

# Active Avoidance: Neural Mechanisms and Attenuation of Pavlovian Conditioned Responding

Emily A. Boeke,<sup>1</sup> Justin M. Moscarello,<sup>2</sup>  Joseph E. LeDoux,<sup>2,3</sup> Elizabeth A. Phelps,<sup>1,2,3</sup> and Catherine A. Hartley<sup>1</sup>

<sup>1</sup>Department of Psychology and <sup>2</sup>Center for Neural Science, New York University, New York, New York 10003, and <sup>3</sup>Nathan Kline Institute, Orangeburg, New York 10962

Patients with anxiety disorders often experience a relapse in symptoms after exposure therapy. Similarly, threat responses acquired during Pavlovian threat conditioning often return after extinction learning. Accordingly, there is a need for alternative methods to persistently reduce threat responding. Studies in rodents have suggested that exercising behavioral control over an aversive stimulus can persistently diminish threat responses, and that these effects are mediated by the amygdala, ventromedial prefrontal cortex, and striatum. In this fMRI study, we attempted to translate these findings to humans. Subjects first underwent threat conditioning. We then contrasted two forms of safety learning: active avoidance, in which participants could prevent the shock through an action, and yoked extinction, with shock presentation matched to the active condition, but without instrumental control. The following day, we assessed subjects' threat responses (measured by skin conductance) to the conditioned stimuli without shock. Subjects next underwent threat conditioning with novel stimuli. Yoked extinction subjects showed an increase in conditioned response to stimuli from the previous day, but the active avoidance group did not. Additionally, active avoidance subjects showed reduced conditioned responding during novel threat conditioning, but the extinction group did not. We observed between-group differences in striatal BOLD responses to shock omission in Avoidance/Extinction. These findings suggest a differential role for the striatum in human active avoidance versus extinction learning, and indicate that active avoidance may be more effective than extinction in persistently diminishing threat responses.

**Key words:** active avoidance; anxiety; coping; instrumental learning; resilience; threat conditioning

## Significance Statement

Extinguished threat responses often reemerge with time, highlighting the importance of identifying more enduring means of attenuation. We compared the effects of active avoidance learning and yoked extinction on threat responses in humans and contrasted the neural circuitry engaged by these two processes. Subjects who learned to prevent a shock through an action maintained low threat responses after safety learning and showed attenuated threat conditioning with novel stimuli, in contrast to those who underwent yoked extinction. The results suggest that experiences of active control over threat engage the striatum and promote a shift from expression of innate defensive responses toward more adaptive behavioral responses to threatening stimuli.

## Introduction

The study of Pavlovian threat conditioning and extinction has been useful for understanding and treating clinical anxiety. In threat conditioning, when a neutral cue is paired with an aversive stimulus (such as shock), the subject typically develops a condi-

tioned response, such as freezing or sweating, to the predictive cue. Threat conditioning provides a model for how threat associations are acquired in anxiety disorders. In extinction, the cue is no longer followed by the aversive outcome and conditioned responding diminishes. Exposure therapy is a common treatment for anxiety based on the principles of extinction. However, extinguished threat responses often return with the mere passage of time (“spontaneous recovery”) or following a shift in context (“renewal”). Similarly, patients' fears often reemerge after exposure therapy, and safety learned in therapy does not always reduce anxiety outside the therapeutic context. The temporary nature of extinction learning has motivated research on alternative techniques for reducing threat responses.

Studies in rodents suggest that exercising behavioral control over aversive stimuli can persistently diminish learned threat re-

Received Oct. 20, 2016; revised March 24, 2017; accepted March 27, 2017.

Author contributions: E.A.B., J.M.M., J.E.L., E.A.P., and C.A.H. designed research; E.A.B. performed research; E.A.B. analyzed data; E.A.B., J.M.M., J.E.L., E.A.P., and C.A.H. wrote the paper.

This work was supported by National Institutes of Health Grant TR01MH097085 to E.A.P. We thank Karolina Lempert, Alexa Tompany, Jennifer Lenow, and Elizabeth Goldfarb for helpful discussions and suggestions.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Catherine A. Hartley, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. E-mail: cah369@nyu.edu.

DOI:10.1523/JNEUROSCI.3261-16.2017

Copyright © 2017 the authors 0270-6474/17/374808-11\$15.00/0

sponses (Cain and LeDoux, 2007; Maier, 2015; LeDoux et al., 2017). These paradigms can be considered a model of “proactive coping,” where individuals exert control to change an aversive situation. The results of these studies suggest that safety learning that incorporates proactive coping could provide a promising alternative to standard extinction. In these studies, a group of rats can learn an action (such as shuttling to another chamber) to terminate either a cue predicting shock or the shock itself. This group is compared with rodents that experience yoked stimulus exposure, but without instrumental control. One or more days later, the rodents that experienced control show less freezing to previous or current threat predictive cues, even when the avoidance action is not possible or no longer instrumentally effective (Cain and LeDoux 2007; Baratta et al., 2007). These studies are notable in that they suggest that control over aversive stimuli may have the potential to promote resilience to previously learned threats, or to threatening stimuli that have not yet been encountered.

Instrumental control over threat has been shown to involve the amygdala, striatum, and ventromedial prefrontal cortex (vmPFC) in rodents. Cue-shock associations are thought to be established in the lateral nucleus of the amygdala and subsequently transmitted to the central nucleus, which drives the expression of Pavlovian threat responses, such as freezing, through descending brainstem projections (LeDoux, 2000). The lateral nucleus also projects to the basal nucleus of the amygdala, which projects to the striatum. These connections are essential for avoidance learning (Amorapanth et al., 2000; Ramirez et al., 2015). vmPFC activation during avoidance/escape learning, as well as subsequent Pavlovian tests, appears to mediate the attenuation of conditioned threat responses, possibly via inhibiting function of the amygdala central nucleus (Baratta et al., 2007, 2008; Moscarello and LeDoux, 2013). The vmPFC is proposed to toggle between these different output paths of the amygdala, shifting the flow of information to promote aversively motivated action instead of Pavlovian freezing responses (Moscarello and LeDoux, 2013; Ramirez et al., 2015).

Translation of this research to humans has been limited (Collins et al., 2014; Hartley et al., 2014). Here, we contrasted the effects of active avoidance versus yoked extinction on subsequent threat responses, assessed via skin conductance responses (SCR), and examined their respective neural substrates using fMRI. Our study had four phases: Acquisition, Avoidance or Extinction, Retrieval, and Novel Acquisition. We predicted that active avoidance subjects (“Masters”) would show reduced spontaneous recovery and reduced threat conditioning with new stimuli, compared with subjects who underwent yoked passive extinction (“Yokes”). Based on the neural model stemming from rodent studies, we hypothesized that Master subjects would show greater conditioning-related BOLD signal in vmPFC and striatum and less BOLD signal in amygdala during Avoidance learning relative to Yoke subjects during Extinction, and that these neural differences would persist during tests of spontaneous recovery (during Retrieval) and Novel Acquisition the following day.

## Materials and Methods

**Subjects.** Fifty-six subjects (35 female, see Inclusions/exclusions), with an average age of 23 ( $\pm 0.5$ ) years, were included in analyses. Subjects were randomly assigned to be Master subjects ( $N = 28$ ) or Yoke subjects ( $N = 28$ ).

**Procedure overview.** The experiment consisted of 4 phases across 2 d (sessions were scheduled 23–25 h apart). Each Yoke subject was paired to a Master subject. Throughout all phases, Yoke subjects experienced the

same stimulus counterbalancing as their respective Master subject. On day 1, subjects underwent the Acquisition and Avoidance/Extinction phases. In Avoidance/Extinction, subjects in the Master group learned to avoid the shock by performing an action (correctly changing the location of a circle in a grid), and Yoke subjects underwent yoked extinction while performing a motor control task. On day 2, subjects underwent the Retrieval and Novel Acquisition phases. All phases were separated by 5–6 min breaks in which resting state or structural data were collected (analyses not presented here). Each experimental phase started and ended with 20 s fixation. Skin conductance was recorded in all phases, and stimuli were presented in E-Prime 2.0.

**Stimuli and procedure.** Written informed consent was obtained from all subjects before the start of the study in accordance with the policies of New York University’s Committee on Activities Involving Human Subjects.

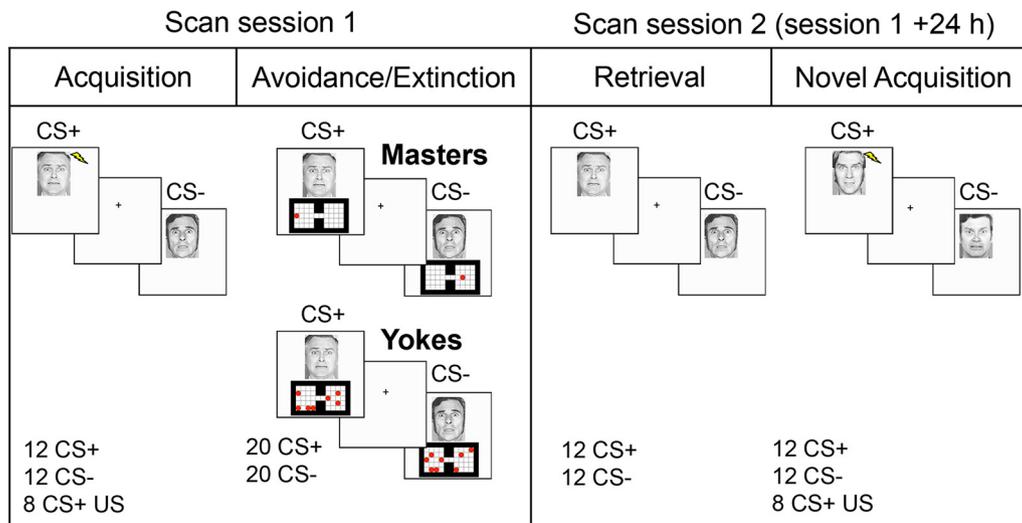
Master subjects were given a chance to practice moving a circle around in the grid, to ensure that they knew how to use the buttons to manipulate the position of the circle before entering the MRI scanner. However, they were not told anything about avoiding shocks before entering the scanner.

The level of the shock, which served as an unconditioned stimulus (US), was adjusted before the subject entered the scanner. The level was set to be “uncomfortable, but not painful,” by first testing the shock at the lowest level and increasing the level with the subject’s permission until an appropriate level was reached. The shock (200 ms, 50 pulses/s) was produced by a stimulator (Grass Instruments) via electrodes on the subject’s right wrist. After the Acquisition scan, 5 subjects asked for the shock level to be decreased, and the level was adjusted. We also checked the shock level immediately before Novel Acquisition (after the test of spontaneous recovery during Retrieval phase) to recalibrate it if necessary (recalibration occurred for 7 subjects). Fisher’s exact test showed no difference in the proportion of subjects requiring recalibration at either time point across Master and Yoke subjects ( $p$  values = 1).

During acquisition, subjects viewed conditioned stimuli (CSs;  $CS^+$  predicts shock on 40% of trials and  $CS^-$  predicts no shock), which consisted of two pairs of black and white images of faces with fearful expressions from the Ekman pictures of facial affect (Ekman and Friesen, 1976) on a white background (Fig. 1). Subjects were instructed that one face would sometimes be followed by a shock (the “threat face”) while another would not (the “no threat face”), but they were not told which face was the “threat face.” Within each pair, the identity of the  $CS^+$  was counterbalanced across subjects. Each subject saw one pair of faces in the first 3 phases of the experiment and one pair in the last phase of the experiment. Pair assignment was counterbalanced across subjects. During Acquisition, subjects saw 8  $CS^+$  stimuli that coterminated with a shock, 12  $CS^+$  stimuli unpaired with shock, and 12  $CS^-$  stimuli in a fixed pseudorandom order. CSs were presented for 6 s, with an intertrial interval (ITI) of 8–12 s. During ITIs, subjects viewed a white screen with a fixation cross.

Before Avoidance/Extinction, subjects were told that they would again see the faces from the previous scan, but this time, a grid would be present below the faces. Master subjects were told that they would see a circle in the grid, and that they would be able to manipulate the position of the circle within the grid with the arrow keys on the keyboard. They were told that performing a certain action with the circle in the grid would allow them to avoid the shock, and that they should perform this action whenever the threat face was on the screen. They were not told what the specific action was that would prevent the shock. Master subjects were additionally told to press a button ~8 times (the expected average number of button presses during  $CS^+$  trials) when the  $CS^-$  was presented, to equate for button pressing during the  $CS^+$  trials. They were again allowed to practice moving the circle in the grid before they began the scan. Yoke subjects were told that, on a given trial, they may see a number of red circles and that they should press a button once for every circle that appeared on that trial. They were also informed that their actions would have no influence on the stimuli.

In the Avoidance/Extinction phase, subjects saw 20  $CS^+$  and 20  $CS^-$  faces, lasting 6 s each with a 10–14 s ITI, in a fixed pseudorandom order (Fig. 1). In this phase, a grid appeared below the CSs, and it was present



**Figure 1.** Design. In the Acquisition phase, subjects saw a CS<sup>+</sup> face that sometimes coterminated with shock, and a CS<sup>-</sup> face that was not paired with shock. In Avoidance/Extinction, Master subjects were able to prevent the shock on CS<sup>+</sup> trials by moving the circle to the opposite “chamber” from which it started. Yoke subjects each were paired to a Master subject. Yoke subjects received identical CS and shock exposure as their Master subjects, and they performed a motor control task. On the subsequent day, subjects saw the CSs in the Retrieval phase, but they were not shocked. Novel Acquisition was identical to initial Acquisition, but subjects saw a new pair of CS<sup>+</sup> and CS<sup>-</sup> stimuli.

continuously, including during ITIs. The grid consisted of 2 “chambers” composed of 20 cells each, connected by a 2-cell “tunnel” (Fig. 1). During CS<sup>+</sup> trials, Master subjects could move the circle by one cell per press of an arrow on the keyboard. On a given CS<sup>+</sup> trial, if the subjects moved the circle to the opposite “chamber” from where the circle had started, they were not shocked at the end of the trial. If they failed to perform this action, they were shocked immediately after the 6 s CS<sup>+</sup> presentation. Yoke subjects saw the same grid but saw multiple (static) red circles in the grid on each trial. The number of red circles was matched to the number of button presses made by their respective Master subjects on that trial. Yokes were asked to press a button once for every circle present, and thus they were matched to Masters on each trial for number of button presses. Yoke subjects were shocked on trials when their Master subjects were shocked. Since Master subjects typically learned quickly, this phase was considered an extinction phase for Yokes. At the end of day 1, subjects were asked to rate the shock on a scale from 1 to 10, with 10 being “extremely aversive” and 1 being “not at all aversive.”

On day 2, subjects were told that the shock would be set at the same level as previously. Before the Retrieval phase, subjects were instructed that they would see the same faces, but that there would be no grid and they would be making no responses. We were interested in the initial response to the CS<sup>+</sup> during Retrieval; however, subjects often show a nonspecific orienting response to the first stimulus presented in a new phase. Thus, the Retrieval phase started with one CS<sup>-</sup> presentation to capture the orienting response, which was not included in analysis. This CS<sup>-</sup> was followed by 12 CS<sup>+</sup> and 12 CS<sup>-</sup> stimuli, in a fixed pseudorandom order, with 6 s presentation time, and a 8–12 s ITI. No shocks were administered in this phase. The Novel Acquisition phase was the same as the Acquisition phase, but with a new pair of stimuli for the CS<sup>+</sup> and CS<sup>-</sup>.

After the day 2 scan, subjects filled out the Internal Control Index (ICI) (Duttweiler, 1984), Spielberger State and Trait Anxiety Index (STAI) (Spielberger, 1993), the Intolerance of Uncertainty Scale (IUS) (Freeston et al., 1994), the Perceived Stress Scale (PSS) (Cohen et al., 1983), and a brief questionnaire about the experiment, including a question in which subjects were asked to rate the extent to which they were able to control the onset of the shock (from 1, “not at all,” to 5, “completely”) during the Avoidance/Extinction phase.

**Skin conductance acquisition and analysis.** Skin conductance data were acquired at a 200 Hz sampling rate and amplified with a Biopac Systems skin conductance module. Disposable Ag/AgCl Electrodes were attached to the subject’s left palm. Before analysis (with AcqKnowledge software, RRID: SCR\_014279), the data were filtered with a 1 Hz low pass FIR

filter. The SCR for each trial was computed as the trough to peak amplitude for the first deflection with a latency of 0.5–6.5 s after CS onset. SCRs < 0.02  $\mu$ S were scored as 0. For the Acquisition phases, CS<sup>+</sup> trials followed by US were not included in analysis. Data were square root transformed and range corrected by dividing by the mean square root transformed US response during initial Acquisition.

SCR values from all four phases were analyzed in a 2 (CS type)  $\times$  2 (group)  $\times$  4 (phase) repeated-measures ANOVA. Where Mauchly’s test indicated that sphericity assumptions were violated, Greenhouse-Geisser  $\epsilon$  correction was applied. Significant interactions were investigated with *post hoc* ANOVAs and two-tailed *t* tests.

**Inclusions/exclusions.** All subjects enrolled ( $N = 97$ ) reported no past or present psychiatric or neurological disorders and were not taking psychotropic medications. Subjects were age 18–35 years, right-handed, fluent in English, had normal or corrected vision, were not pregnant, had no contraindications for MRI, and had not participated in another study involving shock in at least 6 months. Subjects with <4 skin conductance responses to unreinforced CS<sup>+</sup> or CS<sup>-</sup> trials during Acquisition were considered nonresponders and did not participate in the day 2 scan ( $N = 17$ ). Nine additional subjects who did not acquire a conditioned response (average CS<sup>+</sup> SCR > CS<sup>-</sup> SCR during Acquisition) were not invited to participate on day 2. Of the remaining subjects, 5 Master subjects, who received at least one shock during the last 5 CS<sup>+</sup> trials of the scan, were considered not to have learned the avoidance response and were not invited to participate on day 2. Eight subjects were excluded for other reasons (illness on day 2,  $N = 1$ ; equipment problems,  $N = 2$ ; incomplete SCR data,  $N = 1$ ; noncompliance,  $N = 1$ ; requesting to stop mid-scan,  $N = 2$ ; and inability to identify the CS<sup>+</sup> after Acquisition,  $N = 1$ ). Two of the remaining subjects were excluded for excessive head motion (>2.5 mm) in the Avoidance/Extinction scan. Two Masters (and their paired Yokes) were excluded from fMRI analyses of the Novel Acquisition scan only (one for head movement, one for stimulus presentation timing problems).

**MRI acquisition.** Subjects were scanned on a 3T Siemens Allegra head-only scanner and a Siemens standard head coil. We began both days by collecting an auto-align scout and a brief structural scan for positioning slices for functional scans. Functional scans (TR = 2000 ms; effective TE = 30 ms; flip angle = 82°, 34 contiguous 3 mm slices, matrix = 80  $\times$  64; FOV = 240  $\times$  192 mm; acquisition voxel size = 3  $\times$  3  $\times$  3 mm; 271 volumes in Acquisition/Novel Acquisition, 374 volumes in Avoidance/Extinction, 215 volumes in Retrieval) were acquired using a customized multiecho EPI sequence to mitigate the effects of susceptibility artifacts near vmPFC, described in detail previously (Hackel et al., 2015). For all

functional scans, slices were aligned to ACPC and then tilted up by 10 degrees to maximize signal recovery near vmPFC and positioned so that the medial temporal lobe was covered (in some subjects, the superior frontal and parietal lobes were not completely covered). To aid in both the estimation of the parameter maps for the functional data and the alignment between the functional and high-resolution anatomy, we collected a calibration scan (multiecho GRE) before the first functional run. This calibration scan had the same parameters and slice prescription as the functional runs. Day 1 ended with a T1 MPRAGE scan (176 sagittal slices, 1 mm isotropic voxels, TR = 2500 ms, TE 3.93 ms, flip angle 8°). On day 2, an additional T1 scan and a DTI scan were collected (not analyzed for this manuscript).

**MRI data analysis.** The data were preprocessed and analyzed with FMRIB Software Library (FSL, RRID: SCR\_002823) (Smith et al., 2004). The first 4 time points of each BOLD run were removed to allow for T1 equilibration effects. Data were motion corrected, smoothed with a Gaussian kernel at 5 mm FWHM, and high pass filtered (120 s). Transformation from native space to subjects' anatomical space (using the skull-stripped T1 image) and transformation from anatomical space to MNI 152 template space ( $2 \times 2 \times 2$  voxel size) were computed using FMRIB's Linear Image Registration Tool. Data remained in native space for first level analyses.

Subjects' data were analyzed with a GLM. Separate models were implemented for each experimental phase. Analyses included regressors for 6 motion parameters. Boxcar functions (6 s) were used to model CS<sup>+</sup> and CS<sup>-</sup> cues. All analyses of BOLD signal during CS presentation included regressors for the unreinforced CS<sup>+</sup> and CS<sup>-</sup>, as well as regressors for the reinforced CS<sup>+</sup> and for the US (stick function regressor) if applicable. Task-related regressors were convolved with a double-gamma function, and their temporal derivatives were included in the model.

Analyses of the Avoidance/Extinction phase included two main analyses and two exploratory analyses. As there was variation in the number of trials in which Master subjects learned the avoidance response, one analysis focused on CS<sup>+</sup> versus CS<sup>-</sup> activation during CS presentation in late Avoidance/Extinction (last 10 trials), after all Master subjects had learned the avoidance task and subjects were no longer being shocked (one Master/Yoke pair that received a shock during this period was not included in analyses of late Avoidance/Extinction). A second analysis examined CS<sup>+</sup> versus CS<sup>-</sup> activation during CS presentation, across all trials of Avoidance/Extinction. As avoidance learning is driven by the absence of reinforcement, we also examined neural responses during the time point of the offset of the CS, when the subject would have received the shock if it had not been avoided. These two analyses examined the last 10 trials and all trials of Avoidance/Extinction, respectively, but CS<sup>+</sup> and CS<sup>-</sup> regressors were stick functions marking the offset of the CS. CS offset analyses included a boxcar regressor for all CSs, as well as a regressor for the US.

In all analyses, the contrasts of interest were nonreinforced CS<sup>+</sup> > CS<sup>-</sup> (or CS<sup>+</sup> > CS<sup>-</sup> offset) and the reverse contrast. Data were analyzed at the group level using randomize, FSL tool for nonparametric permutation inference on neuroimaging data (Winkler et al., 2014). Data were corrected for multiple comparisons (controlling family-wise error [FWE]) with an extent-based cluster correction via randomize, with a cluster-forming *t*-stat threshold of 3.0 and a cluster-level  $\alpha$  of 0.05. For reference, all tables additionally include voxelwise FWE corrected *p* values (computed by randomize) of the peak voxel of each cluster that survives correction. All permutation tests consisted of 5000 iterations. For the Acquisition phase, a single within-group map (one-sample, two-tailed nonparametric *t* test) was produced, collapsing across Master and Yoke subjects, correcting for multiple comparisons across the whole brain. Given our strong a priori hypotheses, for other phases, we performed a small volume correction analysis in the vmPFC, striatum, and amygdala. Master and Yoke subjects' masked maps were compared with a two-sample, one-tailed nonparametric *t* test (as our hypotheses were directional), correcting for multiple comparisons within our regions of interest. Within-group maps were also examined. The region of interest mask was constructed from the Harvard-Oxford Atlas, using the combined bilateral amygdala, caudate, putamen, and vmPFC (accumbens,

subcallosal cortex, frontal medial cortex, anterior cingulate cortex, and paracingulate cortex), thresholded at 25% probability (to allow for overlap between subregions of vmPFC). The anterior cingulate cortex and paracingulate cortex masks were truncated above  $z = 4$ , at the level of the genu of the corpus callosum. Given that we had strong a priori hypotheses about vmPFC and amygdala, but no between-group differences survived small volume corrections in these regions, we also report any uncorrected between-group differences in vmPFC and amygdala (one-tailed nonparametric *t* test,  $p \leq 0.001$ , 10 voxel extent threshold), as well as the within-group clusters that drive the differences.

## Results

### Behavior

As would be expected since groups were randomly assigned, groups did not differ in age (mean  $\pm$  SEM: Masters:  $23.21 \pm 0.75$ , Yokes:  $22.79 \pm 0.68$ ,  $t_{(54)} = 0.424$ ,  $p = 0.67$ ), gender (Masters: 16 female, Yokes: 19 females,  $\chi_{(1)} = 0.686$ ,  $p = 0.41$ ), or questionnaire scores (Masters STAI state:  $34.93 \pm 1.8$ , Yokes:  $34.18 \pm 1.69$ ,  $t_{(54)} = 0.3$ ,  $p = 0.76$ ; Masters STAI trait:  $38.82 \pm 1.3$ , Yokes:  $39.54 \pm 1.57$ ,  $t_{(54)} = -0.35$ ,  $p = 0.73$ ; Masters IUS:  $63.75 \pm 2.88$ , Yokes:  $64.89 \pm 2.96$ ,  $t_{(54)} = -0.28$ ,  $p = 0.78$ ; Masters ICI:  $104.25 \pm 1.95$ , Yokes:  $105.32 \pm 2.09$ ,  $t_{(54)} = -0.38$ ,  $p = 0.71$ ; Masters PSS:  $14.37 \pm 1.01$ , Yokes:  $16.43 \pm 1.23$ ,  $t_{(53)} = -1.28$ ,  $p = 0.21$ ).

Neither Master nor Yoke subjects showed a difference in mean number of button presses on CS<sup>+</sup> versus CS<sup>-</sup> trials (Masters CS<sup>+</sup>:  $10.22 \pm 1.07$ , Masters CS<sup>-</sup>:  $8.56 \pm 0.16$ , Yokes CS<sup>+</sup>:  $8.68 \pm 0.81$ , Yokes CS<sup>-</sup>:  $8.34 \pm 0.22$ ,  $p$  values > 0.13), and Master subjects did not differ from Yoke subjects in mean number of button presses during either CS<sup>+</sup> or CS<sup>-</sup> trials ( $p$  values > 0.26). Master and Yoke subjects did not differ in mean latency to the first response on each trial (Masters:  $0.86 \pm 0.03$  s, Yokes:  $0.87 \pm 0.04$  s,  $t_{(52)} = -0.28$ ,  $p = 0.78$ ). There was no difference in the shock level (Masters:  $35.7 \pm 2$  V, Yokes:  $39.3 \pm 2.1$  V,  $t_{(54)} = -1.22$ ,  $p = 0.23$ ) between Master and Yoke subjects. Subjects also rated how aversive the shock was, and Masters and Yokes did not differ in these ratings (Masters:  $6.43 \pm 0.28$ , Yokes:  $6.96 \pm 0.25$ ,  $t_{(54)} = -1.43$ ,  $p = 0.16$ ).

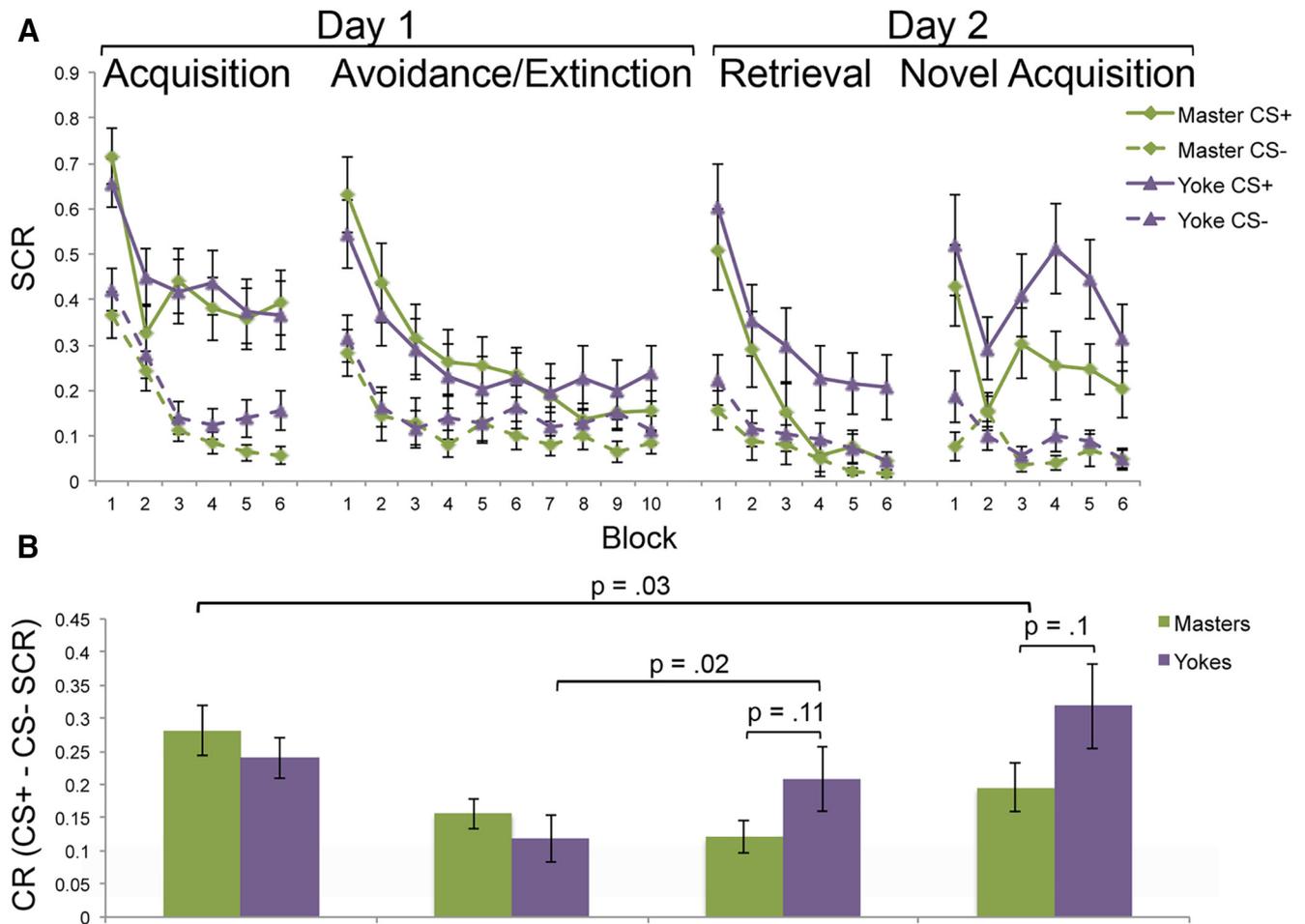
One Yoke subject pressed an incorrect response button and his responses were not recorded, but he reported making the correct number of presses. Another Yoke subject misunderstood the instructions and made approximately half of the correct number of presses on each trial, but this subject's number of presses per trial were still highly correlated with his Master subject's presses (Spearman's  $\rho = 0.77$ ,  $p = 7.9 \times 10^{-9}$ ).

Nine Master subjects received 1–7 shocks during the Avoidance scan. The other 19 Master subjects discovered the correct response on the first trial and continued to avoid being shocked for the rest of the scan.

After the experiment was over, Master subjects rated significantly higher on a scale asking about their subjective sense of control over the shock during the Avoidance/Extinction scan (Masters:  $4.5 \pm 0.16$ , Yokes:  $1.42 \pm 0.17$ ,  $t_{(54)} = 14.27$ ,  $p = 5.8 \times 10^{-20}$ ).

### Skin conductance

Skin conductance throughout the experiment, in blocks of two trials, can be seen in Figure 2A. Conditioned responses during each phase of the experiment are pictured in Figure 2B. The  $2 \times 2 \times 4$  ANOVA revealed main effects of CS type ( $F_{(1,54)} = 98.65$ ,  $p = 8.7 \times 10^{-14}$ ) and phase ( $F_{(2.57,138.92)} = 13.71$ ,  $p = 3.6 \times 10^{-7}$ ), which were qualified by several interactions: CS type  $\times$  phase ( $F_{(2.41, 130.21)} = 8.28$ ,  $p = 0.0002$ ) and CS type  $\times$  group  $\times$  phase ( $F_{(2.41,130.21)} = 3.726$ ,  $p = 0.02$ ).

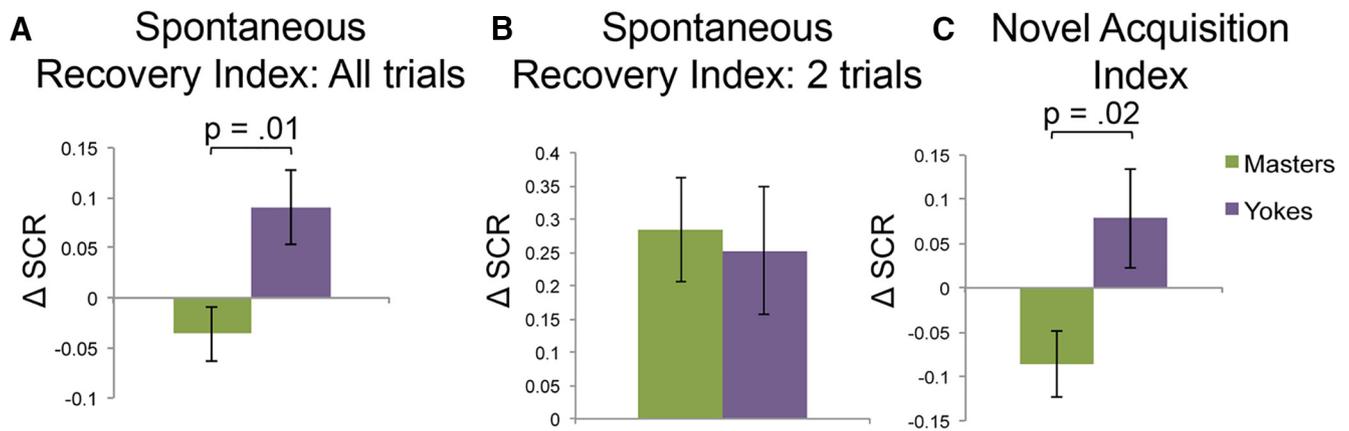


**Figure 2.** Skin conductance results. **A**, Skin conductance data are plotted in two-trial blocks. **B**, Average conditioned responses ( $CS^+ - CS^-$  skin conductance responses) during each phase of the experiment. CR, conditioned response. Error bars show SEM.

As we hypothesized that Master subjects would show reduced spontaneous recovery compared with Yoke subjects, we expected a between-group difference in change in conditioned response from Avoidance/Extinction to Retrieval. Thus, we conducted a  $2 \times 2 \times 2$  *post hoc* ANOVA only including the Avoidance/Extinction and Retrieval phase. In addition to a main effect of CS type ( $F_{(1,54)} = 48.8, p = 4 \times 10^{-9}$ ), a trend-level effect of phase ( $F_{(1,54)} = 3.12, p = 0.08$ ), and a trend-level phase  $\times$  group interaction ( $F_{(1,54)} = 3.08, p = 0.09$ ), we observed a significant three-way interaction ( $F_{(1,54)} = 7.59, p = 0.01$ ), suggesting that the groups differed in how their conditioned responses changed between these two phases. This interaction can be visualized as the Spontaneous Recovery Index, or change in conditioned response (average  $CS^+$  SCR -  $CS^-$  SCR) from Avoidance or Extinction to Retrieval, which is plotted in Figure 3A. Yoke subjects show an increase in conditioned response from safety learning to Retrieval ( $t_{(27)} = -2.45, p = 0.02$ ), whereas Masters show no significant change in conditioned response ( $t_{(27)} = 1.32, p = 0.2$ ). To determine which phase was driving the interaction with group, we conducted within-phase  $2 \times 2$  ANOVAs. In Avoidance/Extinction, we found a main effect of CS type ( $F_{(1,54)} = 41.44, p = 3 \times 10^{-8}$ ) but no interaction with group ( $F_{(1,54)} = 0.79, p = 0.38$ ). In Retrieval, we found a main effect of CS type ( $F_{(1,54)} = 36.54, p = 1 \times 10^{-7}$ ), a trend-level effect of group ( $F_{(1,54)} = 2.23, p = 0.14$ ), and a trend-level interaction between CS type and group ( $F_{(1,54)} = 2.64, p = 0.11$ ). During Retrieval, Masters showed lowered  $CS^+$

responses than Yokes at a trend level ( $t_{(54)} = -1.62, p = 0.11$ ). The  $CS^-$  responses were not significantly different between groups ( $t_{(54)} = -1.02, p = 0.31$ ). As spontaneous recovery is often limited to the earliest stage of retrieval, we also conducted a  $2 \times 2 \times 2$  ANOVA examining only the last 2-trial block of Avoidance/Extinction and the first 2-trial block of Retrieval. We observed a main effect of CS type ( $F_{(1,54)} = 59.54, p = 3 \times 10^{-10}$ ) and phase ( $F_{(1,54)} = 31.51, p = 7 \times 10^{-7}$ ), as well as a CS type  $\times$  phase interaction ( $F_{(1,54)} = 18.77, p = 0.00007$ ), but no three-way interaction ( $F_{(1,54)} = 0.06, p = 0.8$ ), depicted in Figure 3B by the difference in mean conditioned response between these two blocks. These tests suggest that, although both groups show evidence of spontaneous recovery in early Retrieval, Masters show reduced spontaneous recovery compared with Yokes across the entire Retrieval phase, and that the between-group difference in change in conditioned response is predominantly driven by  $CS^+$  responses during Retrieval.

We also expected a greater decrement in conditioned responding from Acquisition to Novel Acquisition in Masters compared with Yokes, so we conducted a  $2 \times 2 \times 2$  ANOVA only including the Acquisition and Novel Acquisition phases. Here, we found main effects of CS type ( $F_{(1,54)} = 95.53, p = 2 \times 10^{-13}$ ) and phase ( $F_{(1,54)} = 19.9, p = 0.00004$ ) as well as a significant three-way interaction ( $F_{(1,54)} = 6.02, p = 0.02$ ). The three-way interaction can be visualized as the Novel Acquisition index, or change in average  $CS^+$  minus  $CS^-$  response from Acquisition to Novel



**Figure 3.** Spontaneous Recovery Indices and Novel Acquisition Index. **A**, Spontaneous Recovery Index, as measured by the average conditioned response across all Retrieval trials minus the average conditioned response across all Avoidance/Extinction trials. The two-sample *t* test *p* value is equivalent to that of the CS type  $\times$  group  $\times$  phase (Avoidance/Extinction vs Retrieval) ANOVA interaction. **B**, Spontaneous Recovery Index, as measured by the average conditioned response from the first two trials of Retrieval minus the average conditioned response from the last two trials of Avoidance/Extinction. **C**, Novel Acquisition Index, the conditioned response during Novel Acquisition minus the conditioned response during Initial Acquisition. The two-sample *t* test *p* value is equivalent to that of the CS type  $\times$  group  $\times$  phase (Acquisition vs Novel Acquisition) ANOVA interaction. Error bars show SEM.

Acquisition, which is plotted in Figure 3C. Masters show a decrease in conditioned response from Acquisition to Novel Acquisition ( $t_{(27)} = 2.31, p = 0.03$ ), but Yokes show no change in conditioned response ( $t_{(27)} = -1.4, p = 0.17$ ). We again further probed the interaction with within-phase ANOVAs in Acquisition and Novel Acquisition. In Acquisition, there was a main effect of CS type ( $F_{(1,54)} = 116.64, p = 4 \times 10^{-15}$ ) but no interaction with group ( $F_{(1,54)} = 0.721, p = 0.4$ ). In Novel Acquisition, there was a significant main effect of CS type ( $F_{(1,54)} = 47.8, p = 6 \times 10^{-9}$ ), trend effect of group ( $F_{(1,54)} = 2.68, p = 0.11$ ), and a trend-level interaction between CS type and group ( $F_{(1,54)} = 2.76, p = 0.1$ ). During Novel Acquisition, Masters showed decreased CS<sup>+</sup> responses compared with Yokes, at a trend level ( $t_{(54)} = -1.71, p = 0.09$ ), and comparable CS<sup>-</sup> responses ( $t_{(54)} = -0.91, p = 0.37$ ). Thus, Master subjects show an attenuation of conditioned responding from Acquisition to Novel Acquisition, whereas Yoke subjects do not, and between-group difference in the change in conditioned responses are predominantly driven by CS<sup>+</sup> responses in Novel Acquisition.

To rule out potential confounds for these behavioral results, we tested for correlations between differential SCR during Retrieval and Novel Acquisition and behavioral measures, such as shock level, shock ratings, mean number of button presses per trial, age, and questionnaire scores (STAI, ICI, PSS, IUS), which did not significantly differ between groups but were individually variable. We found no significant correlations between any of these factors and differential SCR in either of the day 2 phases (*p* values >0.09).

We were additionally interested in whether differential SCR during Retrieval and Novel Acquisition related to subjects' self-reported subjective sense of control during the Avoidance/Extinction phase. However, we found no correlation between this rating and differential SCR during Retrieval or Novel Acquisition, across the entire group or in Master subjects alone (*p* values >0.26).

As an additional control analysis, we tested whether the primary results remained significant when excluding 9 pairs of subjects who received one or more shocks during Avoidance/Extinction, as these Yoked subjects experienced an abbreviated extinction phase. In the 2 (CS type)  $\times$  2 (group)  $\times$  4 (phase) ANOVA, we again found significant main effects of CS type

( $F_{(1,36)} = 93.59, p = 1 \times 10^{-11}$ ) and phase ( $F_{(3,108)} = 14.6, p = 5 \times 10^{-8}$ ), as well as a significant CS type  $\times$  phase interaction ( $F_{(3,108)} = 9.79, p = 0.00009$ ) and a significant CS type  $\times$  phase  $\times$  group interaction ( $F_{(3,108)} = 3.35, p = 0.02$ ). The *post hoc* ANOVA investigating only the Avoidance/Extinction and Retrieval phases revealed, as in the analysis with all subjects, a significant main effect of CS type ( $F_{(1,36)} = 49.08, p = 3 \times 10^{-8}$ ), a trend-level CS type  $\times$  phase interaction ( $F_{(1,36)} = 2.76, p = 0.11$ ), and a significant CS type  $\times$  phase  $\times$  group interaction ( $F_{(1,36)} = 7.91, p = 0.01$ ). Similarly, the *post hoc* ANOVA investigating the Initial and Novel Acquisition phases yielded significant main effects of CS type ( $F_{(1,36)} = 73.67, p = 3 \times 10^{-10}$ ), phase ( $F_{(1,36)} = 24.99, p = 0.00002$ ), and a significant interaction between CS type, phase, and group ( $F_{(1,36)} = 4.75, p = 0.04$ ). Thus, the group differences in both Spontaneous Recovery and attenuation of acquisition are also evident in this subset of our full sample.

## fMRI

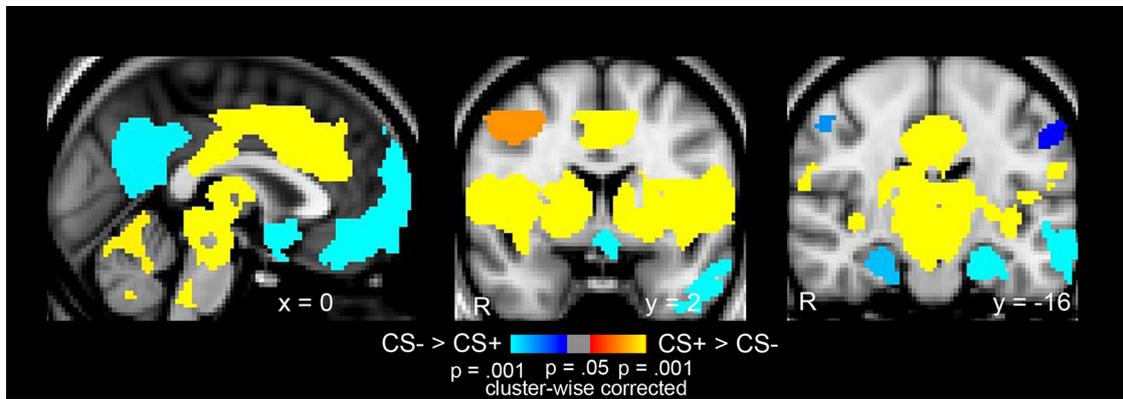
### Acquisition

During the Acquisition phase, subjects showed CS<sup>+</sup> versus CS<sup>-</sup> BOLD signal in regions typically identified in threat conditioning studies, similar to the results of a recent meta-analysis (Fullana et al., 2016) (Figure 4; Table 1).

### Avoidance/Extinction: responses during CS<sup>+</sup> versus CS<sup>-</sup> presentation

In the analyses of Avoidance/Extinction during cue presentation (both during late Avoidance/Extinction and across the whole phase), Master and Yoke subjects did not show differences in CS<sup>+</sup> versus CS<sup>-</sup> BOLD response in our regions of interest (vmPFC, striatum, and amygdala) that survived small volume correction. Master subjects showed CS<sup>+</sup> > CS<sup>-</sup> BOLD responses in putamen, caudate, and mPFC (Tables 2, 3), and no other within-group clusters survived small volume correction in these analyses.

During late Avoidance/Extinction, in the analysis of CS<sup>+</sup> > CS<sup>-</sup> responses during cue presentation, we observed a Masters > Yokes between-group difference in vmPFC that did not survive small volume correction but did survive an uncorrected threshold of  $p \leq 0.001$ , 10 voxels (MNI: -4, 34, -8, uncorrected *p* value of peak  $\leq 0.0002$ , 15 voxels; Fig. 5). This difference appears to be



**Figure 4.** CS<sup>+</sup> versus CS<sup>−</sup> BOLD responses during Acquisition, across the whole group (*N* = 56). FWE corrected, using nonparametric clusterwise correction for comparisons across the whole brain, with a cluster-forming threshold of *t* = 3, clusterwise *p* < 0.05, two-tailed. Colored pixels represent the clusterwise corrected *p* values.

**Table 1.** Whole brain-corrected results during Acquisition phase in the whole group (*N* = 56)<sup>a</sup>

Region	MNI ( <i>x, y, z</i> )	Voxels	Clusterwise corrected <i>p</i>	Voxelwise corrected <i>p</i>
<b>CS<sup>+</sup> &gt; CS<sup>−</sup></b>				
R frontal operculum, extending to L frontal operculum, bilateral insula, caudate, putamen, anterior cingulate cortex, thalamus, brainstem	38, 20, 8	23206	≤0.0004	≤0.0004
R supramarginal gyrus	60, −42, 34	3077	0.002	≤0.0004
L cerebellum	−32, −58, −30	1470	0.004	0.002
R superior frontal gyrus, extending to R middle frontal gyrus	32, 48, 24	1223	0.007	≤0.0004
L middle frontal gyrus, extending to L superior frontal gyrus	−30, 42, 26	641	0.02	≤0.0004
R precentral gyrus	46, 0, 48	611	0.022	≤0.0004
R cerebellum	34, −58, −30	384	0.042	0.085
R precuneus	14, −66, 36	372	0.045	0.001
<b>CS<sup>−</sup> &gt; CS<sup>+</sup></b>				
L superior frontal gyrus, extending to vmPFC	−12, 64, 20	7135	≤0.0004	≤0.0004
L hippocampus, extending to L amygdala	−22, −18, −20	3978	0.002	≤0.0004
L precuneus	−2, −60, 32	2619	0.003	≤0.0004
L superior lateral occipital cortex	−38, −74, 36	1744	0.006	≤0.0004
R hippocampus, extending to R amygdala	22, −10, −20	1071	0.013	≤0.0004
R superior lateral occipital cortex	52, −64, 26	831	0.018	≤0.0004
R postcentral gyrus	58, −6, 32	716	0.022	≤0.0004
R superior lateral occipital cortex	40, −88, 8	449	0.04	0.457
L postcentral gyrus	−58, −12, 36	400	0.046	≤0.0004

<sup>a</sup>FWE corrected, using nonparametric clusterwise correction for comparisons across the whole brain with a cluster-forming threshold of *t* = 3, clusterwise *p* ≤ 0.05, two-tailed. Voxelwise whole-brain corrected *p* values are listed for the peak voxel of each cluster. In a two-tailed permutation test with 5000 iterations, *p* values can only be resolved down to 0.0004.

**Table 2.** SVC results during late Avoidance/Extinction phase, during CS<sup>+</sup> versus CS<sup>−</sup> presentation<sup>a</sup>

Region	MNI ( <i>x, y, z</i> )	Voxels	Clusterwise corrected <i>p</i>	Voxelwise corrected <i>p</i>
<b>Masters CS<sup>+</sup> &gt; CS<sup>−</sup></b>				
L putamen	−30, −6, −2	110	0.013	0.013

<sup>a</sup>FWE corrected, using nonparametric clusterwise correction for comparisons within bilateral amygdala, vmPFC, caudate, accumbens, and putamen with a cluster-forming threshold of *t* = 3, clusterwise *p* ≤ 0.05 (one-tailed). Voxelwise SVC *p* values are listed for the peak voxel of each cluster. *N* = 27 in each group.

**Table 3.** SVC results during entire Avoidance/Extinction phase, during CS<sup>+</sup> versus CS<sup>−</sup> presentation<sup>a</sup>

Region	MNI ( <i>x, y, z</i> )	Voxels	Clusterwise corrected <i>p</i>	Voxelwise corrected <i>p</i>
<b>Masters CS<sup>+</sup> &gt; CS<sup>−</sup></b>				
R caudate (extending to R putamen)	6, 8, 4	203	0.004	0.0004
L caudate	−8, 12, 0	71	0.032	0.043
R mPFC	8, 42, 4	62	0.039	0.001

<sup>a</sup>FWE corrected, using nonparametric clusterwise correction for comparisons within bilateral amygdala, vmPFC, caudate, accumbens, and putamen with a cluster-forming threshold of *t* = 3, clusterwise *p* ≤ 0.05 (one-tailed). Voxelwise SVC *p* values are listed for the peak voxel of each cluster. *N* = 28 in each group.

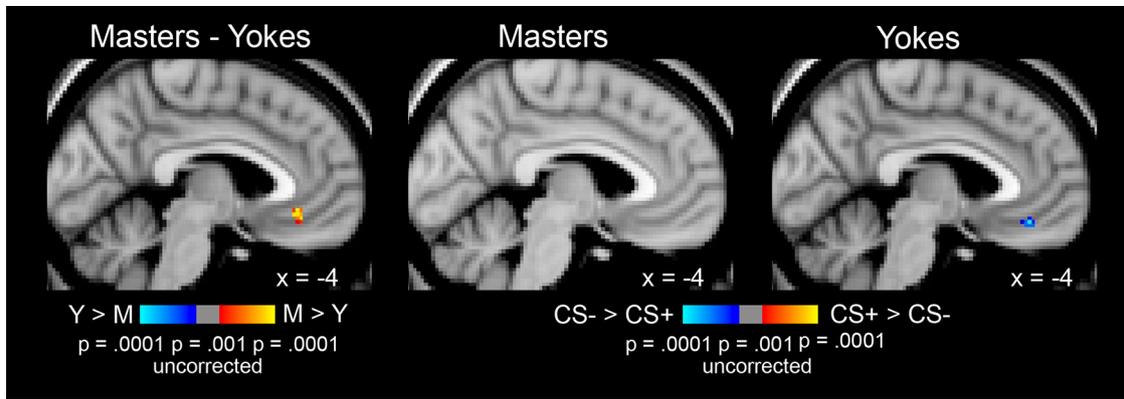
driven by CS<sup>−</sup> > CS<sup>+</sup> BOLD vmPFC signal in the Yokes (MNI: −4, 38, −14, uncorrected *p* value of peak ≤ 0.0002, 15 voxels).

*Avoidance/Extinction: responses during CS<sup>+</sup> versus CS<sup>−</sup> offset*

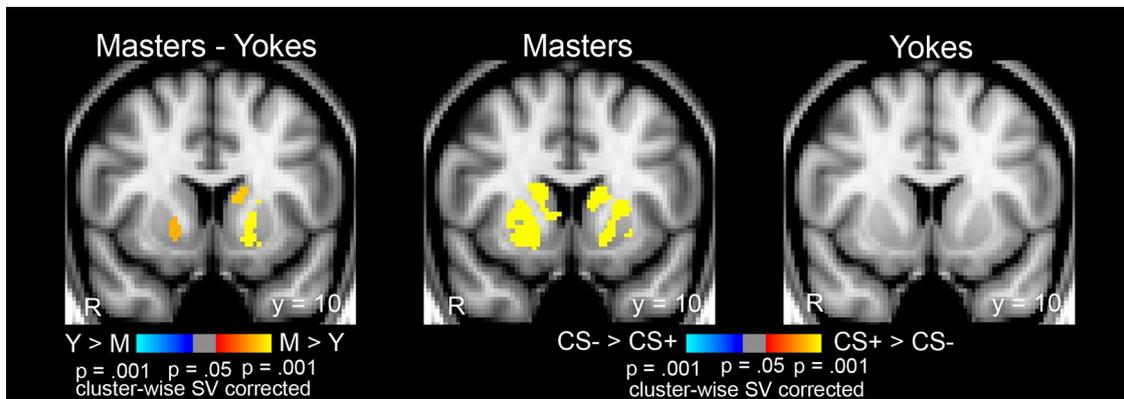
In the analysis of CS<sup>+</sup> versus CS<sup>−</sup> offset in late Avoidance/Extinction, Master subjects showed increased CS<sup>+</sup> versus CS<sup>−</sup> BOLD signal compared with Yoke subjects in left caudate and bilateral putamen (driven by bilateral caudate and putamen CS<sup>+</sup> > CS<sup>−</sup> offset responses in Master subjects; Fig. 6; Table 4). A similar pattern emerged when examining all trials of Avoidance/Extinc-

tion: Masters showed greater CS<sup>+</sup> > CS<sup>−</sup> BOLD signal than Yokes in left caudate and putamen (Table 5).

In the uncorrected analysis of CS<sup>+</sup> > CS<sup>−</sup> offset during all trials of Avoidance/Extinction, there was a Master > Yoke between-group difference in the mPFC (MNI: 10, 44, 4, uncorrected *p* value of peak ≤ 0.0002, 25 voxels). The between-group difference was driven by CS<sup>+</sup> > CS<sup>−</sup> offset signal in the Master subjects (for within Master group small volume corrected [SVC] mPFC result, see Table 5).



**Figure 5.** BOLD responses in late Avoidance/Extinction during CS<sup>+</sup> versus CS<sup>-</sup> presentation. *N* = 27 in each group. Uncorrected threshold of *p* = 0.001, one-tailed, 10 voxels. Image is masked by regions of interest (in this slice, vmPFC). M, Masters; Y, Yokes.



**Figure 6.** BOLD responses in late Avoidance/Extinction, CS<sup>+</sup> versus CS<sup>-</sup> offset. *N* = 27 in each group. FWE corrected, using nonparametric clusterwise correction for comparisons across voxels in regions of interest, with a cluster-forming threshold of *t* = 3, clusterwise *p* ≤ 0.05, one-tailed. Image is masked by regions of interest (in this slice, bilateral caudate, putamen, accumbens). Colored pixels represent the clusterwise SVC *p* values. M, Masters; Y, Yokes; SV, small volume.

**Table 4.** SVC results during late Avoidance/Extinction phase, CS<sup>+</sup> versus CS<sup>-</sup> offset<sup>a</sup>

Region	MNI ( <i>x, y, z</i> )	Voxels	Clusterwise corrected <i>p</i>	Voxelwise corrected <i>p</i>
<b>Masters &gt; Yokes</b>				
L putamen	-20, 4, 8	328	0.001	0.018
L caudate	-12, 8, 10	87	0.012	0.062
R putamen	22, -2, 6	71	0.016	0.05
<b>Masters CS<sup>+</sup> &gt; CS<sup>-</sup></b>				
L putamen (extending to L caudate)	-22, -2, 6	767	0.0004	0.001
R caudate (extending to R putamen)	18, 6, 16	741	0.0004	0.0004

<sup>a</sup>FWE corrected, using nonparametric clusterwise correction for comparisons within bilateral amygdala, vmPFC, caudate, accumbens, and putamen with a cluster-forming threshold of *t* = 3, clusterwise *p* ≤ 0.05 (one-tailed). Voxelwise SVC *p* values are listed for the peak voxel of each cluster. *N* = 27 in each group.

**Table 5.** SVC results during entire Avoidance/Extinction phase, CS<sup>+</sup> versus CS<sup>-</sup> offset<sup>a</sup>

Region	MNI ( <i>x, y, z</i> )	Voxels	Clusterwise corrected <i>p</i>	Voxelwise corrected <i>p</i>
<b>Masters &gt; Yokes</b>				
L putamen	-18, 4, 6	111	0.009	0.101
L caudate	-14, 6, 12	58	0.027	0.038
<b>Masters CS<sup>+</sup> &gt; CS<sup>-</sup></b>				
R putamen (extends to R caudate)	24, -4, 4	914	≤0.0002	0.003
L putamen (extends to L caudate)	-22, -6, 8	692	0.0004	0.001
R mPFC	10, 44, 4	95	0.017	0.014
<b>Yokes CS<sup>-</sup> &gt; CS<sup>+</sup></b>				
L vmPFC	-2, 28, -18	80	0.019	0.135

<sup>a</sup>FWE corrected, using nonparametric clusterwise correction for comparisons within bilateral amygdala, vmPFC, caudate, accumbens, and putamen with a cluster-forming threshold of *t* = 3, clusterwise *p* ≤ 0.05 (one-tailed). Voxelwise SVC *p* values are listed for the peak voxel of each cluster. In a one-tailed permutation test with 5000 iterations, *p* values can only be resolved down to 0.0002. *N* = 28 in each group.

**Retrieval and Novel Acquisition**

No regions showed a significant between-group difference in BOLD response that survived small volume correction during Retrieval or Novel Acquisition. In both phases, both Master and Yoke within-group maps show a similar pattern to Acquisition (Tables 6, 7).

However, during Retrieval, we observed between-group differences in vmPFC and orbitofrontal cortex at the uncorrected threshold (MNI: -6, 42, -24, uncorrected *p* value of peak ≤ 0.0002, 10 voxels, MNI: 12, 20, -20, uncorrected *p* value of peak ≤ 0.0002, 12 voxels). These between-group vmPFC clusters were driven by CS<sup>-</sup> > CS<sup>+</sup> BOLD signal in Yokes (for within Yoke group SVC vmPFC result, see Table 6).

**Discussion**

**Summary of findings**

Prior rodent research has demonstrated that behavioral control over an aversive stimulus reduces Pavlovian threat responses, even when control is no longer possible, and involves the vmPFC and striatum (Baratta et al., 2007; Cain and LeDoux, 2007; Moscarello and LeDoux, 2013; Ramirez et al., 2015). In this study, we found that control reduced threat responding and engaged similar neurocircuitry in humans. Whereas yoked subjects' conditioned responding increased from Extinction to Retrieval, sub-

**Table 6. SVC results during Retrieval phase<sup>a</sup>**

Region	MNI (x, y, z)	Voxels	Clusterwise corrected <i>p</i>	Voxelwise corrected <i>p</i>
Masters CS <sup>+</sup> > CS <sup>-</sup>				
R caudate	12, 2, 14	89	0.016	0.237
L caudate	-10, 2, 14	75	0.023	0.075
Masters CS <sup>-</sup> > CS <sup>+</sup>				
L vmPFC	-6, 56, -4	297	0.001	0.124
L amygdala	-24, -8, -18	135	0.005	0.003
R amygdala	24, -2, -26	47	0.048	0.192
Yokes CS <sup>+</sup> > CS <sup>-</sup>				
R caudate	10, 10, 12	81	0.015	0.074
Yokes CS <sup>-</sup> > CS <sup>+</sup>				
R vmPFC	8, 42, -14	257	0.001	0.084

<sup>a</sup>FWE corrected, using nonparametric clusterwise correction for comparisons within bilateral amygdala, vmPFC, caudate, accumbens, and putamen with a cluster-forming threshold of  $t = 3$ , clusterwise  $p \leq 0.05$  (one-tailed). Voxelwise SVC *p* values are listed for the peak voxel of each cluster.  $N = 28$  in each group.

**Table 7. SVC results during Novel Acquisition<sup>a</sup>**

Region	MNI (x, y, z)	Voxels	Clusterwise corrected <i>p</i>	Voxelwise corrected <i>p</i>
Masters CS <sup>+</sup> > CS <sup>-</sup>				
L putamen	-22, 8, 10	311	0.003	0.044
R putamen	20, 8, 2	234	0.006	0.003
R caudate	12, 4, 14	171	0.01	0.005
Masters CS <sup>-</sup> > CS <sup>+</sup>				
R vmPFC	2, 56, -8	939	0.0004	$\leq 0.0002$
R vmPFC	2, 14, -8	173	0.01	0.008
R amygdala	30, 0, -22	133	0.014	0.02
L hippocampus/amygdala	-22, -10, -18	127	0.015	0.014
Yokes CS <sup>+</sup> > CS <sup>-</sup>				
L putamen	-30, -2, 0	248	0.005	0.049
L caudate	-8, 8, 12	69	0.03	0.116
R caudate	10, 10, 4	48	0.048	0.304
Yokes CS <sup>-</sup> > CS <sup>+</sup>				
R vmPFC	2, 20, -8	438	0.002	0.003

<sup>a</sup>FWE corrected, using nonparametric clusterwise correction for comparisons within bilateral amygdala, vmPFC, caudate, accumbens, and putamen with a cluster-forming threshold of  $t = 3$ , clusterwise  $p \leq 0.05$  (one-tailed). In a one-tailed permutation test with 5000 iterations, *p* values can only be resolved down to 0.0002. Voxelwise SVC *p* values are listed for the peak voxel of each cluster.  $N = 26$  in each group.

jects who underwent signaled active avoidance showed no change in conditioned response from Avoidance to Retrieval, suggesting that active avoidance was more effective than the extinction procedure in producing a lasting reduction in threat responses. However, we did not observe a between-group difference in spontaneous recovery as measured during early Retrieval. The results suggest that, although the threat memory recovered early in Retrieval in Master subjects, it attenuated more rapidly than in Yoke subjects. We also present a novel finding that avoidance learning can attenuate Pavlovian threat responses during subsequent threat conditioning with previously unseen stimuli. Master subjects showed reduced conditioned responding (compared with initial acquisition) during an acquisition phase with novel stimuli, even though no avoidance response could be performed during that phase. In contrast, yoked extinction subjects showed no change in acquisition across days. These behavioral differences were accompanied by group differences in BOLD signal in the striatum, and at a threshold uncorrected for multiple comparisons, the vmPFC, during Avoidance/Extinction. A between-group vmPFC difference was also observed, at an uncorrected threshold, during Retrieval.

### Effect of active avoidance on threat responses

A small experimental literature on avoidance in humans and a large clinical literature have yielded conflicting findings regard-

ing whether avoidance attenuates threat responses or patients' fear. Consistent with the results of experiments in rodents discussed above, some studies in humans (Hartley et al., 2014; C. A. Hartley, C. A. Coelho, E. A. Boeke, F. Ramirez, and E. A. Phelps, unpublished observations) have found that learning to prevent shocks eliminated spontaneous recovery. Other experiments in humans, as well as clinical studies, have found that avoidance or safety behavior during extinction or exposure therapy can hinder the reduction of threat responses or fear (Lovibond et al., 2009; Blakey and Abramowitz, 2016; Volders et al., 2012). The interpretation of these studies is that subjects attribute the lack of aversive outcome to their action, instead of learning that the CS or anxiety-inducing situation is safe ("protection from extinction") (Chorazyna, 1962). Various factors, such as whether the behavioral response is learned or instructed, active or passive, whether the omission of the aversive outcome is contingent upon one's action, and whether the action engenders a strong sense of internal control, may modulate the effects of avoidance. Further studies are required to fully understand the discrepancies in results.

### Neurocircuitry of active avoidance

Previous fMRI studies have implicated the striatum in avoidance of shock, pain, or aversive images or sounds (Jensen et al., 2003; Delgado et al., 2009; Levita et al., 2012; Seymour et al., 2012; Bolstad et al., 2013; Collins et al., 2014; Eldar et al., 2016; Schlund et al., 2016). Several studies observed striatal activation during cue presentation, but they compared the BOLD responses within-subjects during a threatening cue associated with an avoidance response and a nonthreatening cue in a single cohort of subjects (Jensen et al., 2003; Levita et al., 2012; Bolstad et al., 2013; Collins et al., 2014). Here, we observed striatal BOLD responses in Masters during threat cue presentation, relative to the nonthreat cue, and the absence of significant between-group differences suggests that Yokes also showed subthreshold striatal BOLD responses during threat cue presentation. Thus, cue-related striatal responses observed in previous studies of avoidance may be primarily attributable to the threat association with the cue, rather than the avoidance response itself. In contrast, we found that subjects undergoing Avoidance showed greater striatal responses at cue offset, when the no-shock "outcome" occurred, compared with subjects undergoing passive extinction. This signal in the Master subjects may reflect a prediction error, or the discrepancy between expectation and outcome. Consistent with previous findings suggesting distinct localization of Pavlovian and instrumental learning signals (O'Doherty et al., 2004), this finding suggests that shock omission outcomes resulting from action elicit greater responses in the caudate and putamen than when these events are passively observed. As the majority of subjects learned the avoidance response in one trial, our study did not lend itself to trial-by-trial modeling of learning signals and these analyses were exploratory. However, three decision-making studies (Seymour et al., 2012; Roy et al., 2014; Eldar et al., 2016) that involved avoiding pain or shock examined prediction errors to successful and unsuccessful avoidance, and two found evidence for striatal prediction errors to successful avoidance of shock.

Our finding of a between-group difference in vmPFC BOLD signal during cue presentation in late Avoidance/Extinction and Retrieval (at an uncorrected threshold) is also consistent with the rodent literature suggesting a key role for the vmPFC in avoidance learning and its subsequent behavioral consequences (Baratta et al., 2007; Moscarello and LeDoux, 2013). Research in rats has suggested that vmPFC activity may mediate the persistent effects of

behavioral control on reduction of Pavlovian responses (Baratta et al., 2007; Moscarello and LeDoux, 2013). We observed a between-group difference during Retrieval in vmPFC, suggesting that a similar mechanism may underlie the long-term consequences of control over threat in humans. Between-group differences during cue presentation in Avoidance/Extinction and Retrieval appeared to be driven by  $CS^- > CS^+$  BOLD signals in Yoke subjects, with Masters showing reduced differentiation between  $CS^+$  and  $CS^-$  in vmPFC. However, the vmPFC differences during Avoidance/Extinction and Retrieval observed here should be interpreted cautiously, as they did not survive small volume correction.

The amygdala plays a central role in the acquisition of threat associations and their modulation through extinction and avoidance (Phelps and LeDoux, 2005). However, we did not observe any differences in amygdala BOLD signal between Masters and Yokes. The central nucleus and its brainstem-mediated behavioral output (freezing) are inhibited in avoidance learning (Moscarello and LeDoux, 2013). However, the basal nucleus of the amygdala is engaged during avoidance learning and may transfer information about CS–US association to the striatum (Amorapanth et al., 2000; Ramirez et al., 2015). Because these small subregions may exhibit opposing-directional effects, our ability to find differences in the amygdala may have been limited by the resolution possible with standard fMRI techniques.

### Limitations

The present study has several limitations. One limitation is that, because of the yoked design, subjects in the yoked extinction condition did not undergo a standard extinction procedure, where the removal of the US is the only feature that differentiates extinction from acquisition. During the “Extinction” phase, the grid was present on the screen and subjects performed a novel task, details that make this phase distinct from the Acquisition and Retrieval phases. Additionally, some subjects received shocks in the early part of “Extinction.” Although other studies of extinction have observed reextinction of threat responses during retrieval (Phelps et al., 2004), Yoke subjects in this study exhibited incomplete reextinction. In this study, the motor control task or the presence of shocks during early “Extinction” might have contributed to increased threat responses in the Retrieval phase in the Yoke subjects and heightened the difference in spontaneous recovery between the two groups. However, our control SCR analyses that included only subjects who did not receive shocks in Avoidance/Extinction suggest that the subjects who did receive shocks are not driving the between-group differences in spontaneous recovery and attenuation of acquisition. Additionally, some have criticized the yoked control design (Church, 1964) because individual differences in the effectiveness of the US or other factors may lead to systematic between-group differences. Another limitation is the differential instruction between groups. To promote learning, we chose to tell Master subjects that they could learn a response to prevent the shocks; however, the correct avoidance action itself was not instructed. It is possible that behavioral or neural differences could be influenced by these differences in instruction. However, a previous study observed similar behavioral effects without any differential instruction between groups (Hartley et al., 2014).

In conclusion, the present study extends the literature on the neural substrates of human avoidance learning, using a novel between-subjects design that allows for dissociation between aversive reinforcement history and avoidance. Signaled active avoidance provided an effective way to regulate threat responses

that appeared to last longer and generalize more broadly than passive extinction. The present evidence that coping actively with threat can attenuate subsequent threat responding, even in uncontrollable or novel contexts, has potential implications for the treatment of anxiety disorders. Interventions that foster a sense of control over threat might serve to ameliorate fearful responding in anxiety or foster greater resilience in those at high risk for anxiety disorders. Moreover, our findings suggest that using active coping strategies in the face of everyday challenges may help to promote psychological resilience.

### References

- Amorapanth P, LeDoux JE, Nader K (2000) Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nat Neurosci* 3:74–79. [CrossRef Medline](#)
- Baratta MV, Christianson JP, Gomez DM, Zarza CM, Amat J, Masini CV, Watkins LR, Maier SF (2007) Controllable versus uncontrollable stressors bi-directionally modulate conditioned but not innate fear. *Neuroscience* 146:1495–1503. [CrossRef Medline](#)
- Baratta MV, Lucero TR, Amat J, Watkins LR, Maier SF (2008) Role of the ventral medial prefrontal cortex in mediating behavioral control-induced reduction of later conditioned fear. *Learn Mem* 15:84–87. [CrossRef Medline](#)
- Blakey SM, Abramowitz JS (2016) The effects of safety behaviors during exposure therapy for anxiety: critical analysis from an inhibitory learning perspective. *Clin Psychol Rev* 49:1–15. [CrossRef Medline](#)
- Bolstad I, Andreassen OA, Reckless GE, Sigvartsen NP, Server A, Jensen J (2013) Aversive event anticipation affects connectivity between the ventral striatum and the orbitofrontal cortex in an fMRI avoidance task. *PLoS One* 8:e68494. [CrossRef Medline](#)
- Cain CK, LeDoux JE (2007) Escape from fear: a detailed behavioral analysis of two atypical responses reinforced by CS termination. *J Exp Psychol Anim Behav Process* 33:451–463. [CrossRef Medline](#)
- Chorzyna H (1962) Some properties of conditioned inhibition. *Acta Biol Exp* 22:5–13. [Medline](#)
- Church RM (1964) Systematic effect of random error in the yoked control design. *Psychol Bull* 62:122–131. [CrossRef Medline](#)
- Cohen S, Kamarck T, Mermelstein R (1983) A global measure of perceived stress. *J Health Soc Behav* 24:385–396. [CrossRef Medline](#)
- Collins KA, Mendelsohn A, Cain CK, Schiller D (2014) Taking action in the face of threat: neural synchronization predicts adaptive coping. *J Neurosci* 34:14733–14738. [CrossRef Medline](#)
- Delgado MR, Jou RL, Ledoux JE, Phelps EA (2009) Avoiding negative outcomes: tracking the mechanisms of avoidance learning in humans during fear conditioning. *Front Behav Neurosci* 3:33. [CrossRef Medline](#)
- Duttweiler PC (1984) The internal control index: a newly developed measure of locus of control. *Educ Psychol Measure* 44:209–221. [CrossRef](#)
- Ekman P, Friesen W (1976) *Measuring facial movement: environmental psychology and nonverbal communication*. San Francisco: Human Sciences.
- Eldar E, Hauser TU, Dayan P, Dolan RJ (2016) Striatal structure and function predict individual biases in learning to avoid pain. *Proc Natl Acad Sci U S A* 113:4812–4817. [CrossRef Medline](#)
- Freeston MH, Rhéaume J, Letarte H, Dugas MJ, Ladouceur R (1994) Why do people worry? *Pers Individual Differences* 17:791–802. [CrossRef](#)
- Fullana MA, Harrison BJ, Soriano-Mas C, Vervliet B, Cardoner N, Ávila-Parcet A, Radau J (2016) Neural signatures of human fear conditioning: an updated and extended meta-analysis of fMRI studies. *Mol Psychiatry* 21:500–508. [CrossRef Medline](#)
- Hackel LM, Doll BB, Amodio DM (2015) Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice. *Nat Neurosci* 18:1233–1235. [CrossRef Medline](#)
- Hartley CA, Gorun A, Reddan MC, Ramirez F, Phelps EA (2014) Stressor controllability modulates fear extinction in humans. *Neurobiol Learn Mem* 113:149–156. [CrossRef Medline](#)
- Jensen J, McIntosh AR, Crawley AP, Mikulis DJ, Remington G, Kapur S (2003) Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron* 40:1251–1257. [CrossRef Medline](#)
- LeDoux JE (2000) Emotion circuits in the brain. *Annu Rev Neurosci* 23:155–184. [CrossRef Medline](#)
- LeDoux JE, Moscarello J, Sears R, Campese V (2017) The birth, death and

- resurrection of avoidance: a reconceptualization of a troubled paradigm. *Mol Psychiatry* 22:24–36. [CrossRef Medline](#)
- Levita L, Hoskin R, Champi S (2012) Avoidance of harm and anxiety: a role for the nucleus accumbens. *Neuroimage* 62:189–198. [CrossRef Medline](#)
- Lovibond PF, Mitchell CJ, Minard E, Brady A, Menzies RG (2009) Safety behaviours preserve threat beliefs: protection from extinction of human fear conditioning by an avoidance response. *Behav Res Ther* 47:716–720. [CrossRef Medline](#)
- Maier SF (2015) Behavioral control blunts reactions to contemporaneous and future adverse events: medial prefrontal cortex plasticity and a corticostriatal network. *Neurobiol Stress* 1:12–22. [CrossRef Medline](#)
- Moscarello JM, LeDoux JE (2013) Active avoidance learning requires prefrontal suppression of amygdala-mediated defensive reactions. *J Neurosci* 33:3815–3823. [CrossRef Medline](#)
- O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454. [CrossRef Medline](#)
- Phelps EA, LeDoux JE (2005) Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* 48:175–187. [CrossRef Medline](#)
- Phelps EA, Delgado MR, Nearing KI, LeDoux JE (2004) Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* 43:897–905. [CrossRef Medline](#)
- Ramirez F, Moscarello JM, LeDoux JE, Sears RM (2015) Active avoidance requires a serial basal amygdala to nucleus accumbens shell circuit. *J Neurosci* 35:3470–3477. [CrossRef Medline](#)
- Roy M, Shohamy D, Daw N, Jepma M, Wimmer GE, Wager TD (2014) Representation of aversive prediction errors in the human periaqueductal gray. *Nat Neurosci* 17:1607–1612. [CrossRef Medline](#)
- Schlund MW, Brewer AT, Magee SK, Richman DM, Solomon S, Ludlum M, Dymond S (2016) The tipping point: value differences and parallel dorsal-ventral frontal circuits gating human approach-avoidance behavior. *Neuroimage* 136:94–105. [CrossRef Medline](#)
- Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R (2012) Serotonin selectively modulates reward value in human decision-making. *J Neurosci* 32:5833–5842. [CrossRef Medline](#)
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23:S208–S219. [CrossRef Medline](#)
- Spielberger CD (1993) *State-trait anxiety inventory: a comprehensive bibliography*. Palo Alto, CA: Consulting Psychologists.
- Volders S, Meulders A, De Peuter S, Vervliet B, Vlaeyen JWS (2012) Safety behavior can hamper the extinction of fear of movement-related pain: an experimental investigation in healthy participants. *Behav Res Ther* 50:735–746. [CrossRef Medline](#)
- Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014) Permutation inference for the general linear model. *Neuroimage* 92:381–397. [CrossRef Medline](#)