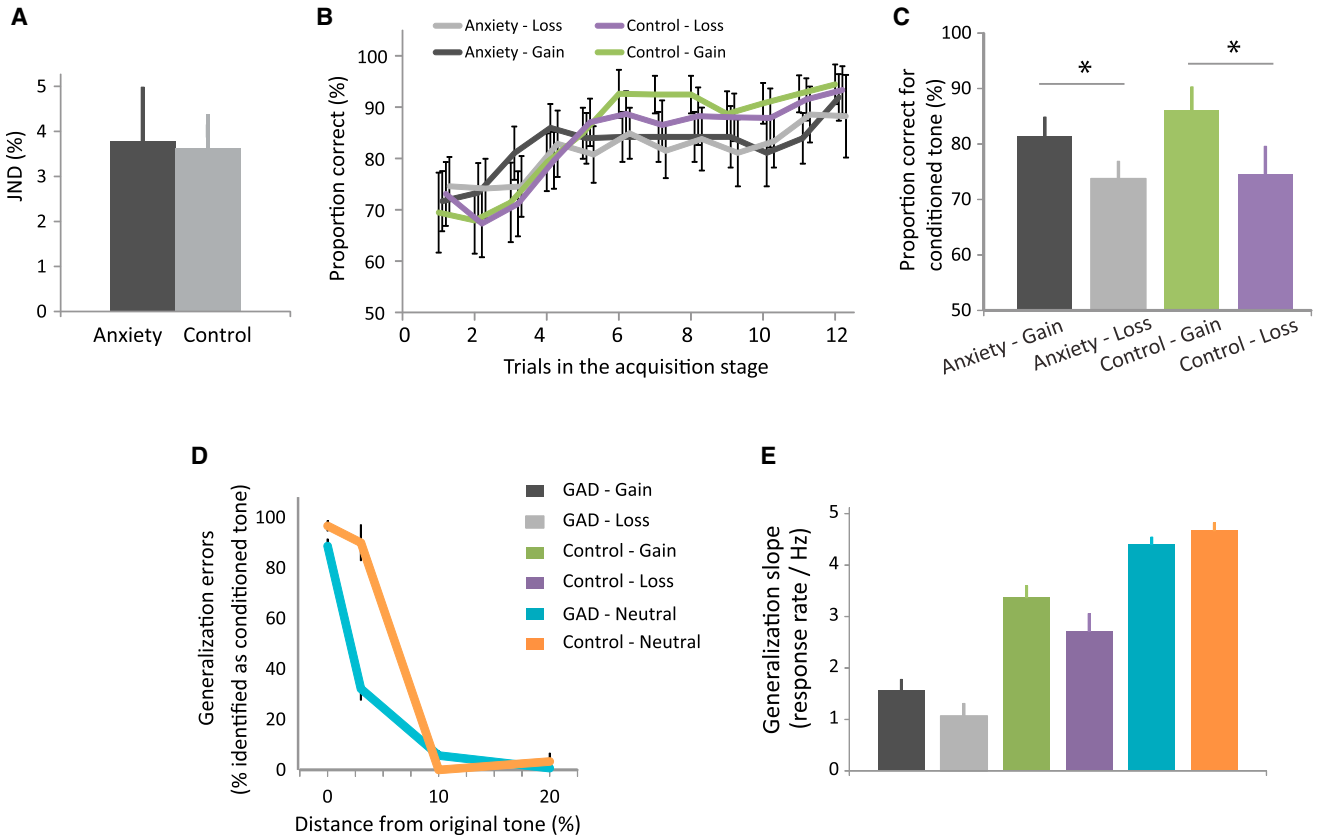

Figure 2. The Wider Stimulus Generalization in Anxiety Is Not Observed for Neutral Stimuli and Cannot Be Attributed to Differences in Learning or Basic Auditory Performance

(A) Baseline discrimination thresholds (just-noticeable difference [JND]) for GAD patients and control group ($p > 0.5$, t test). JND was tested with a standard 2AFC auditory task using adaptive staircase procedure (n = 14 GAD, 16 controls).

(B) Learning rates were similar across groups and valence ($p > 0.1$, group and valence effects and interactions, $p < 0.005$, acquisition stage main effect three-way ANOVA). Shown are correct response rates in the conditioning phase. All subjects quickly learned to discriminate between the 300 and 700 Hz pure tones that served as CSs (counterbalanced across subjects) and associate them correctly. Behavior reached plateau around the fifth trial and was not different across anxiety or control and gain or loss (n = 28 GAD, 16 controls).

(C) During generalization, there was no difference between anxiety patients and controls in correct responses to the conditioned tone (either gain or loss) ($p > 0.1$, group main effect, two-way ANOVA), indicating that CS loss and CS gain were learned to similar extents in patients and controls and that there were no between-group differences in attentional factors (n = 28 GAD, 16 controls). The fact that the CS loss was associated with fewer correct responses than the CS gain is a result of the wide generalization and less discrimination observed around it. In other words, if a subject cannot discriminate between two close tones (e.g., the CS loss and +3% from it), then fewer correct identifications would occur for both.

(D) Proportion of trials identified as the original tone but for a neutral-unpaired tone (in % Hz, same as Figures 1B and 1C), for anxiety patients (n = 10) and controls (n = 10). Notice that controls had similar narrow generalization (see E below), further suggesting that the wider generalization for loss and gain (Figures 1B and 1C) is due to affective conditioning.

(E) Generalization slopes of patients and controls for the gain, loss, and neutral tones (left four bars are similar to Figure 1D).

See also Figure S1.

we conducted two types of analyses. The first analysis was conducted on a wide set of a priori regions of interest (ROIs) selected due to their implication in emotional learning, emotional modulation, and sensory-auditory representation: all prefrontal cortices (PFCs), the striatum (caudate and putamen), auditory cortices (BA41, BA42, BA21, BA22), insula (BA13), hippocampus, amygdala, and the thalamus. The second analysis was conducted using a whole-brain approach (this analysis replicated the findings in the above-mentioned regions and further revealed additional frontal and midbrain activity detailed in the Supplemental Information).

To identify regions that play a differential role in generalization between anxiety and controls, we performed three kinds of analyses. The first examined differential activity during conditioning. The second examined differential activity during conditioning but that correlates with later individual width of generalization. The third examined activity during the generalization itself and uses neural discrimination measures to compare activity between the conditioned and its surrounding stimuli. The rationale behind this set of analyses follows the rationale of our task design: the affective manipulation occurs during the conditioning itself rather than during the generalization, and we therefore look for activity during conditioning that induces plasticity that is later reflected during generalization in the (in)ability to discriminate the stimuli. All analyses employed

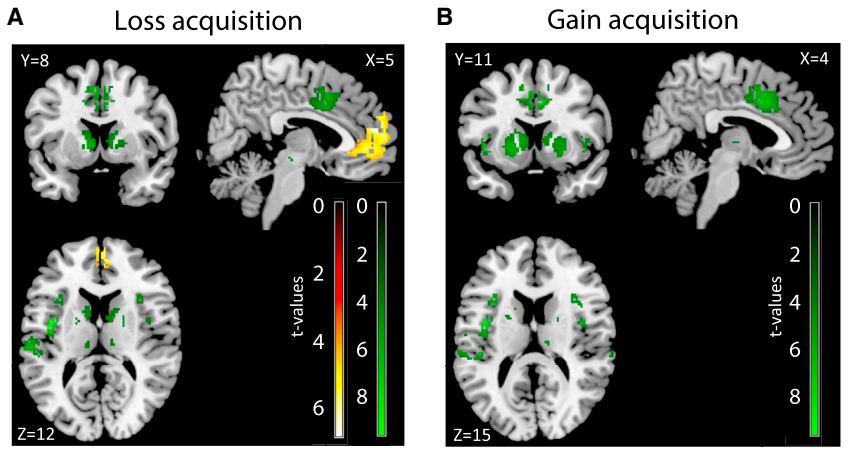

Figure 3. Differential Activations during Conditioning

(A) During loss-conditioning trials, activations were higher in controls in the ventro-medial-prefrontal cortex (vmPFC) (yellow) and BA10 and higher in anxiety patients in dACC, caudate, insula, and auditory cortices (green).

(B) During gain-conditioning trials, no regions showed increased activation in controls, but activations were higher in anxiety patients in dACC, caudate and putamen, insula, and auditory cortices (green). Activations are thresholded at $p < 0.05$, FWE SVC, $k > 10$ ($n = 16$ GAD, 16 controls).

the contrast between the anxiety group and the control group as the main objective of the current study.

Differential Activity during Conditioning between Anxiety and Controls

To determine group differences in activations during the conditioning phase, a two-sample t test of loss > baseline and gain > baseline contrasts was conducted within a general linear model framework that incorporated both groups (Figure 3; Table S1). Activations were higher in anxiety in dorsal anterior cingulate cortex (ACC) (BA24,32) and caudate and in the primary auditory cortex (BA41). In contrast, activations were higher in controls mainly in ventro-medial PFC and ventral ACC (BA10,24). These differential activations in PFC are compatible with the roles of the ACC in promoting anxiety and fear and the role of the ventro-medial cortices in fear modulation and safety [19–22].

Activity during Conditioning that Contributes to Wider Generalization

To identify the regions in which activity during conditioning contributed to subsequent wider generalization, we calculated individual slopes of generalization curves and used them as covariate with loss > baseline or gain > baseline contrasts, within the GAD versus controls model. We emphasize that neural activity is taken from the conditioning phase, but behavior is taken from the generalization phase that comes only later. This analysis allows us to examine the regions that separately contribute to either gain or loss generalization.

In the loss > baseline contrast, activations in the dorsal-ACC (BA24 and 32) were correlated with individual generalization slopes (Figure 4A; Table S2). There was a significant relationship, larger in GAD relative to controls, between activations in these regions and the extent of overgeneralization, suggesting that the higher the activity during the loss-conditioning trials, the more the individual subsequently generalized for the loss-conditioned tone (Figure 4A; Figure S2A; notice these correlation plots report the size of the effect and its directionality and do not constitute additional non-independent significance tests). In contrast, in the gain > baseline condition, the Putamen activity was correlated with later generalization of gain (Figure 4B), across both populations together, but again only for anxiety patients when tested individually (Figure S2B). These findings suggest that there might be separate contributions for affective modulation via the dorsal anterior cingulate cortex (dACC) and the putamen, the former contributes more during negative experiences and the latter more during positive ones; however, each contributes to wider generalization in anxiety patients.

Yet, anxiety patients, just like controls, were able to differentiate loss from gain and further exhibited an even wider generalization in the loss domain when compared to the gain (Figures 1 and 2). To examine this and to identify shared mechanisms that compare valence, we created a loss > gain contrast and used as covariate the difference in individual slopes of generalization (loss minus gain). This revealed that activity in the amygdala during conditioning was correlated with later individual differences in generalization between loss and gain (Figure 5), as well as in the posterior hippocampus (Figure S3).

If the amygdala is involved as a shared mechanism across valence for generalization, and the ACC and putamen contribute valence-specific information for generalization (at least in the current generalization task), then there might be matching functional link during learning. To test this, a functional connectivity analysis (psychophysiological interaction [PPI]) was performed using the amygdala as seed, revealing increased connectivity between the right amygdala and right dACC (BA24) during loss-conditioning trials only and between right amygdala and putamen during gain-conditioning trials only (Figure S4; Table S3). No significant functional connectivity was found when using the posterior hippocampus as seed.

Neural Discrimination during Generalization

All analyses until now examined activity during conditioning, either independently or with relation to later behavioral generalization. To identify regions that contributed directly during the generalization, we tested whether overall differences in generalization between the four experimental conditions (anxiety or control × loss or gain) would be reflected in neural discrimination measures between the CSs and their respective surrounding GS tones. For each subject, neural discrimination was quantified as the correlation between spatial patterns of activity [23] evoked during the generalization stage by the CS and the GSs. All

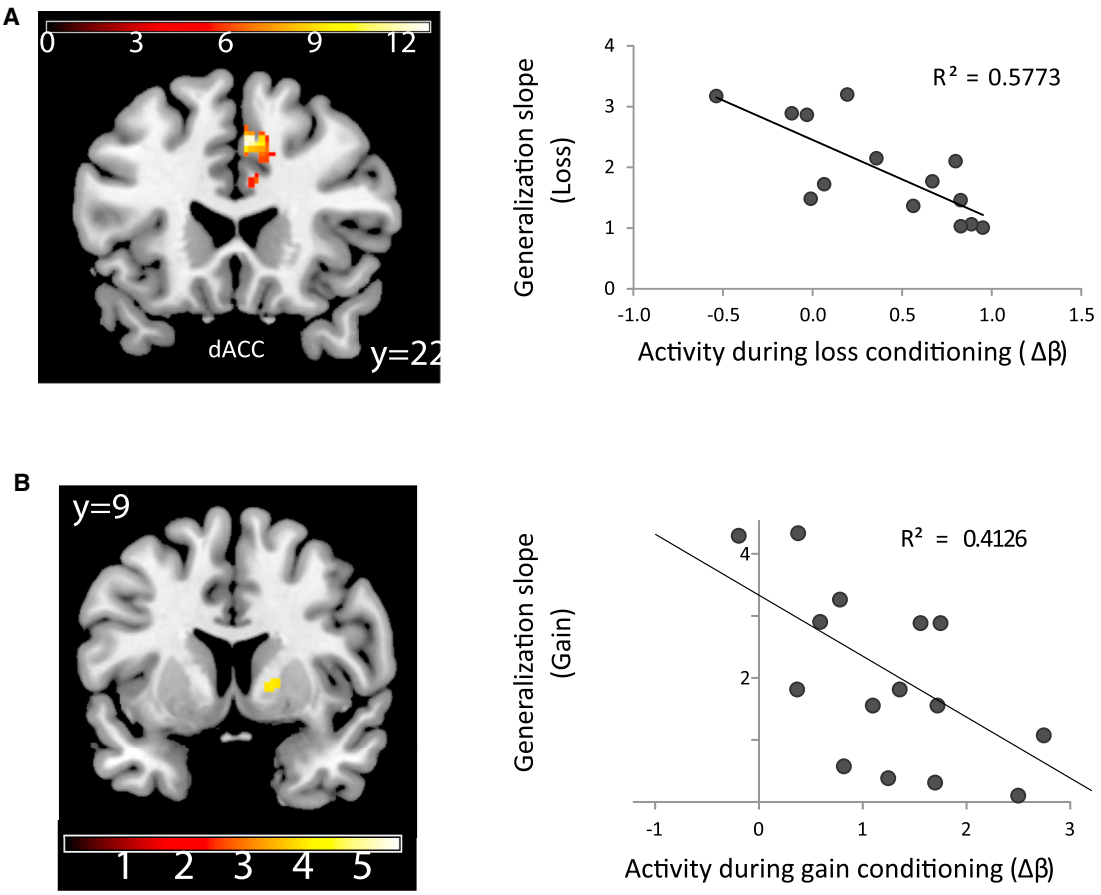

Figure 4. Activations in ACC and Putamen during Conditioning Are Correlated with Later Individual Generalization Extent of Loss and Gain, Respectively

(A) Activation map when using generalization slopes as covariate with a loss > baseline contrast in a GAD versus control model: the dorsal ACC activity during conditioning of loss was correlated with the later individual loss generalization. Activations are thresholded at $p < 0.05$, FWE SVC, $k > 10$.

(B) Same as (A) when using gain trials during conditioning and gain generalization: the putamen activity during conditioning was correlated with later generalization. The amygdala showed significant activation during conditioning that was significantly more correlated with the later individual generalization behavior of GAD patients relative to controls.

Correlations between individual activations of anxiety subjects and their behavioral slopes are shown at the right. Notice these correlation plots report the effect size and directionality and do not constitute additional non-independent tests.

See also Figure S2.

regions implicated in any of the previous analyses were examined. We found that in the primary auditory cortex (BA41), the correlation was significantly higher in anxiety than in controls for both gain and loss (Figure 6A; group main effect $F = 13.85$ $p < 0.001$, two-way ANOVA). However, only in the amygdala did neural discrimination match the behavior across all four conditions (Figure 6B, left; group main effect $F = 10.03$ $p < 0.01$, two-way ANOVA); namely, high discrimination (low correlation) for control gain that exhibited a narrow generalization; smaller for control loss, exhibiting wider generalization [11, 17]; even smaller for anxiety gain that exhibited even wider generalization; and lowest discrimination for anxiety loss that indeed exhibited the widest generalization. Moreover, the amygdala showed a significant correlation at an individual level between generalization slopes and neural discrimination (Figure 6B, right; $r = -0.45$, $p < 0.001$). These patterns are likely due to conditioning because comparing early to late trials showed that the largest reduction in neural discrimination occurred for CS loss in the amygdala (Figures S5D–S5F). In contrast, neural discrimination remained high for comparing CS loss or CS gain with a well-separated unpaired neutral CS (Figures S5A–S5C), reinforcing the idea that detection improves (Figure S1F) [24], even though discrimination deteriorates. Other regions that contributed during conditioning did not show such relationship between neural discrimination and behavior during generalization (Figure S6).

DISCUSSION

Our study shows that anxiety patients overgeneralize after conditioning with affective stimuli and provides evidence that they have altered perception of simple stimulus features. There are two important factors of the experimental design that support this finding: first, we used simple stimuli of basic features as pure frequency tones; second, and importantly, the

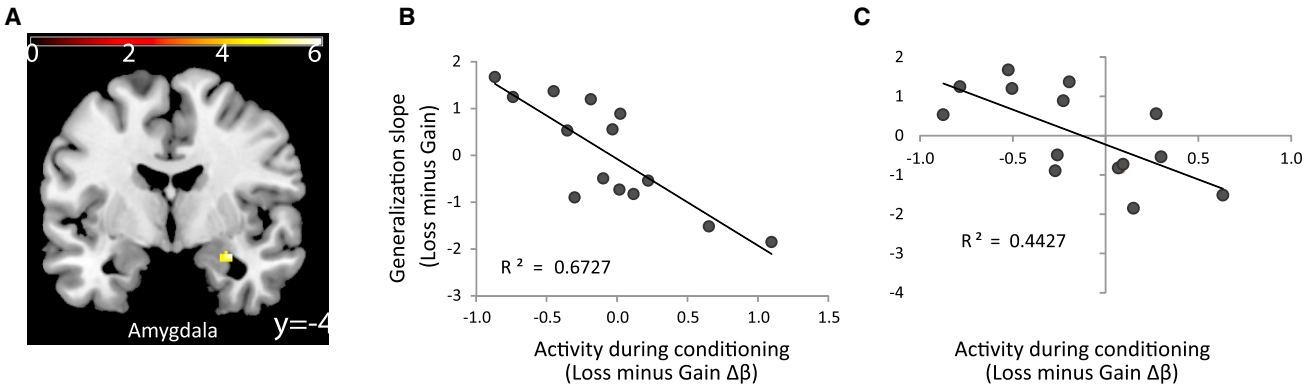

Figure 5. Activations in Amygdala during Conditioning Are Correlated with Later Individual Generalization Difference between Loss and Gain

(A) Activation map when using generalization slopes as covariate with a loss > gain contrast in a GAD versus control model: the amygdala showed significant activation during conditioning that was significantly more correlated with the later individual generalization behavior of GAD patients only. Activations are thresholded at $p < 0.05$, FWE SVC, $k > 10$.

(B) Correlations between individual activations of anxiety subjects and their behavioral slopes are shown. Notice this correlation plot reports the effect size and directionality and does not constitute additional non-independent tests.

(C) Same as (B) but using activations from anatomical ROIs of the amygdala.

See also Figures S3 and S4.

generalization occurred in a safe context when the stimuli no longer entail danger. We then identified neural pathways that participate in affective modulation during conditioning: the putamen for gain, the ACC for loss, and the amygdala for both. Finally, we show that amygdala activity during generalization itself correlates with extent of behavioral generalization across valence and anxiety, and auditory cortex activity differentiated between anxiety and controls (but not valence). The combined evidence suggests that in anxiety patients, affective stimuli shape early sensory representations during conditioning, and these altered representations later result in less discrimination between the conditioned stimulus and new safe stimuli.

In the current paradigm, the generalization phase takes place after conditioning already occurred and the stimulus no longer predicts a negative outcome. Yet, even in this safe context, patients still have wide generalization. This is an important addition to previous studies where the aversive conditioning occurs during the generalization as well [9, 13, 25–27], and wider generalization can therefore be attributed to a decision bias. Moreover, if losing is a source of anxiety (as during conditioning), then in the generalization phase they should try to minimize it by making fewer errors, hence resulting in narrower generalization. However, patients still overgeneralize. This implies that the generalization effects were due, at least in part, to fundamental changes in how patients perceived the stimulus and to what extent they could discriminate it from other stimuli [11, 17].

An additional finding was that patients overgeneralize to gain-conditioned stimulus as well. This could reflect the fact that patients constantly screen for ominous future scenarios, and a gain situation automatically activates the prospect of loss. Alternatively, it could stem from differential sensitivity to reward and altered processing in reward circuits [28–30], in line with our finding of putamen activity during conditioning that correlated with gain generalization. An additional likely interpretation, one that is not mutually exclusive and supported by the reward sensitivity as well, is that there is heightened sensitivity in patients to any affective stimulus and its impact on early sensory mechanisms. This option is in line with the hypothesis and findings of the current study, specifically with the occurrence of generalization in a later safe context, but also from the observed neural activations.

Indeed, we identified a network in which activity during conditioning correlates with the generalization extent in patients only. The ACC activity during loss-conditioning trials contributed to later less discrimination around the loss-conditioned tone, as well as correlated with amygdala activity. This is in line with the body of evidence showing that medial-prefrontal cortex (mPFC)-amygdala interactions are required for correct fear discrimination and safety learning [31–33]. On the other hand, putamen activity during gain-conditioning trials contributed to later less discrimination around the gain-conditioned tone, and it was also correlated with the amygdala. Interestingly, the amygdala and hippocampus activity during conditioning also correlated with later generalization but only when taking the difference between gain and loss. For the hippocampus, this is in accordance with its well-established role in many forms of generalization [25, 34–37] and with its role in trait anxiety [38].

These pathways likely contribute to modulate stimulus representations that later underlie generalization patterns, and we therefore used neural discrimination measures to unveil the regions in which activity directly matched behavioral discrimination of individual subjects. The auditory cortex matched differences between anxiety and controls, but not across valence, suggesting that in individuals with anxiety, there is a larger or faster tendency to modulate representations in auditory cortex. This is in line with evidence from animal models demonstrating that both aversive and rewarding information can induce plasticity [39–43], one that is retained even following extinction [44, 45]. A natural candidate to drive such change is the aforementioned correlation between amygdala activity during conditioning and later generalization across valence [46].

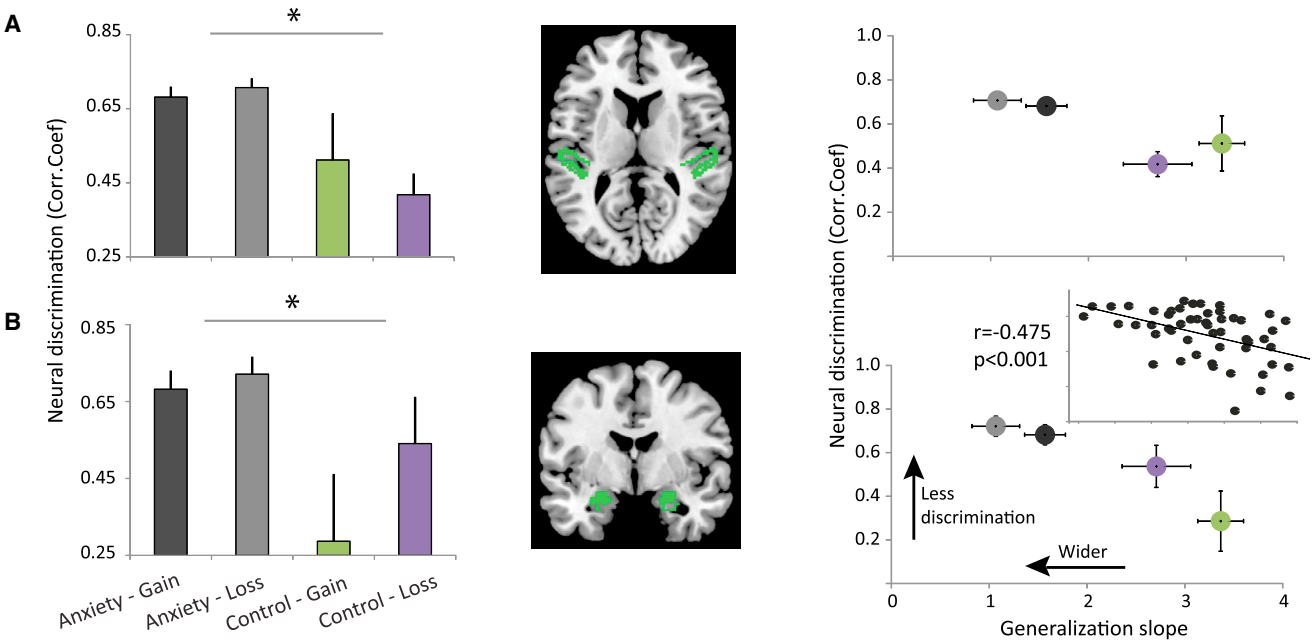

Figure 6. Neural Discriminability Corresponds to Generalization in the Amygdala

Neural discriminability between the CS and surrounding tones (GS), measured as the correlation between the CS- and GS-evoked multivoxel spatial activity patterns. Shown are the correlations for the four conditions (left) and the relationship between these correlations and generalization slopes (taken from Figure 1D). (A and B) In the auditory cortex (A), neural discriminability matched generalization in terms of differentiating anxiety patients from controls; in the amygdala (B), neural discriminability was correlated with generalization across all conditions and, importantly, at an individual level as well (inset, n = 16 GAD, 16 control). See also Figures S5 and S6.

In addition to its role during conditioning, the amygdala showed tight correlation between neural and behavioral discrimination across both anxiety and valence, providing evidence for its additional role in production of generalization due to plasticity that occurred in previous conditioning [17, 47–49]. This is much in line with its role in auditory pathways as well as in production of fear responses and with recent evidence for specific changes in auditory tuning curves within the amygdala [50]. Whereas other regions such as the insula were found to contribute to perceptual generalization in healthy subjects [17, 25], we did not find differences between anxiety and controls; yet, it is important to keep in mind that such null results can stem from sample size and other study limitations.

To conclude, overgeneralization following aversive experiences and/or in anxiety populations can result from both perceptual stimulus generalization [8, 15] and other cognitive sources (e.g., choice bias) [1, 7, 25], and the final generalization profile depends on several factors [16]. Here, we provide evidence that an affective stimulus induces changes in the core representation of the conditioned stimulus space and does so to a larger extent in some individuals via amygdala and primary sensory regions, therefore making these individuals more prone to exhibit anxiety symptoms.

EXPERIMENTAL PROCEDURES

Subjects

The patient group included GAD ambulatory patients (n = 28). All patients were diagnosed according to DSM-IV with the duration of illness ranging from 1 year to more than 15 years (age range, 19–45, mean age ± SEM = 27.96 ± 1.63; 14 females). All participants had normal hearing acuity (a subjective hearing threshold of 40 dB). The Hamilton Anxiety Rating Scale was used in order to estimate the anxiety symptoms prior to testing with our paradigm. The separate group of major depressive disorder (MDD) patients consisted of 11 patients under medication, and we used Hamilton Depression Rating Scale, 17 items for estimating depressive symptoms. All clinical diagnoses and testing were performed and followed by D.I. (MD, PhD), a certified psychiatrist and the head of the department of day hospitalization at the Jerusalem Mental Health Center.

Twenty-five patients were under SSRI or SNRI treatment, and five patients from the GAD group were without treatment prior to taking the test. All patients with MDD were medicated with either SSRI or SNRI. All the GAD patients that undertook the MRI were on SSRI or SNRI treatment. The SSRI treatment consisted of 19 patients on the following drugs: fluoxetine 20 mg (three patients), sertraline 50–100 mg (four patients), paroxetine 10–20 mg (six patients), escitalopram 10 mg (six patients). The SNRI treatment included venlafaxine in six patients at 150–225 mg. Patients did not use benzodiazepines more than 1 week prior to testing. No other drugs (i.e., mood stabilizers or antipsychotic drugs) were used. Patients in the GAD group showing comorbidity included two patients with panic disorder and one patient with obsessive compulsive disorder (OCD). There is history of major trauma in two patients without overt posttraumatic stress disorder (PTSD) symptoms. Patients in the GAD group could have symptoms of depression but did not reach Hamilton scores of minor MDD.

The control group consisted of 16 healthy subjects that were recruited through advertisement (age range, 23–32, mean age ± SEM = 25.3 ± 1.12; seven females). Most of the control group (n = 14) was tested ~4 years ago, and these data were already published [17]. However, the exact same procedures and equipment were used then and now. This is, therefore, not a replication of previous findings, but rather we use the control group as comparison to identify differences between them and the new cohort of patients.

Experiments were conducted under a Helsinki approval to R.P. and D.I. and with signed consents. Subjects were paid a flat amount for their participation.

Paradigm

The experiment consisted of two parts, an acquisition phase, designed to assign positive, neutral, or negative value to three different pure tones by learning, and a generalization phase—the target of this study—designed to test differences in the generalization around the positive and negative tones [5].

All participants underwent the behavioral version of the experiment, and 16 of 28 GAD patients (age range, 20–45, mean age ± SEM = 30.31 ± 2.00; eight females) and all 16 healthy subjects were also scanned (fMRI) during the experiment. For the acquisition phase, subjects were instructed that in each trial they would hear one of three tones (300, 500, or 700 Hz for 300 ms). One tone would gain them money, if they pressed one of two keys within 2.5 s after it (a "positive" tone); one tone would always lead to zero outcome (CS−); and one tone would lose them money, unless they pressed the other key after it (a "negative" tone). Tones and key assignment were counterbalanced across subjects. The experiment was preceded by visual examples of all possible scenarios and trial types.

Subjects learned by trial and error which tone was the positive and which was the negative and received visual feedback telling them whether they earned or lost in the trial. Learning in this paradigm is very fast, and subjects reach a plateau within few trials and maintain it for the rest of the experiment [5, 17]. In addition, there were Pavlovian trials (classical conditioning; informed by the appearance of the word "helpless" on the screen), in which the positive tone resulted in gain and the negative tone resulted in loss independent of the response. The randomly intermingled classical trials were implanted to ensure that the tones retain their valence, and the instrumental trials were implanted to ensure continuous active involvement and allow a measure of generalization afterward. Similar generalization behavior is observed with the use of classical conditioning only [17]. There was an equal amount of trials (21) per each tone, randomly interleaved, and hence there were 63 trials overall in the acquisition stage. The loss amount was 0.5 NIS and the gain was of 1 NIS to account for possible loss-aversion bias of a factor of 2 [5, 17, 51] (in other words, to equalize intensity of value); preliminary testing showed that an equal amount for gain and loss does not change the main results.

Subjects were informed when the acquisition stage was over and when a new different stage began. In this generalization stage, subjects were instructed that many different tones would be presented, one in each trial. In every trial, if they hear the positive or negative tone from the previous phase, they should press the same key that was previously associated with it; however, if they hear a tone that is not the exact same tone as either the positive or the negative tone, they should press a third middle key. They were informed that they would gain 1 NIS for a correct press (either pressing the appropriate key for a negative or positive tone, or the middle key for a different tone) or lose this amount for any mistake (pressing the positive or negative keys for a different tone, or the middle key for the positive or negative tone). There was no feedback reported to the subjects in this generalization stage, to avoid any changes in value of the tones that was acquired during the acquisition stage. Notice that the original tones are still referred to as positive or negative tones in the generalization phase, yet they have completely identical requirements and outcomes in this stage and therefore should entail identical decision-making policies. The generalization stage comprised 15 trials of the positive tone, 15 trials of the negative tone, and 96 of other tones (i.e., that required a press on the middle key). These 96 trials consisted of 12 CS− trials and 84 trials of tones that were ±3%, ±10%, and ±20% of the original (positive and negative) tones. There were 126 trials overall in the generalization stage. The acquisition and generalization stages appeared four times consecutively in a loop, to keep the original tones reinforced and avoid extinction effects.

Data Analysis

Behavioral

We quantified the proportion of correct and incorrect responses in the generalization stage, and individual generalization slopes were measured as the drop in response rate per 1 Hz of tone (the overall slope of the generalization curve in the range tested in our paradigm). The slopes were then included in either one-way ANOVA tests or two-way ANOVAs when both valence (gain and loss) and group (GAD and controls) factors were analyzed. A two-way ANOVA was also used to look for differences in learning rates during acquisition.

Baseline auditory discrimination thresholds (just-noticeable difference [JND]) were collected from GAD patients (n = 14) and healthy subjects (n = 16) using a two-alternative forced-choice (2AFC) paradigm and an adaptive "two-down, one-up" staircase converging procedure.

To examine the effect of conditioning on detection, we measured thresholds using a 2AFC, transformed up-down staircase implemented in the PSYCHOACOUSTICS toolbox. Each trial included two intervals: a 20 ms sine tone (the signal) presented within a band of band-pass noise of 300 ms (400–1,600 Hz) and the band-pass noise only. A different group of healthy subjects (n = 18) had to identify which interval included the tone, which varied in intensity adaptively, before and after conditioning with monetary outcomes (gain, loss, neutral). This experimental paradigm was run only on healthy subjects.

Neuroimaging

Functional images were acquired by T2*-weighted gradient-echo echo-planar imaging (35 slices, flip angle = 75°, time repetition [TR] = 2,000 ms, time echo [TE] = 30 ms, voxel size = 3 × 3 × 4 mm) on a 3 T MRI scanner (Tim Trio, Siemens). Anatomical images were acquired using a T1-weighted MPRAGE sequence (176 slices, flip angle = 9°, TR = 2,300 ms, TE = 2.98 ms, FOV 256, voxel size 1.0 × 1.0 × 1.1 mm).

We used statistical parametric mapping (SPM8; Wellcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm) for image preprocessing. Images were realigned to the first volume, unwarped, normalized to a standard EPI template based on the MNI reference brain, resampled to 2 × 2 × 2 mm voxels, and spatially smoothed with an isotropic 8 mm FWHM Gaussian kernel, and for estimation of statistical maps using a general linear model approach with six rigid-body realignment parameters as nuisance covariates.

All analyses were first conducted using a wide set of a priori regions implicated in emotional learning, emotional modulation, and sensory-auditory representation: all PFCs, the striatum (caudate and putamen), auditory cortices (BA41, BA42, BA21, BA22), insula (BA13), hippocampus, amygdala, and the thalamus, thresholded at $p < 0.05$, FWE small volume correction, $k > 10$. Analyses were repeated using a whole-brain approach, thresholded at $p < 0.05$, FWE corrected, $k > 10$. Anatomical ROIs were defined based on known anatomical landmarks according to the Talairach Daemon Atlas using the SPM WFU PickAtlas tool.

To trace valence-related brain regions, regressors were constructed for conditioning trial types. Regressors were modeled as box-car functions, from trial onset until the participant's response, and convolved with the hemodynamic response function. Feedback screens were modeled separately as 1.5 s epochs. Second-level random effects analyses were performed in order to identify differences in activations between anxiety and healthy control subjects (two-sample t test). To identify brain regions mediating the observed effect of monetary outcome on the extent of generalization, we added individual generalization slopes as covariates to the second-level random-effects analyses of the corresponding contrasts.

A psychophysiological interaction (PPI) analysis was conducted to identify regions that show a significant condition-specific difference in functional connectivity with the amygdala. The conditions of interest were gain and loss acquisition trials. The regressors in the PPI analysis included (1) the activation time course of the volume of interest (i.e., physiological variable; the blood-oxygen-level-dependent [BOLD] signal); (2) a regressor representing the psychological variable of interest (i.e., the different experimental conditions); and (3) a regressor representing the cross product of the previous two (the PPI). The first two regressors were added as covariates to the model, while the last regressor was the regressor of interest.

The neural discrimination measurement was based on multivoxel pattern analysis (in accordance with [23, 52]) and included three steps: (1) individual, tone-responsive primary auditory cortex and amygdala voxels were identified for each subject by contrasting tone > baseline during the acquisition phases, thresholded at $p < 0.1$ uncorrected and constrained with the respective anatomical ROI. (2) fMRI data from the generalization phases (independent of the acquisition phases used for ROI localization) were passed through the full pre-processing procedure except for normalization and smoothing, to maintain individual voxel information. (3) fMRI time series data were extracted for each generalization trial type (3 CSs and 12 other tones), for each voxel within the individual functional ROIs. Trial-type-specific vectors were constructed, consisting of peak BOLD signal intensities of all respective trials

within a 5-TR window (0–10 s after tone onset), and normalized to baseline (by subtracting the minimal fMRI activity in the interval between two TRs before tone onset and one TR before the peak signal). This produced 15 pattern vectors for each CS+ (gain and loss), 12 pattern vectors for CS−, and 84 vectors for all surrounding tones, the length of which corresponded to the number of individual functional voxels. Trial-specific vectors were then averaged across all trials to form mean pattern vectors, and these were used to calculate pairwise spatial correlation coefficients between the CS and other tones.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2016.01.023.

AUTHOR CONTRIBUTIONS

R.P., O.L., and D.I. designed the study. O.L. and D.I. conducted the experiments. O.L. analyzed the data. R.P. and O.L. wrote the paper.

ACKNOWLEDGMENTS

We thank Dr. Edna Furman-Haran and Nachum Stern for MRI procedures. This work was supported by I-CORE #51/11, ISF #26613, EP7 human brain project, and Minerva-Foundation grants.

Received: December 23, 2015
Revised: January 7, 2016
Accepted: January 11, 2016
Published: March 3, 2016

REFERENCES

1. Guttman, N., and Kalish, H.I. (1956). Discriminability and stimulus generalization. J. Exp. Psychol. *51*, 79–88.
2. Bouton, M.E. (2006). Learning and Behavior: A Contemporary Synthesis (Sinauer Associates Inc).
3. Watson, J.B., and Rayner, R. (1920). Conditioned emotional reactions. J. Exp. Psychol. *3*, 1–14.
4. Green, D., and Swets, J. (1989). Signal Detection Theory and Psychophysics (Peninsula Pub).
5. Schechtman, E., Laufer, O., and Paz, R. (2010). Negative valence widens generalization of learning. J. Neurosci. *30*, 10460–10464.
6. Jovanovic, T., and Ressler, K.J. (2010). How the neurocircuitry and genetics of fear inhibition may inform our understanding of PTSD. Am. J. Psychiatry *167*, 648–662.
7. Pitman, R.K., Rasmusson, A.M., Koenen, K.C., Shin, L.M., Orr, S.P., Gilbertson, M.W., Milad, M.R., and Liberzon, I. (2012). Biological studies of post-traumatic stress disorder. Nat. Rev. Neurosci. *13*, 769–787.
8. Lissek, S. (2012). Toward an account of clinical anxiety predicated on basic, neurally mapped mechanisms of Pavlovian fear-learning: the case for conditioned overgeneralization. Depress. Anxiety *29*, 257–263.
9. Lissek, S., Rabin, S., Heller, R.E., Lukenbaugh, D., Geraci, M., Pine, D.S., and Grillon, C. (2010). Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. Am. J. Psychiatry *167*, 47–55.
10. Dunsmoor, J.E., Prince, S.E., Murty, V.P., Kragel, P.A., and LaBar, K.S. (2011). Neurobehavioral mechanisms of human fear generalization. Neuroimage *55*, 1878–1888.
11. Resnik, J., Sobel, N., and Paz, R. (2011). Auditory aversive learning increases discrimination thresholds. Nat. Neurosci. *14*, 791–796.
12. Levy-Gigi, E., Kéri, S., Myers, C.E., Lencovsky, Z., Sharvit-Benbaji, H., Orr, S.P., Gilbertson, M.W., Servatius, R.J., Tsao, J.W., and Gluck, M.A. (2012). Individuals with posttraumatic stress disorder show a selective deficit in generalization of associative learning. Neuropsychology *26*, 758–767.
13. Morey, R.A., Dunsmoor, J.E., Haswell, C.C., Brown, V.M., Vora, A., Weiner, J., Stjepanovic, D., Wagner, H.R., 3rd, LaBar, K.S., and LaBar, K.S.; VA Mid-Atlantic MIRECC Workgroup (2015). Fear learning circuitry is biased toward generalization of fear associations in posttraumatic stress disorder. Transl. Psychiatry *5*, e700.
14. Hartley, C.A., and Phelps, E.A. (2012). Anxiety and decision-making. Biol. Psychiatry *72*, 113–118.
15. Struyf, D., Zaman, J., Vervliet, B., and Van Diest, I. (2015). Perceptual discrimination in fear generalization: Mechanistic and clinical implications. Neurosci. Biobehav. Rev. *59*, 201–207.
16. Dunsmoor, J.E., and Paz, R. (2015). Fear generalization and anxiety: behavioral and neural mechanisms. Biol. Psychiatry *78*, 336–343.
17. Laufer, O., and Paz, R. (2012). Monetary loss alters perceptual thresholds and compromises future decisions via amygdala and prefrontal networks. J. Neurosci. *32*, 6304–6311.
18. Dunsmoor, J.E., and LaBar, K.S. (2013). Effects of discrimination training on fear generalization gradients and perceptual classification in humans. Behav. Neurosci. *127*, 350–356.
19. Milad, M.R., and Quirk, G.J. (2012). Fear extinction as a model for translational neuroscience: ten years of progress. Annu. Rev. Psychol. *63*, 129–151.
20. Etkin, A., Egner, T., and Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cogn. Sci. *15*, 85–93.
21. Rauch, S.L., Shin, L.M., and Phelps, E.A. (2006). Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research–past, present, and future. Biol. Psychiatry *60*, 376–382.
22. Greenberg, T., Carlson, J.M., Cha, J., Hajcak, G., and Mujica-Parodi, L.R. (2013). Ventromedial prefrontal cortex reactivity is altered in generalized anxiety disorder during fear generalization. Depress. Anxiety *30*, 242–250.
23. Li, W., Howard, J.D., Parrish, T.B., and Gottfried, J.A. (2008). Aversive learning enhances perceptual and cortical discrimination of indiscriminable odor cues. Science *319*, 1842–1845.
24. Padmala, S., and Pessoa, L. (2008). Affective learning enhances visual detection and responses in primary visual cortex. J. Neurosci. *28*, 6202–6210.
25. Onat, S., and Büchel, C. (2015). The neuronal basis of fear generalization in humans. Nat. Neurosci. *18*, 1811–1818.
26. Dunsmoor, J.E., Mitroff, S.R., and LaBar, K.S. (2009). Generalization of conditioned fear along a dimension of increasing fear intensity. Learn. Mem. *16*, 460–469.
27. Lissek, S., Kaczkurkin, A.N., Rabin, S., Geraci, M., Pine, D.S., and Grillon, C. (2014). Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. Biol. Psychiatry *75*, 909–915.
28. Cha, J., Carlson, J.M., Dedora, D.J., Greenberg, T., Proudfit, G.H., and Mujica-Parodi, L.R. (2014). Hyper-reactive human ventral tegmental area and aberrant mesocorticolimbic connectivity in overgeneralization of fear in generalized anxiety disorder. J. Neurosci. *34*, 5855–5860.
29. Guyer, A.E., Choate, V.R., Detloff, A., Benson, B., Nelson, E.E., Perez-Edgar, K., Fox, N.A., Pine, D.S., and Ernst, M. (2012). Striatal functional alteration during incentive anticipation in pediatric anxiety disorders. Am. J. Psychiatry *169*, 205–212.
30. Kessel, E.M., Kujawa, A., Hajcak Proudfit, G., and Klein, D.N. (2015). Neural reactivity to monetary rewards and losses differentiates social from generalized anxiety in children. J. Child Psychol. Psychiatry *56*, 792–800.
31. Klavir, O., Genud-Gabai, R., and Paz, R. (2013). Functional connectivity between amygdala and cingulate cortex for adaptive aversive learning. Neuron *80*, 1290–1300.
32. Likhtik, E., Stujenske, J.M., Topiwala, M.A., Harris, A.Z., and Gordon, J.A. (2014). Prefrontal entrainment of amygdala activity signals safety in learned fear and innate anxiety. Nat. Neurosci. *17*, 106–113.
33. Livneh, U., and Paz, R. (2012). Amygdala-prefrontal synchronization underlies resistance to extinction of aversive memories. Neuron *75*, 133–142.
34. Sangha, S., Narayanan, R.T., Bergado-Acosta, J.R., Stork, O., Seidenbecher, T., and Pape, H.C. (2009). Deficiency of the 65 kDa isoform

of glutamic acid decarboxylase impairs extinction of cued but not contextual fear memory. J. Neurosci. *29*, 15713–15720.

35. Kheirbek, M.A., Klemenhagen, K.C., Sahay, A., and Hen, R. (2012). Neurogenesis and generalization: a new approach to stratify and treat anxiety disorders. Nat. Neurosci. *15*, 1613–1620.
36. Kaouane, N., Porte, Y., Vallée, M., Brayda-Bruno, L., Mons, N., Calandreau, L., Marighetto, A., Piazza, P.V., and Desmedt, A. (2012). Glucocorticoids can induce PTSD-like memory impairments in mice. Science *335*, 1510–1513.
37. Wimmer, G.E., and Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. Science *338*, 270–273.
38. Satpute, A.B., Mumford, J.A., Naliboff, B.D., and Poldrack, R.A. (2012). Human anterior and posterior hippocampus respond distinctly to state and trait anxiety. Emotion *12*, 58–68.
39. Edeline, J.M., and Weinberger, N.M. (1993). Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. Behav. Neurosci. *107*, 82–103.
40. Aizenberg, M., and Geffen, M.N. (2013). Bidirectional effects of aversive learning on perceptual acuity are mediated by the sensory cortex. Nat. Neurosci. *16*, 994–996.
41. Armony, J.L., Servan-Schreiber, D., Romanski, L.M., Cohen, J.D., and LeDoux, J.E. (1997). Stimulus generalization of fear responses: effects of auditory cortex lesions in a computational model and in rats. Cereb. Cortex *7*, 157–165.
42. Bieszczad, K.M., and Weinberger, N.M. (2010). Representational gain in cortical area underlies increase of memory strength. Proc. Natl. Acad. Sci. USA *107*, 3793–3798.
43. David, S.V., Fritz, J.B., and Shamma, S.A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. Proc. Natl. Acad. Sci. USA *109*, 2144–2149.
44. Apergis-Schoute, A.M., Schiller, D., LeDoux, J.E., and Phelps, E.A. (2014). Extinction resistant changes in the human auditory association cortex following threat learning. Neurobiol. Learn. Mem. *113*, 109–114.
45. Armony, J.L., Quirk, G.J., and LeDoux, J.E. (1998). Differential effects of amygdala lesions on early and late plastic components of auditory cortex spike trains during fear conditioning. J. Neurosci. *18*, 2592–2601.
46. Chavez, C.M., McGaugh, J.L., and Weinberger, N.M. (2009). The basolateral amygdala modulates specific sensory memory representations in the cerebral cortex. Neurobiol. Learn. Mem. *91*, 382–392.
47. Shaban, H., Humeau, Y., Herry, C., Cassasus, G., Shigemoto, R., Ciocchi, S., Barbieri, S., van der Putten, H., Kaupmann, K., Bettler, B., and Lüthi, A. (2006). Generalization of amygdala LTP and conditioned fear in the absence of presynaptic inhibition. Nat. Neurosci. *9*, 1028–1035.
48. Ghosh, S., and Chattarji, S. (2014). Neuronal encoding of the switch from specific to generalized fear. Nat. Neurosci. *18*, 112–120.
49. Ciocchi, S., Herry, C., Grenier, F., Wolff, S.B., Letzkus, J.J., Vlachos, I., Ehrlich, I., Sprengel, R., Deisseroth, K., Stadler, M.B., et al. (2010). Encoding of conditioned fear in central amygdala inhibitory circuits. Nature *468*, 277–282.
50. Resnik, J., and Paz, R. (2015). Fear generalization in the primate amygdala. Nat. Neurosci. *18*, 188–190.
51. Sokol-Hessner, P., Hsu, M., Curley, N.G., Delgado, M.R., Camerer, C.F., and Phelps, E.A. (2009). Thinking like a trader selectively reduces individuals' loss aversion. Proc. Natl. Acad. Sci. USA *106*, 5035–5040.
52. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.

