NeuroImage 188 (2019) 445–455

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

The relationship of perceptual discrimination to neural mechanisms of fear generalization

Lauri Tuominen [a,b,*,1], Emily Boeke [a,2], Stephanie DeCross [a,3], Rick PF. Wolthusen [a,c,d], Shahin Nasr [d,e], Mohammed Milad [a,b,4], Mark Vangel [b,d,e], Roger Tootell [b,d,e], Daphne Holt [a,b,d]

[a] Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[b] Harvard Medical School, Boston, MA, USA
[c] Division of Psychological & Social Medicine and Developmental Neurosciences, Faculty of Medicine Carl Gustav Carus of the Technische Universität Dresden, Germany
[d] Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA
[e] Department of Radiology, Massachusetts General Hospital, Boston, MA, USA

ARTICLE INFO

Keywords:
Fear conditioning
Fear generalization
fMRI
Galvanic skin conductance
Insula
Visual perception
Face discrimination

ABSTRACT

The generalization of conditioned fear responses has been shown to decrease as a function of perceptual similarity. However, generalization may also extend beyond the perceptual discrimination threshold, ostensibly due to contributions from processes other than perception. Currently the neural mechanisms that mediate perceptual and non-perceptual aspects of fear generalization are unclear.

To investigate this question, we conducted a Pavlovian fear conditioning and generalization experiment, collecting functional magnetic resonance imaging (fMRI), skin conductance and explicit shock likelihood ratings, in 37 healthy subjects. Face stimuli were initially paired (CS+) or not paired (CS—) with an electrical shock. During the generalization phase, responses were measured to the CS+, CS— and a range of CS + -to—CS— morphs (generalization stimuli), selected for each participant based on that participant's discrimination ability.

Across multiple measurements, we found that fear generalization responses were limited to stimuli that could not be distinguished from the CS + stimulus, thus following a gradient closely linked to perceptual discriminability. These measurements, which were correlated with one another, included skin conductance responses, behavioral ratings, and fMRI responses of anterior insula and superior frontal gyrus. In contrast, responses in areas of the default network, including the posterior cingulate gyrus, angular gyrus and hippocampus, showed a negative generalization function extending to stimuli that were more likely to be distinguished from the CS+. In addition, the generalization gradients of the anterior insula and the behavioral ratings showed some evidence for extension beyond perceptual limits. Taken together, these results suggest that distinct brain areas are involved in perceptual and non-perceptual components of fear generalization.

1. Introduction

Fear generalization is a fear response to a stimulus that is similar – but not identical – to a previously encountered fear-inducing stimulus. Generalization responses are known to decrease in magnitude as perceptual similarity decreases between the stimuli and the original fear-inducing stimulus (Ghirlanda and Enquist, 2003; Guttman and Kalish, 1956). In prior work, fear generalization has often been implicitly assumed to reflect an affective process, in which fear responses extend to stimuli that are perceptually similar to (but distinguishable from) the original, aversive one, in a manner that is independent of perceptual discrimination. However, the role of perceptual discriminability in fear generalization in humans has not been systematically controlled or examined in prior studies (Struyf et al., 2015).

Fear generalization has been typically measured using a Pavlovian fear conditioning paradigm, beginning with a conditioning phase in

* Corresponding author. The Royal's Institute for Mental Health Research, Suite 5415, University of Ottawa, 1145 Carling Ave, Ottawa, Ontario, K1Z 7K4, Canada.
E-mail address: lauri.tuominen@theroyal.ca (L. Tuominen).
[1] Permanent address: Institute for Mental Health Research, University of Ottawa, Ottawa, Ontario, Canada.
[2] Permanent address: Department of Psychology, New York University, New York, New York, USA.
[3] Permanent address: Department of Psychology, University of Washington, Seattle, Washington, USA.
[4] Permanent address: Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois, USA.

https://doi.org/10.1016/j.neuroimage.2018.12.034
Received 12 April 2018; Received in revised form 21 November 2018; Accepted 16 December 2018
Available online 18 December 2018

which one experimental stimulus (the CS+) is paired with an aversive stimulus (e.g. electrical shock), compared to a control stimulus (the CS—) that is not paired with the aversive stimulus. This conditioning phase is followed by a generalization phase, during which stimuli of varying similarity to the CS+ (generalization stimuli) are presented, while fear responses to these stimuli are measured.

Within this framework, generalization of fear responses extending beyond the perceptual threshold (to stimuli that are easily distinguishable from the CS+) is thought to reflect a non-perceptual process (e.g., representing contributions of fear learning and memory mechanisms) rather than purely perception. Accordingly, the over-generalization of fear that has been observed in patients with anxiety disorders has been interpreted as a pathogenic mechanism for these disorders related to overactive fear-producing processes (Greenberg et al., 2013a; Lissek et al., 2014a). However, the mechanisms that underlie the wider generalization gradients found in anxiety disorders remain unknown. In fact, it remains unclear whether perceptual processes or other brain functions are responsible for the responses observed in distinct brain networks during fear generalization (Dunsmoor and Paz, 2015; Greenberg et al., 2013a; Lissek et al., 2014a).

Prior neuroimaging studies of fear generalization have tested the effect of fearful faces (Dunsmoor et al., 2011), circles (Lissek et al., 2014a), rectangles (Greenberg et al., 2013a), or neutral faces (Onat and Büchel, 2015) as stimuli, with evenly spaced gradations in *physical* similarity between the CS + stimulus and generalization stimuli. In these studies, at least two distinct patterns of responses have been identified: 1) responses to the CS+ and generalization stimuli that are *greater* than responses to the CS— in the anterior insula and connected areas of prefrontal cortex and 2) responses to the CS+ and generalization stimuli that are *lower* than responses to the CS— in regions of the default network, including the posterior cingulate cortex and hippocampus.

In these studies, conditioned fear responses generalized to the first or second generalization stimuli (i.e., the stimulus that was closest and second-closest in similarity to the CS+) (Dunsmoor et al., 2011; Greenberg et al., 2013a; Lissek et al., 2014a). However, because the perceptual discriminability of the generalization stimuli used in these studies was not measured, the fear generalization gradients in those studies cannot be directly compared to perceptual limits. Thus it remains unclear whether fear generalization (as measured by brain, autonomic or behavioral responses) is limited by perceptual boundaries, or extends beyond them (Struyf et al., 2015). Quantifying fear generalization in terms of perceptual properties, i.e., in units that are directly comparable across different subjects and populations, should permit better isolation of fear-memory related and other processes that could be affected in disease states.

Here, we measured the extent to which brain areas with conditioned fear responses show generalization of those responses at a perceptually-limited level, by controlling the discriminability of the stimuli for each subject. The discrimination threshold, based on the Just Noticeable Difference along the CS+ – CS— continuum, was measured and used to create individualized generalization stimuli that could or could not be distinguished from the CS + for each subject (i.e., above and below the discrimination threshold). These stimuli were presented in a fear conditioning and generalization experimental paradigm (Holt et al., 2014) during fMRI acquisition. In addition, we determined whether the generalization gradients observed in brain responses were linked to the autonomic and behavioral manifestations of fear generalization.

2. Material and methods

2.1. Participants

Thirty-eight subjects (17 female; mean (±SD) age: 28.51 ± 5.81) participated in this study; none had a history of psychiatric or neurological illness. Written informed consent was obtained from all subjects prior to experimental procedures, which had been approved by the Partners Health Care Institutional Review Board. All subjects had normal or corrected-to-normal vision, as assessed using the standard Snellen chart (Snellen, 1862). All data from one subject were excluded from analysis due to poor data quality resulting from technical difficulties during scanning. An additional eight subjects were excluded from the skin conductance analyses based on *a priori* quality control criteria (see below).

2.2. Stimuli

Four distinct images of human faces, and faces created by morphing between them (in 100 steps), were generated using FaceGen 3.4 (Singular Inversions, Canada), as described previously (Holt et al., 2014). All four faces (A, B, C and D) were male, Caucasian, had neutral facial expressions, and were easily differentiable from each other (Fig. 1A).

For each subject, a pair of face images (either A-B or C-D) was selected. One face was assigned to be later associated with a mild electrical shock delivered to the second and third finger of the right hand (the CS+); the other face was used as the stimulus that was unassociated with the shock (the CS—). Subjects were not explicitly informed of these assignments, only that they should "try to pay attention to any patterns they notice" in the stimulus pairings. The assignment of face pair (and the CS+ and CS— assignment within the pair) was counterbalanced across subjects.

To define perceptual thresholds, we measured the Just Noticeable Difference (JND) for each subject, for the subject's assigned face pair (Holt et al., 2014). Measurements of JND provide a standard way to define psychophysical thresholds, also known as a "difference limen", a "differential threshold", or "least perceptual differences", formulated originally well over a century ago by Weber, subsequently revised by Fechner (1860)

Here, the JNDs were measured by testing the accuracy of distinguishing the morphs from the CS+, using a 2-alternative forced choice (2-AFC) discrimination task (Fig. 1B). This 2-AFC task consisted of three runs of 50 trials each. At the beginning of each trial, the CS + stimulus was shown for 500 ms. After a 500 ms inter-trial interval, subjects were presented with the CS+ and a morph stimulus side by side and asked to select which stimulus they had previously seen by pressing the left or right key. The morph stimuli used in this task were 6, 12, 24, 48, and 100 *physical* steps from the CS+. The position of the CS+ and the morph stimulus were pseudo-randomized across trials and the subjects had unlimited time to respond. Responses were followed by a 1 s inter-trial interval. A Weibull function was fitted to the responses: $y = 1\text{-}e^{(-(x/a)^b)}$, in which y was the proportion of correct responses, x was the morph level, and a, and b were parameters for scale and shape, respectively. The JND corresponded to the morph level at which the subject achieved 75% accuracy on the 2-AFC task (Fig. 1C) (Clementz et al., 2007; Parkes et al., 2001; Skrandies and Fahle, 1994). Five generalization stimuli were then chosen for that subject based on this JND value, selected from the original 100 morphed images. These image morphs corresponded to .125, .25, 0.5, 1.0 and 1.5 of the JND for each subject. An example of this morph continuum is shown in Fig. 1D. It is known that perceptual thresholds (and thus JND values) can change with progressive exposure to stimuli over extended periods of time, due to a number of factors including perceptual learning (Schechtman et al., 2010; Resnik et al., 2011). To test whether such variation affected our measurements, we measured JNDs before and after scanning.

2.3. Experimental paradigm and procedure

While in the scanner, two measurements were acquired from each subject: 1) skin conductance responses, and 2) blood oxygen level-dependent (BOLD) signal, during both the fear conditioning and generalization phases. In addition, after the scanning session, subjects viewed

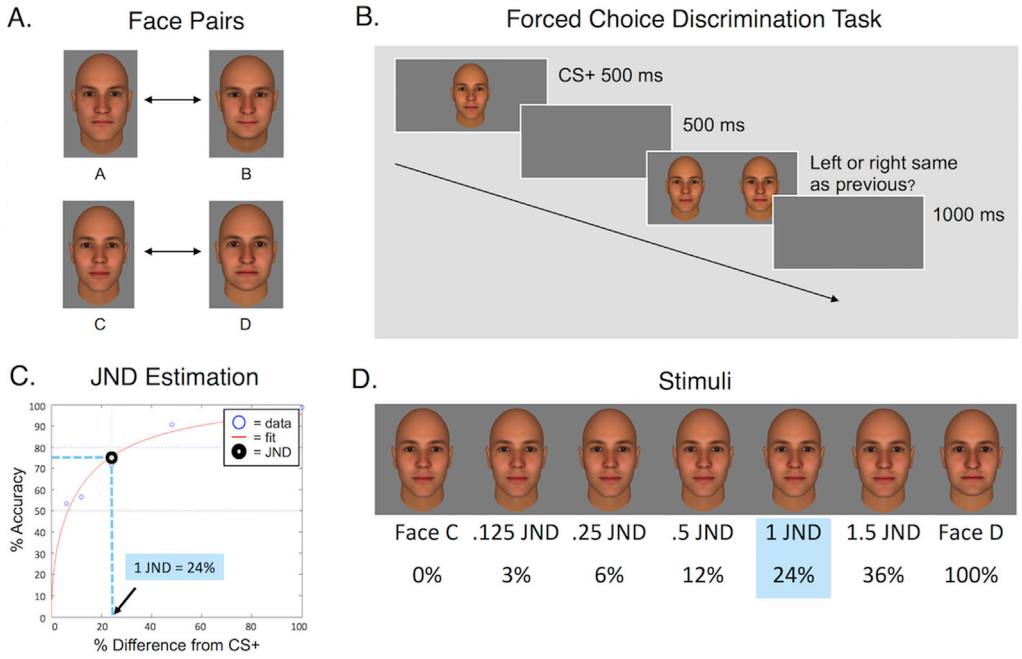

Fig. 1. The experimental design, illustrated in one subject and stimulus set. **A)** Two face pairs used in the experiment. **B)** A schematic of the 2-alternative forced choice (2-AFC) discrimination task used to measure the Just Noticeable Difference (JND) for each subject. **C)** An example of model fitting and estimation of the JND based on 75% accuracy in the 2-AFC discrimination task. **D)** An example of the stimulus set for one subject. The morph corresponding to the JND was the 24th face morph of the 100 equal steps between Face C and D. The selection of the morph stimuli presented in the main fear generalization experiments were based on this value, individually selected for each subject. Percentage values indicate the difference from the CS + stimulus.

each stimulus they had seen during the scan (the CS+, CS— and the five generalization stimuli) and rated the likelihood (0–100%) that each stimulus had been followed by a shock during the scan.

2.3.1. Fear conditioning

This phase consisted of one run (294 s) during which the CS+ and CS— stimuli were each presented 8 times (6 s each), in pseudorandom order with inter-trial intervals of 9–15 s. The CS+ was followed by the unconditioned stimulus (i.e., an electric shock) in 62.5% of the trials, whereas the CS— was never followed by a shock. Subjects were told that each face stimulus may or may not be followed by an electrical shock, and that they will be asked questions about what they had observed following the experiment.

2.3.2. Fear generalization

The generalization phase immediately followed the conditioning phase. It was divided into two runs (642 s in total), during which subjects were presented with the CS+, CS— and five JND levels (0.125–1.5 JND). During this phase, each stimulus was presented 5 times (6 s each), with 9–15 s inter-trial intervals. For each subject, stimuli were presented in one of two different pseudo-random orders, so that no more than two of the same stimuli were presented consecutively. The choice of order was counterbalanced across subjects. Unlike the conditioning phase, during generalization, the CS+ was always followed by an electrical shock (100% reinforcement), to minimize extinction. Instructions for the generalization phase were identical to the fear conditioning instructions above.

2.3.3. Electrical stimulation

Mild electrical shocks (0.5 s in duration) were delivered via a cable attached to the second and third finger of the subject's right hand using a finger stimulator (Coulbourne Instruments, Allentown, PA, USA). Before the scan, the intensity of the electric shock was set by each subject at a level that was "highly annoying but not painful" (Holt et al., 2009, 2014; Milad et al., 2005).

2.4. Skin conductance

Data acquisition and preprocessing: In order to record skin conductance responses (SCRs), two MRI-compatible electrodes were placed on the palm of each subject's left hand at the beginning of the scan session. Skin conductance was measured using a Biopac MP150 acquisition system (BIOPAC Systems, Inc, Goleta, CA) equipped with an EDA100C MRI amplifier. The conductance was collected at a gain of 5 μS/V, using a 1 Hz low pass filter during acquisition, then digitized at 200 Hz. The data was further filtered with a 3 Hz finite impulse response low pass filter, using the AcqKnowledge software package (BIOPAC Systems, Inc, Goleta, CA). Two time windows of the stimulus-related SCR were analyzed: 1) 3–6 s after stimulus onset during the conditioning phase, and 2) 6–9 s after stimulus onset during the generalization phase. Typically, peak skin conductance responses occur within 3–5s following the event (Dawson et al., 2007; Levinson and Edelberg, 1985). A pre-shock time window was selected for the fear conditioning phase, because responses to the electrical shock would have overlapped with most of the delayed responses to the CS+. Selection of these two time windows was based on a prior dataset, which revealed that, during conditioning, sufficiently large peak responses occur during stimulus presentation, whereas during generalization, the peak response occurs during the inter-trial interval following stimulus presentation (Holt et al., 2014). For each trial, we calculated the difference between the peak skin conductance value during the time window and the mean skin conductance level for the 2 s before the stimulus onset. The skin conductance measurements for all events were non-normally distributed (one-sample Kolmogorov-Smirnov test, $p < .05$). Although the ANOVA F-test is fairly robust to non-normality (Harwell et al., 1992), we chose to transform the data (square-root transformation), consistent with standard procedures in SCR analysis. The transformation resulted in distributions that were closer to normal, and these transformed data were used in all subsequent analyses, as in prior studies (LaBar et al., 1995; Milad et al., 2006).

Skin conductance data from four (of 37 total) subjects were excluded from further analyses due to the presence of scanner-induced artifacts. In

addition, subjects were excluded if they were non-responsive based on *a priori* criteria. A subject was considered a responder if at least 2 of the 16 trials of the fear conditioning phase showed a response greater than 0.05 mS (Holt et al., 2014; Turner et al., 2005). Four subjects were found to be non-responders and therefore excluded from the subsequent SCR analyses. Thus, the final sample included in the SCR analyses was composed of 29 subjects.

2.5. Functional MR imaging

MRI data acquisition and preprocessing: Magnetic resonance images (MRI) were acquired using a 3 T Tim Trio Siemens scanner with a 32-channel coil (Erlangen, Germany). T2*-weighted echo-planar images were collected during the fear conditioning and generalization phases (3 mm isotropic, matrix = 64 × 64, 45 slices, TR = 3000 ms, TE = 30 ms, flip angle = 90°). In addition, a T1-weighted 3D scan was collected using an MPRAGE sequence (spatial resolution 1 mm isotropic, matrix = 256 × 256, 176 slices, TR = 2530 ms, TE = 1.64, 3.5, 5.36, and 7.22 ms, flip angle = 7°). Functional MRI data were preprocessed using FreeSurfer version 5.3 (http://surfer.nmr.mgh.harvard.edu) (Fischl, 2012). All functional images were first realigned to correct for head motion, and then corrected for slice-timing differences. Next, the T1-weighted image was used to spatially transform the functional images into a common space (i.e., to fsaverage on the cortical surface and MNI305 in the subcortical volume). Finally, the fMRI data were spatially smoothed using a 3D Gaussian kernel (5 mm FWHM). Correction for temporal drift was incorporated within first-level modeling. Hemodynamic responses were estimated by fitting a convolution of a boxcar regressor with a ɣ function (delta = 2.25s, tau = 1.25s) to the fMRI data. Vertex and voxel-wise statistical tests were conducted by computing contrasts based on a univariate general linear model (GLM). All analyses used head motion parameters as regressors.

Absolute head motion did not exceed 2 mm for any run across all subjects. Across the whole sample, the maximum frame-wise displacement (Power et al., 2012) was 0.95 mm during the conditioning phase and 1.20 mm during the generalization phase; the average frame-wise displacement was 0.088 mm during the conditioning phase and 0.089 mm during the generalization phase. Thus, none of the subjects were excluded from the study due to excessive head motion. The effects of the electric shock on the fMRI responses during the CS + trials were accounted for by using the magnitude of activity evoked 3–9 s following shock onset as a regressor in the fMRI analyses.

2.6. Explicit ratings

Following the scanning, subjects were presented with two trials each of the CS+, CS—, and the five generalization stimuli that they viewed during the scanning, and asked to rate the likelihood that the stimulus had ever been followed by a shock during the fMRI experiment, on a scale of 0–100%. These ratings were collected to test the extent to which subjects 1) retrieved the learned association between the CS + face and the electrical shock, and 2) showed explicit generalization of this fear memory.

2.7. Experimental design and statistical analyses

All statistical analyses were conducted with the Statistics and Machine Learning Toolbox of Matlab R2015b (MathWorks, Inc. Natick, MA, USA), unless stated otherwise.

Skin Conductance: To test for a fear conditioning response, the mean SCRs to the CS+ and CS— stimuli during fear conditioning were compared using a paired *t*-test. Generalization of conditioned fear responses was assessed using repeated measures ANOVA with six within-subject levels (0.125, 0.25, 0.5, 1.0, 1.5 JND and CS—). The CS+ was not included in the tests of fear generalization because responses to the unconditioned stimulus (the electrical shock) likely confounded measurement of responses to the CS + for the time window used in these analyses (see above).

Functional MRI: Brain responses to the CS+ and CS— stimuli during fear conditioning were compared using a two-stage random effects model (Holmes and Friston, 1998) implemented in FreeSurfer. For each subject, vertex and voxel-wise contrasts between CS + vs. CS— events during conditioning were computed by applying a mass univariate general linear model (GLM). In a second GLM, these contrast maps were used to create group-level statistical maps. The group-level maps were then corrected for multiple comparisons within each space (left and right cortical surfaces and the subcortical space) using a Monte Carlo simulation with 10,000 iterations, and across the three spaces (left and right cortical surface and subcortical volume), using a Bonferroni correction. The vertex- and voxel-wise cluster-forming thresholds were set at $p < .001$, and the cluster-wise p-value was set at $p < .05$.

Generalization of BOLD responses was tested using functionally-defined regions-of-interest (ROIs), which included the 26 clusters in the cortex, plus one cluster in the right cerebellum that showed a significant response to the CS + vs. CS— contrast during the conditioning phase. We also examined generalization in the hippocampus, caudate nucleus and thalamus using anatomically-defined ROIs generated from each participant's T1 scan using FreeSurfer (Fischl et al., 2002), because prior studies have reported generalization in these regions, and the stringent level of correction for multiple comparisons in the analysis of the conditioning phase of our experiment may have resulted in Type II error for these relatively small regions. The amygdala was not included among these ROIs because findings of a recent meta-analysis suggested that the involvement of the amygdala in fear conditioning cannot be reliably measured using conventional fMRI in humans (Fullana et al., 2016). In addition because fear generalization and the potentially related process of pattern separation have been localized to distinct subregions of the hippocampus in prior studies (Bakker et al., 2008; Lissek et al., 2014a; Onat and Büchel, 2015), we examined the head, body and tail of the hippocampus independently using the following segmentation algorithm:

https://surfer.nmr.mgh.harvard.edu/fswiki/HippocampalSubfieldsAndNucleiOfAmygdala (Iglesias et al., 2015). As in the SCR analyses, the CS + trials were excluded from the analysis of the fear generalization phase, because of the artefactually higher responses to the CS + due to the close temporal proximity of the shock. Therefore, the BOLD responses were analyzed using repeated measures ANOVA with six within-subject levels (0.125, 0.25, 0.5, 1.0, 1.5 JND, and CS—).

Explicit ratings: Generalization in the explicit ratings was tested using repeated measures ANOVA using seven within-subject levels (CS+, 0.125, 0.25, 0.5, 1.0, 1.5 JND, and CS—), followed up with two-tailed *post hoc* paired t-tests comparing the CS+, the generalization stimuli and the CS—. We also conducted this analysis using the 1.5 JND as the baseline condition, as previously described (Holt et al., 2014), because the average explicit ratings to the CS— were close to 0.

Hypothesis testing: In all analyses, whenever the assumption of sphericity was not met (Mauchly's test $p < .05$), the degrees of freedom in the repeated measures ANOVA were adjusted with a Greenhouse-Geisser correction. *Post hoc* paired t-tests were calculated to compare the response to each generalization stimulus to the response to the CS—. In all analyses, p-values $< .05$ were considered statistically significant and all t-tests were two-sided.

Our goal was to determine whether each type of generalization response (SCR, fMRI or ratings) was - or was not - constrained by the perceptual limits defined in the initial psychophysical assessment (as illustrated in Fig. 1), i.e., whether it met the predictions of the perceptual hypothesis of fear generalization. This question can be tested using two slightly different criteria, here termed the "fixed threshold" criterion, and the "continuous threshold" criterion.

A "fixed threshold" was defined here as the JND at which subjects performed at 75% accuracy (see above), and subsequent statistical tests determined whether the responses were greater than, or less than, this

fixed threshold. According to this criterion, the prediction of the perceptual hypothesis is that responses will be higher than the CS— at JND values of 0.125, 0.25, and 0.5. Conversely, responses are predicted to be indistinguishable from those to the CS— at JND values of 1.5 JND. In this fixed threshold model, responses to 1.0 JND are undefined.

In contrast, the "continuous threshold" criterion accounts for more of the known properties of psychophysical thresholds. Given adequate measurement resolution, it is well known that 1) JND thresholds are graded rather than absolute, and 2) the JND value for a given range of graded stimuli varies with the performance accuracy that is chosen to define the threshold. Here we used a standard threshold (75% correct), at which subjects could discriminate morphed faces from the CS + face, at a level that is intermediate between "never" (50% correct) and "always" (100% correct). At this threshold, the perceptual hypothesis predicts that the face corresponding to 1.0 JND will also produce a response of intermediate (not zero) magnitude. Thus, relative to the responses to the 1.0 JND stimulus, this continuous threshold criterion predicts increased responses at .5 JND and decreased (but greater than zero) responses to the 1.5 JND stimulus.

Gaussian modeling of fear generalization: Previous work has modeled generalization gradients as a Gaussian curve (Ghirlanda and Enquist, 2003). Accordingly here, the extent of generalization in the explicit ratings, the SCR, and the fMRI responses of the anterior insula and superior frontal gyrus, were each examined by fitting these data with a Gaussian function, $f(x) = a*\exp(-(b-x)^2/2c^2))$, where a is the height of the curve and c the width, and b is the location of the peak. The b parameter was fixed at 0, i.e. the CS+. The extent of fear generalization is represented by the parameter c. The Gaussian function was fitted to these data using non-linear least squares; the responses were scaled prior to fitting by subtracting the CS— value from the other values.

2.8. Secondary analyses

Associations among fear generalization measures: We also tested whether the magnitude of generalization of neural responses (measured with fMRI) was correlated with: 1) the explicit shock likelihood ratings and 2) the SCRs during fear generalization. For these analyses, a linear mixed effects model was fitted using the lmer function in the lme4 package (http://cran.r-project.org/web/packages/lme4/index.html) in R version 3.4.0 (R Core Team, 2017). These analyses were limited to the anterior insula and superior frontal gyrus because, similar to the SCR and explicit ratings, these two regions showed higher responses to the CS+ and generalization stimuli compared to the CS— (see Results). Because such responses were essentially identical in left and right hemispheres, data were averaged across hemispheres. To account for differences in magnitudes of overall BOLD, SCR, and explicit rating responses, the data were normalized by subtracting the response to the 1.5 JND from the responses to the 0.125, 0.25, 0.5 and 1.0 JND stimuli. We fitted a separate model for the anterior insula and superior frontal gyrus, with either the SCR or the explicit ratings as the dependent variable, with the intercept as a random factor, and the stimuli by BOLD interaction as fixed effects.

The influence of sex on fear conditioning and generalization: Because of known sex differences in fear conditioning and fear extinction (Merz et al., 2018), we examined whether sex influenced any measure of fear conditioning or generalization, by using repeated measures ANOVA for SCRs and GLM for fMRI responses.

3. Results

3.1. Perceptual threshold (JND) measurements

JND measurements before and after the experiment did not differ (paired *t*-test $t_{36} = -0.197$, $p = .845$).

3.2. Skin conductance responses

Fear Conditioning: During the fear conditioning phase, subjects showed significantly higher SCRs to the CS + compared to the CS—, (paired *t*-test $t_{28} = 3.86$, $p < .001$; Fig. 2A).

Fear Generalization: We found a monotonic decrease in the mean response with higher JND level. A repeated measures ANOVA showed a significant effect of JND level ($F_{3.33,\ 93.22} = 2.97$, $p = .031$; Fig. 2B). *Post hoc* paired t-tests showed that the SCR was higher to the 0.125 JND ($t_{28} = 2.52$ $p = .018$), 0.25 JND ($t_{28} = 2.57$, $p = .016$), and 0.5 JND ($t_{28} = 2.13$, $p = .043$), compared to the CS— (Fig. 2B). The SCR responses to 1.0 JND and 1.5 JND did not significantly differ from the responses to the CS— (1.0 JND vs. CS— $t_{28} = 0.73$, $p = .468$ and 1.5 JND vs. CS— $t_{28} = 1.01$, $p = .319$).

3.3. Functional MRI

Fear Conditioning: The analysis revealed 26 clusters in the cortex (localized using the Desikan-Killany Atlas) which showed a significant CS+ > CS— or CS— > CS + response; these are shown in Fig. 3. One statistically significant cluster was also found within subcortical space (identified using a $p < .001$ cluster forming threshold, a CS— > CS + response), with a peak in the right cerebellar cortex (MNI = 32, −75, −39; 872 mm^3). A more lenient cluster-forming threshold of $p < .05$ showed 3 statistically significant clusters in the subcortical space, including the bilateral thalamus, caudate and cerebellum, and left hippocampus (Supplementary Fig. 1).

Fear Generalization: 22 of the 26 cortical clusters that showed significant activation during conditioning, and 5 of the 11 subcortical regions of interest also showed a significant effect of stimulus type during generalization (Table 1). Eleven of these generalization effects met a Bonferroni-corrected level of significance threshold. Among these regions, the anterior insula and superior frontal gyrus exhibited significantly greater responses to the generalization stimuli compared to the

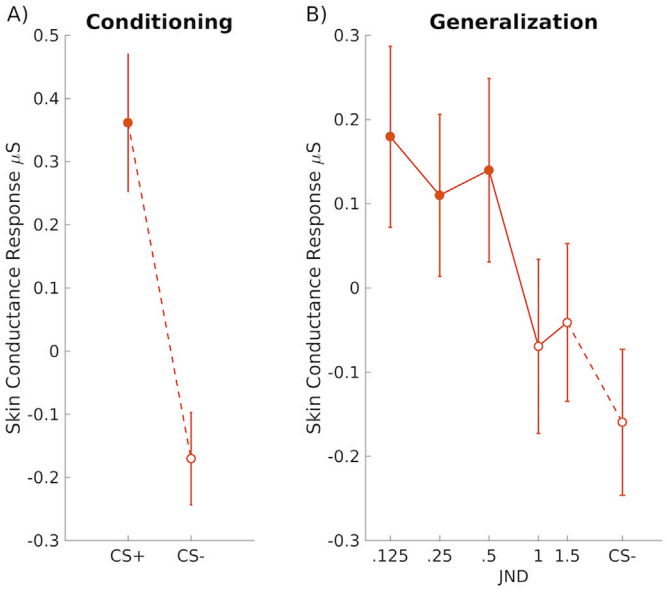

Fig. 2. A) Subjects showed significantly higher skin conductance responses (SCRs) to the CS + stimulus, compared to the CS— stimulus, during the fear conditioning phase. **B)** Subjects showed generalization in the SCRs to the near-CS + stimuli during the fear generalization phase. Filled circles indicate a significant difference in response to the stimulus compared to the CS—, $p < .05$ (paired *t*-test). Error bars represent standard errors of the mean. SCRs are square root transformed.

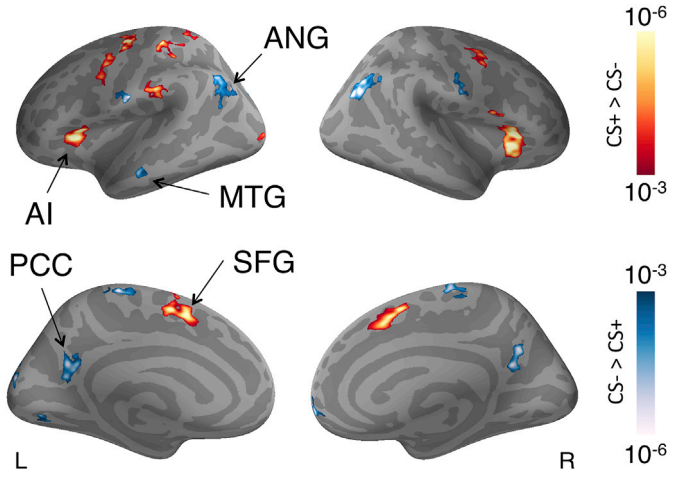

Fig. 3. BOLD responses during the fear conditioning phase to the CS + versus CS— contrast displayed on the cortical surface. Significantly higher responses to the CS + compared to the CS— (red-to-yellow) were found in the anterior insula (AI), superior frontal gyrus (SFG) and in the pre- and postcentral gyrus (see Table 1). Significantly lower activity to the CS + compared to the CS— (blue-to-white) was found in the angular gyrus (ANG), posterior cingulate cortex (PCC) and in the middle temporal gyrus (MTG) among other regions. The maps were Montecarlo corrected for multiple comparisons, using a cluster-forming p-value < .001, and cluster-wise p-value < .05. The color scale indicates vertex-wise p-values.

CS—, with a gradual fall-off of response magnitude beginning between 0.5 and 1.0 JND (Fig. 4A). Specifically, the bilateral anterior insula (repeated measured ANOVA, $F_{3.49,125.63} = 10.04$, $p < .001$) showed significantly greater responses to the 0.125, 0.25, 0.5, 1.0 and 1.5 JND, compared to the CS—, whereas the bilateral superior frontal gyrus (repeated measured ANOVA, $F_{3.71,133,45} = 5.03$, $p = .0011$) showed significantly greater responses to the 0.125, 0.25 and 0.5 JNDs, but not the 1.0 or 1.5 JNDs, compared to the CS— (Table 2).

Other regions, including the bilateral posterior cingulate cortex (repeated measured ANOVA, $F_{4.37,157,34} = 7.32$, $p < .001$), bilateral angular gyrus (repeated measured ANOVA, $F_{4.02,144.57} = 7.79$, $p < .001$; Fig. 4B), and the bilateral head (repeated measures ANOVA, $F_{3.88,139.53} = 6.55$, $p < .001$) and body of the hippocampus (repeated measures ANOVA, $F_{4.20,151.09} = 3.61$, $p = .0067$) (Supplementary Fig. 2) showed significantly lower (and broader) responses to the generalization stimuli compared to the CS—, with differential responses at all JND levels, compared to the CS— (Table 2).

3.4. Explicit ratings

Repeated measures ANOVA showed a significant effect of JND level ($F_{3.78,\ 135.90} = 64.44$, $p < .001$) in the explicit ratings (Fig. 5). *Post hoc* paired t-tests showed that subjects rated the CS+, and all of the JND levels, as more likely to be followed by the shock than the CS— (CS + vs. CS—, $t_{36} = 18.73$, $p < .001$; .125 JND vs. CS—, $t_{36} = 14.96$, $p < .001$; .25 JND vs. CS—, $t_{36} = 17.30$, $p < .001$; .5 JND vs. CS—, $t_{36} = 18.48$, $p < .001$; 1.0 JND vs. CS—, $t_{36} = 9.78$, $p < .001$; 1.5 JND vs. CS—, $t_{36} = 5.11$, $p < .001$).

3.5. Gaussian function

To quantify and compare the extent of fear generalization in our data, we fitted a Gaussian function to the responses (Supplementary Fig. 3). In these analyses, the parameter c reflects the extent of fear generalization (see Methods for details). For the explicit ratings and anterior insula and superior frontal gyrus responses, $c > 0$, consistent with the presence of generalization of the conditioned response to the CS+ in these data. Based on the rank order, the extent of generalization was larger in the anterior insula (c = 40.9, 95% CI = 19.6–62.2) compared to the explicit ratings (c = 34.8, 95% CI = 29.9–39.6). In contrast, the extent of generalization was lower in the superior frontal gyrus (c = 25.0, 95% CI = 9.3–41.4), and in the skin conductance responses (c = 27.2, 95% CI = −0.8–55.2), compared to the explicit ratings.

3.6. Linear mixed models

Using linear mixed models, we found that the anterior insula responses during fear generalization were strongly linked with both the fear generalization SCRs ($F_{1,106.39} = 13.54$, $p < .001$) and the ratings of shock likelihood ($F_{1,136.37} = 8.19$, $p = .0048$). Responses of the superior frontal gyrus during fear generalization were associated with the SCRs ($F_{1,99.66} = 6.20$, $p = .014$) but not with the explicit ratings ($F_{1,139.76} = 2.52$, $p = .11$). We found no significant BOLD by stimulus interactions. The associations between the anterior insula BOLD responses, SCRs, and explicit ratings are illustrated in Supplementary Fig. 4.

3.7. The influence of sex on fear conditioning and generalization

We found no effects of sex in the SCRs ($p > .05$) or the fMRI responses (cluster-forming threshold $p < .001$, Monte-Carlo corrected for multiple comparisons) during the conditioning phase. During the generalization phase, there were no sex by JND-level interactions found in the fMRI responses, SCRs or explicit ratings (all $p > .05$), but significant or near significant main effects of sex were observed: compared to females, males had on average higher SCRs ($t = 2.76$, $p = .010$, $df = 27$) and fMRI responses in the superior frontal gyrus ($t = 2.32$, $p = .026$, $df = 35$), and at a trend-level in the anterior insula ($t = 1.96$, $p = .058$, $df = 35$), but a trend towards lower mean shock likelihood ratings ($t = -1.97$, $p = .056$, $df = 35$).

4. Discussion

4.1. Summary of findings

Here, fear generalization was measured using a Pavlovian fear conditioning and generalization paradigm that accounted for the perceptual discrimination threshold of each subject. Overall, we observed a general pattern, across measurements of skin conductance, shock likelihood estimates, and fMRI responses in specific brain regions (anterior insula and superior frontal cortex) of significant fear generalization when discriminability of stimuli from the CS+ was low (when the generalization stimuli were most similar to the CS+, and least similar to the CS—). Conversely, such measures showed little or no fear generalization when discriminability of stimuli from the CS+ was high. Such results are generally consistent with the perceptual hypothesis of fear generalization.

Other fear generalization measurements corresponded less well to perceptual gradients. Most clearly, certain regions of the default network (posterior cingulate cortex, angular gyrus and hippocampus) showed 1) an *inverted* BOLD response to fear generalization, which also 2) extended beyond the highest JND value tested here. In addition, while the explicit ratings and responses of the anterior insula generally met the predictions of the perceptual hypothesis (according to a continuous threshold model, see Methods), there was also evidence for broadening of these gradients beyond perceptual limits, due to a significantly greater responses to the 1.5 JND stimulus, compared to the CS—. In future investigations, this evidence for possible supra-threshold generalization in the anterior insula and behavior can be investigated further to identify any potential mechanisms.

4.2. Relevance to prior research

The overall objectives of this study reflect a larger and long-running

Table 1
Regions activated during fear conditioning ($p < .001$, corrected) that also showed statistically significant ($p < .05$, uncorrected; for 11 regions, Bonferroni-corrected) responses during fear generalization.

		Fear Conditioning			Fear Generalization (repeated measures ANOVA)		
Region	Hemisphere	Contrast	Size (mm^2)	Peak MNI coordinates (x, y, z)	F	DF	P-value
Anterior Insula	L	CS+ > CS-	297.5	-28.5, 24.8, 8.9	6.2	3.64, 131.06	< .001*
	R	CS+ > CS-	511.6	36, 26.9, 7.9	12.91	3.73, 134.22	< .001*
Superior Frontal Gyrus	L	CS+ > CS-	435.6	-8.7, 6.9, 51	4.09	3.90, 140.24	.004
	R	CS+ > CS-	322.5	7.3, 5.7, 54.1	5.77	3.76, 135.42	< .001*
Postcentral Gyrus	L	CS+ > CS-	300.0	-43, -30.9, 60.4	4.11	4.10, 147.64	.003
Superior Parietal Gyrus	L	CS+ > CS-	195.3	-19.3, -45.9, 66.1	3.6	4.22, 151.92	.007
Occipital Cortex	L	CS+ > CS-	170.5	-25.9, -92.5, -0.6	3.36	4.07, 146.65	.011
	L	CS+ > CS-	122.2	-28.9, -86.9, -16.2	2.63	4.37, 157.38	.032
	R	CS+ > CS-	215.1	27.9, -84.2, -11.1	2.98	4.12, 148.35	.020
Angular Gyrus	L	CS- > CS+	476.0	-46.1, -63.5, 38.3	9.53	3.87, 139.38	< .001*
	R	CS- > CS+	567.1	45.2, -66.9, 28.2	3.73	3.67, 132.05	.008
Posterior Cingulate Cortex	L	CS- > CS+	328.6	-4.5, -63, 20.8	8.05	4.54, 163.59	< .001*
	R	CS- > CS+	229.2	4.7, -59.6, 19.1	5.65	4.00, 143.92	< .001*
Middle Temporal Gyrus	L	CS- > CS+	104.3	-57.2, -16.8, -18.1	7.93	3.35, 120.65	< .001*
Precentral Gyrus	L	CS- > CS+	474.8	-34.6, -11.8, 58.2	3.29	4.04, 145.40	.013
Postcentral Gyrus	L	CS- > CS+	103.1	-63.4, -10.3, 23.2	2.43	4.23, 152.12	.047
	R	CS- > CS+	231.5	60.8, -8.6, 23.6	5.13	3.93, 141.61	< .001*
Paracentral Gyrus	L	CS- > CS+	145.5	-5.0, -33.0, 61.9	3.1	3.81, 137.30	.019
	R	CS- > CS+	164.5	5.6, -22.1, 67.4	3.25	4.23, 152.14	.012
Orbitofrontal	R	CS- > CS+	164.8	9.2, 59.5, -3.8	2.76	4.35, 156.58	.026
Occipital Cortex	L	CS- > CS+	568.3	-22.2, -85.3, 25.3	7.03	3.52, 126.86	< .001*
	L	CS- > CS+	172.2	-11.9, -78.3, -5.1	3.06	4.00, 143.86	.019
Cerebellum	R	CS- > CS+	872**	32, -75, -39	4.10	4.21,151.46	.003
Hippocampal head***	L	-	-	-	6.72	3.99, 143.47	< .001*
	R	-	-	-	4.62	3.88, 139.55	.002
Hippocampal body***	L	-	-	-	5.24	4.56, 164.16	< .001*
Hippocampal tail***	R	-	-	-	2.60	4.36, 156.93	.034

* indicates responses that survived Bonferroni correction for multiple comparisons.
** mm3, *** Anatomically defined regions-of-interest.

debate about whether fear generalization is limited by perceptual boundaries, or whether it extends beyond those limits (Guttman and Kalish, 1956; Struyf et al., 2015). In other words, does fear generalization merely result from an inability to discriminate the CS+ from perceptually similar stimuli (i.e. the perceptual hypothesis)? Alternatively, does it also involve an active mechanism, such as fear responses to stimuli that are similar but clearly distinguishable from the CS+ and/or contributions from other non-perceptual processes (e.g., decision-making and uncertainty assessment mechanisms)?

A prior fMRI study also examined the effects of graded physical changes in stimuli to fear generalization gradients in order to investigate this question (Onat and Büchel, 2015). In contrast to the current results, that study reported evidence for "hyper-sharp" responses in anterior insula during fear generalization. Specifically, responses in the anterior

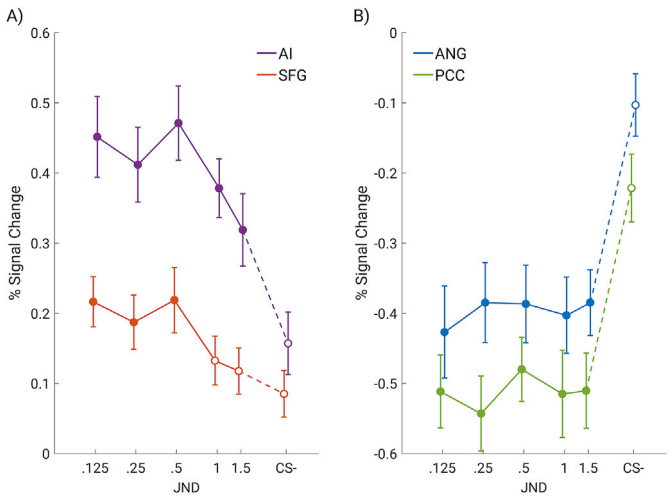

Fig. 4. BOLD responses during the fear generalization phase for selected regions of interest. Data for the two hemispheres are combined in these plots. **A)** For the bilateral anterior insula (AI) and superior frontal gyrus (SFG) there was a positive association between the magnitude of the BOLD responses and the similarity of the generalization stimuli to the CS+. **B)** In contrast, the angular gyrus (ANG) and posterior cingulate cortex (PCC) showed significantly lower responses to the CS+ and CS + -like generalization stimuli, compared to the CS— stimulus. Filled circles indicate a significant difference in response between the stimulus and the CS—, $p < .05$ (paired *t*-test). Error bars represent standard errors of the mean.

insula were found to be more narrowly tuned (i.e., had steeper generalization fall-offs) to systematic physical variations in the generalization stimuli, relative to the behavioral fear generalization responses. However, because perceptual discrimination thresholds were not systematically controlled in the subjects in this study, it is unclear how the observed hyper-sharp responses in anterior insula were related to perception.

4.3. The role of the anterior insula and superior frontal gyrus in fear generalization

The anterior insula and superior frontal gyrus both showed higher responses to the CS+ and generalization stimuli, compared to the CS—. This similarity in the pattern of generalization responses across these two regions is consistent with prior studies showing that these two areas are anatomically connected (Jürgens, 1984; Vogt, 2016; Yeo et al., 2011). The cluster within the superior frontal gyrus showing fear generalization was located in the supplementary motor area (see http://neurosynth.org (Yarkoni et al., 2011) and within the Salience Network (SN; Seeley et al., 2007) according to functional connectivity criteria (Yeo et al., 2011). Recent functional connectivity findings and a diverse array of task-based fMRI studies conducted in humans (Gasquoine, 2014; Sterzer and Kleinschmidt, 2010) suggest that the human SN, and the anterior insula in particular, has a broader and more integrative role in the brain, compared to what was originally suggested by studies conducted in rodents, which emphasized the role of this network in interoception (Menon and Uddin, 2010). Specifically, recent human fMRI data suggest that the anterior insula integrates activity across multiple sensory and cognitive networks (Sidlauskaite et al., 2014). This hub-like function allows the anterior insula to recruit a range of processing and attention resources to execute diverse behaviors (Sterzer and Kleinschmidt, 2010). For example, the human anterior insula is activated during difficult cognitive tasks, such as those involving perceptual decision-making or categorization of ambiguous stimuli, which require effortful processing or resolving conflicts (Seger et al., 2015). Significant correlations between task performance difficulty or stimulus ambiguity, relative to anterior insula responses, have suggested that the activity of the anterior insula is linked to the accumulation of sensory evidence required for making a perceptual decision (Ho et al., 2009; Pedersen et al., 2015), as well as encoding confidence that perceptions are accurate (Hebart et al., 2016; Paul et al., 2015).

Other studies have suggested that anterior insula activity may reflect the subjective experience of sensory input (Wiech et al., 2010). For instance, several fMRI studies of pain processing reported that levels of activity within the anterior insula correlate with explicit ratings of subjective pain intensity (Favilla et al., 2014; Kong et al., 2006; Ploner et al., 2010). Similarly, here we found a correlation between anterior insula responses during fear generalization and explicit shock likelihood ratings, as well as with skin conductance responses. Thus our results add to prior evidence that the anterior insula plays a key role in the subjective awareness of stimuli in the external environment (Craig, 2009). During fear generalization, the anterior insula may contribute to a conscious experience of fear and subsequent initiation of defensive responses to the CS+ and generalization stimuli.

Our data are also consistent with prior evidence suggesting that the anterior insula and the closely connected areas of midline cortex, such as the anterior cingulate cortex and supplementary motor area, have distinct functions (Craig, 2009; Sterzer and Kleinschmidt, 2010). Here, the midline cortical area showing fear generalization (located within the superior frontal gyrus, the supplementary motor area) showed a lower extent and magnitude of fear generalization than the anterior insula. Also, the magnitude of fear generalization responses of the anterior insula, but not of the superior frontal gyrus/supplementary motor area region, correlated with the magnitude of the explicit shock likelihood ratings. These findings are consistent with a role of the anterior insula in awareness and multi-modal sensory integration (here involving visual and tactile sensory modalities). In contrast, the supplementary motor area may play a central role in the rapid initiation of actions in response to salient information in the environment (Nachev et al., 2008).

4.4. The role of the default network in fear generalization

Regions of the default network (DN) (Raichle et al., 2001), including the posterior cingulate cortex, angular gyrus, hippocampus, and the left middle temporal gyrus, showed significant (albeit negative)

Table 2
Paired t-tests comparing each JND level to CS— in different brain regions of interest.

	0.125 JND vs. CS—		0.25 JND vs. CS—		0.5 JND vs. CS—		1 JND vs. CS—		1.5 JND vs. CS—	
	t	p	t	p	t	p	t	p	t	p
AI	4.33	<.001	3.94	<.001	4.64	<.001	4.11	<.001	2.78	.008
SFG	3.01	.005	2.42	.021	2.99	.005	1.30	.200	1.04	.303
PCC	−4.52	<.001	−4.51	<.001	−4.05	<.001	−4.04	<.001	−3.86	<.001
ANG	−4.82	<.001	−4.14	<.001	−4.09	<.001	−4.47	<.001	−4.06	<.001
HIPP head	−2.98	.005	−3.51	.001	−4.71	<.001	−2.78	.008	−2.50	.017
HIPP body	−2.67	.011	−3.16	.003	−2.90	.006	−4.99	<.001	−2.71	.010
HIPP tail	−1.47	.151	−2.57	.014	−1.53	.136	−3.70	<.001	−3.04	.004

AI = anterior insula, SFG = superior frontal gyrus, PCC = posterior cingulate cortex, ANG = angular gyrus, HIPP = hippocampus. t = t-value of a paired *t*-test, p = p-value of a paired *t*-test.

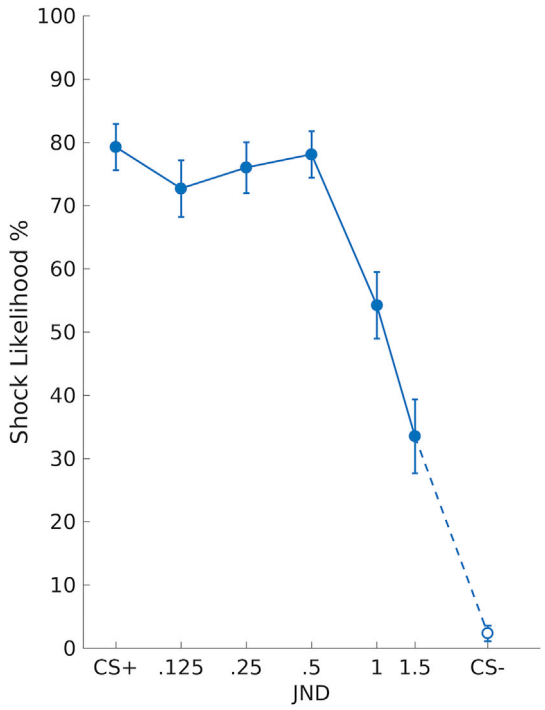

Fig. 5. The explicit ratings of shock likelihood decreased as stimulus similarity (to the CS+) decreased. Filled circles indicate a significant difference in response between the stimulus and the CS—, $p < .05$ (paired *t*-test). Error bars represent standard errors of the mean.

generalization responses at all JND levels tested, i.e. with little evidence for a step-wise gradient of responses). However, the current experiment did not measure fear generalization responses beyond 1.5 JND; thus the extent and pattern of deactivation in the DN regions cannot be precisely determined using these data. Prior studies have shown that the DN is primarily involved in processes requiring introspection that involve shifting attention inward, away from the external environment (Buckner et al., 2008). A lower response to fear-evoking stimuli (compared to the CS—) is consistent with prior findings of deactivation of the DN when attentional resources are shifted to cognitive control or salience detection–related processes, i.e. in response to behaviorally salient information in the environment (Anticevic et al., 2012; Mantini et al., 2011). Thus, DN suppression in response to stimuli that are similar but clearly not identical to threatening stimuli (including the supra-threshold responses here) may play a role in a larger defensive response involving the Salience Network (Goulden et al., 2014; Sridharan et al., 2008), which permits further assessment of ambiguous and potentially threatening stimuli.

4.5. Other types of fear generalization

Depending on the context, generalization of fear responses may also reflect dimensions other than perceptual similarity (Dymond et al., 2015). For example, Dunsmoor and colleagues (2012) showed that following category-based learning, fear responses generalized across an entire category of stimuli, even when individual stimuli were easily distinguished from each other at a perceptual level. Future studies can further characterize the brain networks involved in forms of fear generalization that have little perceptual basis.

4.6. Potential effects of increased arousal on perceptual thresholds

In this study, generalization stimuli were individualized based on each subject's perceptual threshold, measured in a non-threatening neutral context, before the fear conditioning and generalization procedure was conducted. However it is known that emotional states can affect perception (Geuss et al., 2016; Stefanucci and Storbeck, 2009). For example, fear of heights increases height estimates (Teachman et al., 2008). Thus, it can be argued that increased levels of emotional arousal during the fear conditioning and generalization procedures affected discrimination levels (Struyf et al., 2015), increasing perceptual thresholds during and/or after aversive conditioning (Laufer and Paz, 2012; Resnik et al., 2011; Shalev et al., 2018). Such increased thresholds during and following aversive conditioning have been shown to persist for 24 h (Resnik et al., 2011). However, in the current study, perceptual thresholds were similar before and after fear conditioning and generalization, suggesting that perceptual thresholds remained stable throughout the experiment. This stability may be due to the fact that the fear conditioning phase here was relatively short (~5 min). However, to investigate this question further, future studies should measure perceptual thresholds at multiple time points, including during the fear conditioning and generalization procedure itself.

4.7. Sex differences

In an exploratory analysis, we examined the effects of sex on fear conditioning and generalization. Our analyses did not reveal any differences between the males and females during fear conditioning, which is in line with the results of a recent meta-regression analysis (Fullana et al., 2016). However, the literature on sex effects on fear conditioning in humans is somewhat inconsistent, with findings of similar responses in males and females (Guimaraes et al., 1991), as well as higher (Lebron-Milad et al., 2012; Zorawski et al., 2005) and lower (Milad et al., 2006) conditioned responses in females compared to males. We also did not observe sex by JND-level interactions in any of our measures of fear generalization. That is, the pattern of generalization was similar in males and females. However, we did find that males had overall higher SCRs and fMRI responses in the anterior insula and superior frontal gyrus than females, whereas males showed a trend towards lower explicit rating scores than the females, during fear generalization. These findings must be considered preliminary given the small sample sizes for the male and female groups and the known complexity of the effects of gonadal sex hormones (which were not measured in the current study) on fear learning and memory processes. For instance, depending on hormonal status (high estrogen, low estrogen or oral contraception use), females may show either greater or less fear conditioning than males (Hwang et al., 2015). Similarly, depending on menstrual cycle phase, females may show either greater or less fear extinction learning than males (Merz et al., 2018). In future studies, sex effects on fear generalization should be examined further while accounting for menstrual cycle phase, gonadal steroid levels and other factors that may influence sex-related effects on brain function.

4.8. Limitations

We were able to detect fear generalization in skin conductance responses using standard analysis methods, and these skin conductance responses were highly correlated with the magnitude of fear generalization observed in the fMRI data. However, novel approaches such as psychophysiological modeling, which are not limited to measurement of stimulus-evoked peak response amplitudes (Bach et al., 2018), may have even greater power to detect subtle variation in skin conductance responses during fear generalization. Similarly, we did not measure multivariate patterns of BOLD activity; in future studies, this approach could be used to measure generalization of conditioned fear responses of the amygdala (Bach et al., 2011; Staib and Bach, 2018; Visser et al., 2011), which were not detected in the current study.

4.9. Implications for neuropsychiatric disorders

The generalization of fear responses may be disrupted in the context

of certain brain disorders, giving rise to specific symptoms. For example, overactive fear generalization processes have been observed in patients with Panic Disorder (Lissek et al., 2010), Generalized Anxiety Disorder (Lissek et al., 2014b), Social Anxiety Disorder (Ahrens et al., 2016), and Post-Traumatic Stress Disorder (Kaczkurkin et al., 2017). However, some studies failed to detect over-generalization in Generalized Anxiety Disorder (Greenberg et al., 2013b; Tinoco-González et al., 2015). It is possible that discrepancies among these prior studies could be resolved by taking individual perceptual discrimination abilities into account. The experimental paradigm used here provides a means for distinguishing the effects of fear learning and memory-related processes involved in fear generalization from perceptual discrimination ability. Quantitative measurement of the distinct components of fear generalization may permit the characterization of distinct fear generalization phenotypes, which could serve as treatment targets, or used to identify individuals who are at risk for certain neuropsychiatric conditions. Future studies could then investigate whether excessive or diminished generalization in certain psychopathological states is related to perceptual impairments, over or under-retrieval of fear or safety associations, or changes in levels of subjective awareness.

Declarations of interest

None.

Acknowledgements

The authors would like thank Douglas Greve, Scott Orr, Clas Linnman, and Garth Coombs III for highly valuable assistance with the study. The study was supported by NIMH grant R01MH095904 (Holt) and the Sigrid Juselius Fellowship grant (Tuominen).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2018.12.034.

References

Ahrens, L.M., Pauli, P., Reif, A., Mühlberger, A., Langs, G., Aalderink, T., Wieser, M.J., 2016. Fear conditioning and stimulus generalization in patients with social anxiety disorder. J. Anxiety Disord. 44, 36–46.
Anticevic, A., Cole, M.W., Murray, J.D., Corlett, P.R., Wang, X.J., Krystal, J.H., 2012. The role of default network deactivation in cognition and disease. Trends Cognit. Sci. 16, 584–592.
Bach, D.R., Castegnetti, G., Korn, C.W., Gerster, S., Melinscak, F., Moser, T., 2018. Psychophysiological modeling: current state and future directions. Psychophysiology e13214.
Bach, D.R., Weiskopf, N., Dolan, R.J., 2011. A stable sparse fear memory trace in human amygdala. J. Neurosci. 31, 9383–9389.
Bakker, A., Kirwan, C.B., Miller, M., Stark, C.E., 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. Science 319, 1640–1642.
Buckner, R.L., Andrews-Hanna, J.R., Schacter, D.L., 2008. The brain's default network: anatomy, function, and relevance to disease. Ann. N. Y. Acad. Sci. 1124, 1–38.
Clementz, B.A., McDowell, J.E., Dobkins, K.R., 2007. Compromised speed discrimination among schizophrenia patients when viewing smooth pursuit targets. Schizophr. Res. 95, 61–64.
Craig, A.D., 2009. How do you feel–now? The anterior insula and human awareness. Nat. Rev. Neurosci. 10, 59–70.
Dawson, M.E., Schell, A.M., Filion, D.L., 2007. The electrodermal system. Handbook Psychophysiol. 2, 200–223.
Dunsmoor, J.E., Paz, R., 2015. Fear generalization and anxiety: behavioral and neural mechanisms. Biol. Psychiatry 78, 336–343.
Dunsmoor, J.E., Prince, S.E., Murty, V.P., Kragel, P.A., LaBar, K.S., 2011. Neurobehavioral mechanisms of human fear generalization. Neuroimage 55, 1878–1888.
Dymond, S., Dunsmoor, J.E., Vervliet, B., Roche, B., Hermans, D., 2015. Fear generalization in humans: systematic review and implications for anxiety disorder research. Behav. Ther. 46, 561–582.
Favilla, S., Huber, A., Pagnoni, G., Lui, F., Facchin, P., Cocchi, M., Baraldi, P., Porro, C.A., 2014. Ranking brain areas encoding the perceived level of pain from fMRI data. Neuroimage 90, 153–162.
Fechner, G., 1860. Elemente der Psychophysik. Breitkopf & Härtel. Holt, Rinehart & Winston, New York. Leipzig (reprinted in 1964 by Bonset, Amsterdam); English translation by HE Adler (1966): Elements of psychophysics.
Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774–781.
Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.
Fullana, M.A., Harrison, B.J., Soriano-Mas, C., Vervliet, B., Cardoner, N., Àvila-Parcet, A., Radua, J., 2016. Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. Mol. Psychiatr. 21, 500–508.
Gasquoine, P.G., 2014. Contributions of the insula to cognition and emotion. Neuropsychol. Rev. 24, 77–87.
Geuss, M.N., McCardell, M.J., Stefanucci, J.K., 2016. Fear similarly alters perceptual estimates of and actions over gaps. PLoS One 11, e0158610.
Ghirlanda, S., Enquist, M., 2003. A century of generalization. Anim. Behav. 66, 15–36.
Goulden, N., Khusnulina, A., Davis, N.J., Bracewell, R.M., Bokde, A.L., McNulty, J.P., Mullins, P.G., 2014. The salience network is responsible for switching between the default mode network and the central executive network: replication from DCM. Neuroimage 99, 180–190.
Greenberg, T., Carlson, J.M., Cha, J., Hajcak, G., Mujica-Parodi, L.R., 2013a. Neural reactivity tracks fear generalization gradients. Biol. Psychol. 92, 2–8.
Greenberg, T., Carlson, J.M., Cha, J., Hajcak, G., Mujica-Parodi, L.R., 2013b. Ventromedial prefrontal cortex reactivity is altered in generalized anxiety disorder during fear generalization. Depress. Anxiety 30, 242–250.
Guimaraes, F.S., Hellewell, J., Hensman, R., Wang, M., Deakin, J.F., 1991. Characterization of a psychophysiological model of classical fear conditioning in healthy volunteers: influence of gender, instruction, personality and placebo. Psychopharmacology (Berlin) 104, 231–236.
Guttman, N., Kalish, H.I., 1956. Discriminability and stimulus generalization. J. Exp. Psychol. 51, 79–88.
Harwell, M.R., Rubinstein, E.N., Hayes, W.S., Olds, C.C., 1992. Summarizing Monte Carlo results in methodological research: the one-and two-factor fixed effects ANOVA cases. J. Educ. Stat. 17, 315–339.
Hebart, M.N., Schriever, Y., Donner, T.H., Haynes, J.D., 2016. The relationship between perceptual decision variables and confidence in the human brain. Cerebr. Cortex 26, 118–130.
Ho, T.C., Brown, S., Serences, J.T., 2009. Domain general mechanisms of perceptual decision making in human cortex. J. Neurosci. 29, 8675–8687.
Holmes, A., Friston, K., 1998. Generalisability, random effects and population inference. In: Neuroimage: Abstracts of the Fourth International Conference on Functional Mapping of the Human Brain.
Holt, D.J., Boeke, E.A., Wolthusen, R.P., Nasr, S., Milad, M.R., Tootell, R.B., 2014. A parametric study of fear generalization to faces and non-face objects: relationship to discrimination thresholds. Front. Hum. Neurosci. 8, 624.
Holt, D.J., Lebron-Milad, K., Milad, M.R., Rauch, S.L., Pitman, R.K., Orr, S.P., Cassidy, B.S., Walsh, J.P., Goff, D.C., 2009. Extinction memory is impaired in schizophrenia. Biol. Psychiatry 65, 455–463.
Hwang, M.J., Zsido, R.G., Song, H., Pace-Schott, E.F., Miller, K.K., Lebron-Milad, K., Marin, M.-F., Milad, M.R., 2015. Contribution of estradiol levels and hormonal contraceptives to sex differences within the fear network during fear conditioning and extinction. BMC Psychiatry 15, 295.
Iglesias, J.E., Augustinack, J.C., Nguyen, K., Player, C.M., Player, A., Wright, M., Roy, N., Frosch, M.P., McKee, A.C., Wald, L.L., Fischl, B., Van Leemput, K., Initiative, A.s.D.N., 2015. A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. Neuroimage 115, 117–137.
Jürgens, U., 1984. The efferent and afferent connections of the supplementary motor area. Brain Res. 300, 63–81.
Kaczkurkin, A.N., Burton, P.C., Chazin, S.M., Manbeck, A.B., Espensen-Sturges, T., Cooper, S.E., Sponheim, S.R., Lissek, S., 2017. Neural substrates of overgeneralized conditioned fear in PTSD. Am. J. Psychiatry 174, 125–134.
Kong, J., White, N.S., Kwong, K.K., Vangel, M.G., Rosman, I.S., Gracely, R.H., Gollub, R.L., 2006. Using fMRI to dissociate sensory encoding from cognitive evaluation of heat pain intensity. Hum. Brain Mapp. 27, 715–721.
LaBar, K., LeDoux, J., Spencer, D., Phelps, E., 1995. Impaired fear conditioning following unilateral temporal lobectomy in humans. J. Neurosci. 15, 6846–6855.
Laufer, O., Paz, R., 2012. Monetary loss alters perceptual thresholds and compromises future decisions via amygdala and prefrontal networks. J. Neurosci. 32, 6304–6311.
Lebron-Milad, K., Abbs, B., Milad, M.R., Linnman, C., Rougemount-Bücking, A., Zeidan, M.A., Holt, D.J., Goldstein, J.M., 2012. Sex differences in the neurobiology of fear conditioning and extinction: a preliminary fMRI study of shared sex differences with stress-arousal circuitry. Biol. Mood Anxiety Disord. 2, 7-7.
Levinson, D.F., Edelberg, R., 1985. Scoring criteria for response latency and habituation in electrodermal research: a critique. Psychophysiology 22, 417–426.
Lissek, S., Bradford, D.E., Alvarez, R.P., Burton, P., Espensen-Sturges, T., Reynolds, R.C., Grillon, C., 2014a. Neural substrates of classically conditioned fear-generalization in humans: a parametric fMRI study. Soc. Cognit. Affect Neurosci. 9, 1134–1142.
Lissek, S., Kaczkurkin, A.N., Rabin, S., Geraci, M., Pine, D.S., Grillon, C., 2014b. Generalized anxiety disorder is associated with overgeneralization of classically conditioned fear. Biol. Psychiatry 75, 909–915.
Lissek, S., Rabin, S., Heller, R.E., Lukenbaugh, D., Geraci, M., Pine, D.S., Grillon, C., 2010. Overgeneralization of conditioned fear as a pathogenic marker of panic disorder. Am. J. Psychiatry 167, 47–55.
Mantini, D., Gerits, A., Nelissen, K., Durand, J.B., Joly, O., Simone, L., Sawamura, H., Wardak, C., Orban, G.A., Buckner, R.L., Vanduffel, W., 2011. Default mode of brain function in monkeys. J. Neurosci. 31, 12954–12962.
Menon, V., Uddin, L.Q., 2010. Saliency, switching, attention and control: a network model of insula function. Brain Struct. Funct. 214, 655–667.

Merz, C.J., Kinner, V.L., Wolf, O.T., 2018. Let's talk about sex … differences in human fear conditioning. Current Opinion in Behavioral Sciences 23, 7–12.
Milad, M.R., Goldstein, J.M., Orr, S.P., Wedig, M.M., Klibanski, A., Pitman, R.K., Rauch, S.L., 2006. Fear conditioning and extinction: influence of sex and menstrual cycle in healthy humans. Behav. Neurosci. 120, 1196–1203.
Milad, M.R., Orr, S.P., Pitman, R.K., Rauch, S.L., 2005. Context modulation of memory for fear extinction in humans. Psychophysiology 42, 456–464.
Nachev, P., Kennard, C., Husain, M., 2008. Functional role of the supplementary and pre-supplementary motor areas. Nat. Rev. Neurosci. 9, 856.
Onat, S., Büchel, C., 2015. The neuronal basis of fear generalization in humans. Nat. Neurosci. 18, 1811–1818.
Parkes, L., Lund, J., Angelucci, A., Solomon, J.A., Morgan, M., 2001. Compulsory averaging of crowded orientation signals in human vision. Nat. Neurosci. 4, 739.
Paul, E.J., Smith, J.D., Valentin, V.V., Turner, B.O., Barbey, A.K., Ashby, F.G., 2015. Neural networks underlying the metacognitive uncertainty response. Cortex 71, 306–322.
Pedersen, M.L., Endestad, T., Biele, G., 2015. Evidence accumulation and choice maintenance are dissociated in human perceptual decision making. PLoS One 10, e0140361.
Ploner, M., Lee, M.C., Wiech, K., Bingel, U., Tracey, I., 2010. Prestimulus functional connectivity determines pain perception in humans. Proc. Natl. Acad. Sci. U. S. A. 107, 355–360.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage 59, 2142–2154.
R Core Team, 2017. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L., 2001. A default mode of brain function. Proc. Natl. Acad. Sci. Unit. States Am. 98, 676–682.
Resnik, J., Sobel, N., Paz, R., 2011. Auditory aversive learning increases discrimination thresholds. Nat. Neurosci. 14, 791–796.
Seeley, W.W., Menon, V., Schatzberg, A.F., Keller, J., Glover, G.H., Kenna, H., Reiss, A.L., Greicius, M.D., 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J. Neurosci. 27, 2349–2356.
Seger, C.A., Braunlich, K., Wehe, H.S., Liu, Z., 2015. Generalization in category learning: the roles of representational and decisional uncertainty. J. Neurosci. 35, 8802–8812.
Schechtman, E., Laufer, O., Paz, R., 2010. Negative valence widens generalization of learning. J. Neurosci. 30, 10460–10464.
Shalev, L., Paz, R., Avidan, G., 2018. Visual aversive learning compromises sensory discrimination. J. Neurosci. 38, 2766–2779.
Sidlauskaite, J., Wiersema, J.R., Roeyers, H., Krebs, R.M., Vassena, E., Fias, W., Brass, M., Achten, E., Sonuga-Barke, E., 2014. Anticipatory processes in brain state switching - evidence from a novel cued-switching task implicating default mode and salience networks. Neuroimage 98, 359–365.
Skrandies, W., Fahle, M., 1994. Neurophysiological correlates of perceptual learning in the human brain. Brain Topogr. 7, 163–168.
Snellen, H., 1862. Probebuchstaben zur Bestimmung der Sehschärfe, Utrecht.
Sridharan, D., Levitin, D.J., Menon, V., 2008. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. Proc. Natl. Acad. Sci. U. S. A. 105, 12569–12574.
Staib, M., Bach, D.R., 2018. Stimulus-invariant auditory cortex threat encoding during fear conditioning with simple and complex sounds. Neuroimage 166, 276–284.
Stefanucci, J.K., Storbeck, J., 2009. Don't look down: emotional arousal elevates height perception. J. Exp. Psychol. Gen. 138, 131–145.
Sterzer, P., Kleinschmidt, A., 2010. Anterior insula activations in perceptual paradigms: often observed but barely understood. Brain Struct. Funct. 214, 611–622.
Struyf, D., Zaman, J., Vervliet, B., Van Diest, I., 2015. Perceptual discrimination in fear generalization: mechanistic and clinical implications. Neurosci. Biobehav. Rev. 59, 201–207.
Teachman, B.A., Stefanucci, J.K., Clerkin, E.M., Cody, M.W., Proffitt, D.R., 2008. A new mode of fear expression: perceptual bias in height fear. Emotion 8, 296–301.
Tinoco-González, D., Fullana, M.A., Torrents-Rodas, D., Bonillo, A., Vervliet, B., Blasco, M.J., Farré, M., Torrubia, R., 2015. Conditioned fear acquisition and generalization in generalized anxiety disorder. Behav. Ther. 46, 627–639.
Turner, S.M., Beidel, D.C., Roberson-Nay, R., 2005. Offspring of anxious parents: reactivity, habituation, and anxiety-proneness. Behav. Res. Ther. 43, 1263–1279.
Visser, R.M., Scholte, H.S., Kindt, M., 2011. Associative learning increases trial-by-trial similarity of BOLD-MRI patterns. J. Neurosci. 31, 12021–12028.
Vogt, B.A., 2016. Midcingulate cortex: structure, connections, homologies, functions and diseases. J. Chem. Neuroanat. 74, 28–46.
Wiech, K., Lin, C.S., Brodersen, K.H., Bingel, U., Ploner, M., Tracey, I., 2010. Anterior insula integrates information about salience into perceptual decisions about pain. J. Neurosci. 30, 16324–16331.
Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665.
Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 1125–1165.
Zorawski, M., Cook, C.A., Kuhn, C.M., LaBar, K.S., 2005. Sex, stress, and fear: individual differences in conditioned learning. Cognit. Affect Behav. Neurosci. 5, 191–201.