



Avoiding negative outcomes: tracking the mechanisms of avoidance learning in humans during fear conditioning

Mauricio R. Delgado^{1*}, Rita L. Jou², Joseph E. LeDoux³ and Elizabeth A. Phelps^{2,3}

¹ Department of Psychology, Rutgers University, Newark, NJ, USA

² Department of Psychology, New York University, New York, NY, USA

³ Center for Neural Science, New York University, New York, NY, USA

Edited by:

Daeyeol Lee, Yale
University School of Medicine, USA

Reviewed by:

Jeansok J. Kim,
University of Washington, USA
Hackjin Kim, Korea University, Korea

*Correspondence:

Mauricio R. Delgado, Department of
Psychology, Rutgers University, 340
Smith Hall, 101 Warren Street, Newark,
NJ 07102, USA.
e-mail: delgado@psychology.rutgers.edu

Previous research across species has shown that the amygdala is critical for learning about aversive outcomes, while the striatum is involved in reward-related processing. Less is known, however, about the role of the amygdala and the striatum in learning how to exert control over emotions and avoid negative outcomes. One potential mechanism for active avoidance of stressful situations is postulated to involve amygdala–striatal interactions. The goal of this study was to investigate the physiological and neural correlates underlying avoidance learning in humans. Specifically, we used a classical conditioning paradigm where three different conditioned stimuli (CS) were presented. One stimulus predicted the delivery of a shock upon stimulus offset (CS+), while another predicted no negative consequences (CS–). A third conditioned cue also predicted delivery of a shock, but participants were instructed that upon seeing this stimulus, they could avoid the shock if they chose the correct action (AV+). After successful learning, participants could then easily terminate the shock during subsequent stimulus presentations (AV–). Physiological responses (as measured by skin conductance responses) confirmed a main effect of conditioning, particularly showing higher arousal responses during pre (AV+) compared to post (AV–) learning of an avoidance response. Consistent with animal models, amygdala–striatal interactions were observed to underlie the acquisition of an avoidance response. These results support a mechanism of active coping with conditioned fear that allows for the control over emotional responses such as fears that can become maladaptive and influence our decision-making.

Keywords: amygdala, striatum, negative reinforcement, instrumental conditioning, fMRI, reward, punishment

INTRODUCTION

The ability to modify and control our emotional responses is critical for adaptive function and goal-directed behavior. Although learning to fear a potentially dangerous situation is important, it is equally important to be able to modify this fear when new information is available, or use this fear to motivate adaptive action that diminishes the potential threat. Recent research examining the neural systems of regulating fear in humans has highlighted passive extinction techniques (Milad and Quirk, 2002; Knight et al., 2004; Phelps et al., 2004) and the use of cognitive strategies (Kalisch et al., 2005; Ochsner and Gross, 2005; Delgado et al., 2008b). These techniques focus on modifying the fear response in the presence of the fear-eliciting event. Another common response used to regulate fear, however, is to take an action to avoid the potential danger and diminish the fear response. Given how frequently action is used to cope with potential threat and fear outside the laboratory, surprisingly little research conducted in humans has examined the neural system mediating the active coping of fear. Research in non-human animals has suggested that active coping of fear may involve amygdala and striatal interactions (Killcross et al., 1997; Everitt et al., 1999; LeDoux and Gorman, 2001; Cardinal et al., 2002). The goal of the present study is to investigate if an amygdala–striatal circuitry underlies active coping of fear in humans.

The amygdala has been the focus of investigations of aversive learning, particularly Pavlovian fear conditioning paradigms in which fear is expressed passively, such as through autonomic responses (for review see Phelps and LeDoux, 2005). In contrast, the human striatum has been highlighted in investigations of reward-related processing, such as instrumental paradigms, that involve decision-making and action-contingencies (for review see Montague and Berns, 2002; O’Doherty, 2004; Knutson and Cooper, 2005; Delgado, 2007; Rangel et al., 2008). The human striatum, however, is also implicated in aversive learning (e.g., Jensen et al., 2003; Seymour et al., 2004, 2007; Menon et al., 2007; Delgado et al., 2008a). Studies from non-human animals have led to the hypothesis that one role for the striatum in aversive learning may be to aid in the acquisition of avoidance actions that diminish exposure to a fear-eliciting event (LeDoux and Gorman, 2001). For example, an investigation in rodents examining the amygdala subnuclei that mediate fear-motivated action found that the basal nucleus, which projects to the striatum (Mogenson et al., 1980; Robbins et al., 1989), is necessary for the acquisition of a fear-reducing action, but is not necessary for more passive expressions of conditioned fear (Amorapanth et al., 2000). In contrast, the central nucleus, which projects to the brainstem and hypothalamus, is necessary for the passive expression of fear, but not for learning a fear-reducing action. These results suggest that

partially independent neural circuits mediate active and passive means of fear expression and that the amygdala's connectivity with the striatum allows for active coping strategies to develop and diminish fear induced by a conditioned stimulus (Everitt et al., 1991; Amorpant et al., 2000; Cain and LeDoux, 2008). In support of this hypothesis, both dorsal and ventral striatum in rats have been implicated in various types of avoidance learning (Winocur and Mills, 1969; Allen and Davison, 1973; Neill et al., 1974; McCullough et al., 1993; Li et al., 2004).

In humans, neuroimaging experiments suggest that the striatum is involved in the expectation of an aversive stimulus, whether an opportunity to avoid exists or not (Jensen et al., 2003; Delgado et al., 2008a). However, less is known about the potential striatal-amygdala interactions that may underlie avoidance learning in the human brain. In this experiment, we used a modified aversive conditioning paradigm in conjunction with blood oxygenated level dependent (BOLD) and autonomic measures to explore the acquisition of an avoidance learning response. Participants were instructed that they could: (a) avoid a potential shock by learning a behavioral response (i.e., a button press), and (b) express the behavioral response after successful learning to prevent future shock delivery. As suggested by animal models (e.g., LeDoux and Gorman, 2001), we hypothesized that interactions of the amygdala and striatum would underlie a measure of successful avoidance learning, comparing BOLD responses pre- and post-learning.

MATERIALS AND METHODS

PARTICIPANTS

Thirty-two participants were initially recruited for this study (19F/13M, $M = 19.8$, $SD = 2.2$). Nine participants were excluded from further analysis due to excessive motion during scanning ($N = 4$, more than 2 mm of movement), failure to learn the task ($N = 3$) or equipment malfunction during session ($N = 2$, shocks not delivered). The final behavioral and neuroimaging analyses were conducted on 23 participants (15F/8M, mean age = 19.9, $SD = 2.6$). Participants responded to posted advertisement and all participants gave informed consent. The experiment was approved by the University Committee on Activities Involving Human Subjects at New York University.

PROCEDURE

The experiment consisted of an aversive conditioning paradigm with instruction. Participants were presented with three types of colored squares (e.g., blue, yellow, green) that served as conditioned stimuli (CS). Two of the CSs were fully predictable and led to the delivery of a mild shock to the wrist (the unconditioned stimulus, US) with either 100% (CS+ trials) or 0% (CS- trials) probability (certain trials). The third CS also predicted delivery of a mild shock, however participants were afforded a chance to avoid the shock if they learned the appropriate response (avoidable or AV trials). The AV trials were further subdivided into two types of trials according to learning success. During pre-learning trials (AV+ trials) participants attempted to learn how to avoid the negative outcome. Post-learning trials included subsequent presentations of the CS where successful avoidance of the US was maintained by the previously learned response (AV- trials). Thus, the four

conditions (CS+, CS-, AV+, AV-) comprised two 2 classes of CS (certain and avoidable) that varied with respect to conditioned response (aversive and safe; **Figure 1**).

Each CS presentation lasted 10 s and was broken down into a CS and a response phase. During the CS phase (4–6 s), participants were presented with the type of CS and instructed to just observe and wait for the response phase. The CS phase was the task period of interest and measures of physiological and BOLD responses reflect activation at this time point, uncontaminated by any motor responses or shock delivery. The response phase was cued by a question-mark in the middle of the colored square (4–6 s). At this time, participants were instructed to make a behavioral response (i.e., a button press). A mild shock was delivered during CS+ and AV+ trials for 200 ms that co-terminated with the end of the response phase. The trial ended with a 14-s inter-trial interval, for a total trial time of 24 s. Each session contained 24 trials, evenly divided into 6 trials for each type of condition (CS+, CS-, AV+, AV-).

For AV trials, participants had a chance to avoid the shock with the correct response during the response phase. Specifically, they were told that one of eight button presses could terminate the shock delivery. Participants were given an MRI compatible button box with four buttons and used their right hand to make one response per trial. They were further instructed that the “correct button” could be the first or second time a button was pressed, thus creating eight possible correct buttons and diminishing excessive motor coordination issues associated with the use of multiple button boxes. Participants were also asked to make a non-contingent button press during the response phase for certain trials (CS+, CS-) to control for motor requirements.

Prior to scanning, participants were instructed what each CS predicted (certain or avoidable outcome). Unbeknownst to participants, however, the correct button press during AV trials was always the

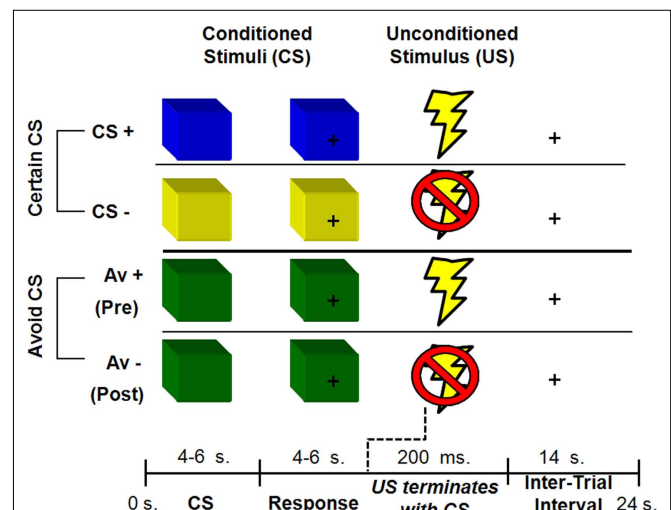


FIGURE 1 | Human avoidance paradigm. Participants were presented with three types of CS. Both CS+ and CS- predicted a certain outcome (an aversive shock or no shock respectively). The AV+ condition predicted a potential shock but afforded the participant an opportunity to avoid a shock with the correct behavioral response. An AV- trial referred to trials post-learning of an avoidance response. Colors were counterbalanced across scanning sessions.

response made in the sixth AV+ trial, irrespective of which button was pressed. The correct button press was then repeated post-learning, during the remaining six AV− trials. This ensured that all participants experienced the same schedule of reinforcement, with each session containing 24 trials, evenly divided into six trials for each type of condition (CS+, CS−, AV+, AV−). Participants who failed to learn the contingency (i.e., failed to repeat the correct button, and thus never experiencing AV− trials), typically reported not paying attention, and were excluded from all further analysis ($N = 3$).

The US constituted mild shocks delivered to the right wrist through a stimulating bar electrode connected to a Grass Medical Instruments stimulator. The stimulator was shielded for magnetic interference and grounded through an RF filter. Participants used a work-up procedure to set the appropriate shock level prior to the experimental session. Specifically, participants experienced a mild shock (10 V, 200 ms, 50 pulses/s) which was gradually increased up to a fixed maximum (60 V). They were instructed to set a level that was deemed uncomfortable, but not painful (mean shock level = 25.69 V, SD = 8.91).

Task events were programmed using E-PRIME software, v1.0 (PST, Pittsburgh, PA, USA). The color of the CSs was counterbalanced across sessions. Stimuli were presented in a black background and projected onto a screen which was visible inside the scanner through a mirror in the head coil. Right handed responses were made using an MRI compatible button box. At the end of the experimental session, participants were debriefed and compensated.

PHYSIOLOGICAL SET-UP, ASSESSMENT AND BEHAVIORAL ANALYSIS

Skin conductance responses (SCRs) were acquired from the participant's middle phalanges of the second and third fingers in the left hand and amplified by BIOPAC Systems skin conductance module. Shielded Ag–AgCl electrodes were grounded through an RF filter panel and served to acquire data. AcqKnowledge software was used to analyze the analog skin conductance waveforms. The level of SCR response was assessed for each trial as the base to peak amplitude difference in skin conductance of the largest deflection in the 0.5–4.5 s latency window after onset of the CS (see LaBar et al., 1995). A minimum response criterion of 0.02 μ S was used with lower responses scored as 0. Raw scores were square-root transformed prior to statistical analysis to normalize the distributions (LaBar et al., 1998). Acquired SCRs during the CS phase were then averaged per participant and per type of trial (CS+, CS−, AV+, AV−). A 2 × 2 repeated measures ANOVA with participants as a random factor was used to test for a main effect of type of CS (certain, avoidable) and conditioned response (aversive, safe). *Post hoc* paired *t*-tests were then conducted to probe differences between the contrast of interest, AV+ and AV− trials.

Additional behavioral data was acquired in the form of reaction time during the response phase, using an MRI compatible button box with four buttons. The primary analysis of the reaction time data was a paired *t*-test comparison of certain (CS+, CS−) and AV trials (AV+, AV−), hypothesized to differ with respect to motivation. Since the schedule of reinforcement was predetermined to reflect learning after six trials, accuracy differences were not expected, and participants who did not learn were excluded as previously described.

FMRI ACQUISITION AND ANALYSIS

A 3T Siemens Allegra head-only scanner and a Siemens standard head coil were used for data acquisition at NYU's Center for Brain Imaging. Anatomical images were acquired using a T1-weighted protocol (256 × 256 matrix, 176 1-mm sagittal slices) Functional images were acquired using a single-shot gradient echo EPI sequence (TR = 2000 ms, TE = 20 ms, FOV = 192 cm, flip angle = 75°, bandwidth = 4340 Hz/px, echo spacing = 0.29 ms). Thirty-five contiguous oblique-axial slices (3 mm × 3 mm × 3 mm voxels) parallel to the AC–PC line were obtained. Analysis of imaging data was conducted using Brain Voyager software (Brain Innovation, Maastricht, The Netherlands). The data was initially corrected for motion (using a threshold of 2 mm or less), and slice scan time using sinc interpolation was applied. Further, spatial smoothing was performed using a three-dimensional Gaussian filter (4-mm FWHM), along with voxel-wise linear detrending and high-pass filtering of frequencies (three cycles per time course). Structural and functional data of each participant was then transformed to standard Talairach stereotaxic space (Talairach and Tournoux, 1988).

A random-effects analysis was performed on the functional data using a general linear model (GLM) on 23 participants. There were four regressors of interest in the CS phase (CS+, CS−, AV+, AV−). There were also six regressors of no interest that modeled the response phase (separated into four types of trials according to condition) and the shock delivery (CS+_US and AV+_US). The principal contrast served to identify regions of interest (ROIs) involved in processing anticipated aversive outcomes during the CS phase, using a conservative threshold of FDR < 0.01 along with a cluster threshold of 10 contiguous voxels. Specifically, trials where an aversive outcome was expected (AV+, CS+) were compared to the most non-aversive control condition (CS−), as some residue conditioned fear could exist in AV− trials. Given the *a priori* hypothesis with respect to amygdala–striatal interactions, an amygdala ROI was functionally defined with this contrast using a more lenient threshold of $p < 0.025$ (uncorrected) along with a cluster threshold of four contiguous voxels (Buchel et al., 1998; LaBar et al., 1998). Mean parameter estimates reflecting effect size of a particular condition were extracted from ROIs in the striatum and amygdala for further analysis. A correlation analysis was then conducted comparing learning changes differences (i.e., mean parameter estimates for AV+ minus mean parameter estimates for AV−) between the functionally defined amygdala and striatum ROIs. Additionally, the time course of activation across the entire functional run for each individual participant was extracted from the amygdala ROI and used in an exploratory connectivity analysis. The time-course data was z-transformed and used as a single predictor in a GLM. The resulting activation map was thresholded at FDR < 0.001 and identified regions which hemodynamic patterns correlated with the seed amygdala ROI. Finally, an exploratory analysis was performed comparing AV+ and AV− trials, investigating brain regions associated with avoidance learning changes, and identified ROIs at $p < 0.005$ with four or more contiguous voxels.

RESULTS

BEHAVIORAL AND PHYSIOLOGICAL RESULTS

SCRs were acquired during the CS phase as a measure of physiological arousal. A main effect of conditioned response was observed [$F(1, 21) = 42.34, p < 0.0001$; **Figure 2**] suggesting that participants

were more aroused during presentation of CS that predicted aversive (CS+, AV+) compared to safer outcomes (CS-, AV-) consistent with earlier studies of human aversive conditioning (LaBar et al., 1998; Phelps et al., 2004). *Post hoc* paired *t*-tests revealed a differential response between CS+ and CS- trials [$t(22) = 4.71, p < 0.0002$] as expected based on previous findings. Importantly, a difference was also observed between AV+ and AV- [$t(22) = 4.73, p < 0.0002$], suggesting that learning an avoidable response in this paradigm decreases previously documented conditioned fear responses. No main effect of type of CS [$F(1, 21) = 0.3, p = 0.59$] or interaction [$F(2, 21) = 1.3, p = 0.27$] was observed using SCRs. Instead, differences across type of CS were seen in the reaction time data in the response phase. Specifically, participants were faster to make a

response to avoidable ($M = 593.55, SD = 161$) compared to certain ($M = 671.93, SD = 146.47$) trials [$F(1, 21) = 12.24, p < 0.002$], suggesting behavioral reactions were faster when motivation to respond was high (Delgado et al., 2004). These effects were particularly strong when comparing safe trials [AV- and CS-; $t(22) = 4.24, p < 0.0005$], but also approaching significance when contrasting the aversive trials [AV+ and CS+; $t(22) = 1.5, p = 0.14$]. No effects of conditioning [$F(1, 21) = 1.06, p = 0.31$] or interaction [$F(1, 21) = 0.31, p = 0.58$] were observed in the reaction time data.

NEUROIMAGING RESULTS

The main analysis used to identify ROIs involved a contrast of aversive trials, where an aversive outcome was expected (AV+, CS+), and the most non-aversive control condition (CS-). This contrast yielded positive activation patterns in an array of cortical regions (Table 1), but central to this investigation, we observed activation in the ventral and dorsal striatum bilaterally. Mean parameter estimates for individual participants were then extracted for the striatum ROIs for further analysis. Within the left ventral striatum ROI identified in this contrast (Figure 3A), which included the putamen, a differential response between AV+ and AV- trials was found using *post hoc t*-tests [$t(22) = 3.25, p < 0.005$]. This pattern also characterized BOLD responses in the right dorsal striatum ROI [Figure 3B; $t(22) = 2.37, p < 0.05$]. Interestingly, a differential response between AV+ and CS+ trials was seen in the left ventral striatum ROI [$t(22) = 2.22, p < 0.05$], but not within the voxels defined in the right dorsal striatum ROI [$t(22) = 0.94, p = 0.36$].

An amygdala ROI was functionally defined with the same contrast of aversive and safe trials, but using a more lenient threshold given the *a priori* hypothesis with respect to amygdala-striatal interactions.

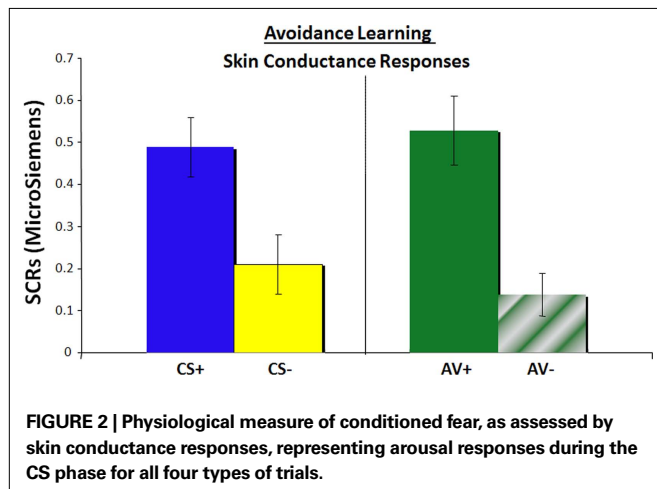
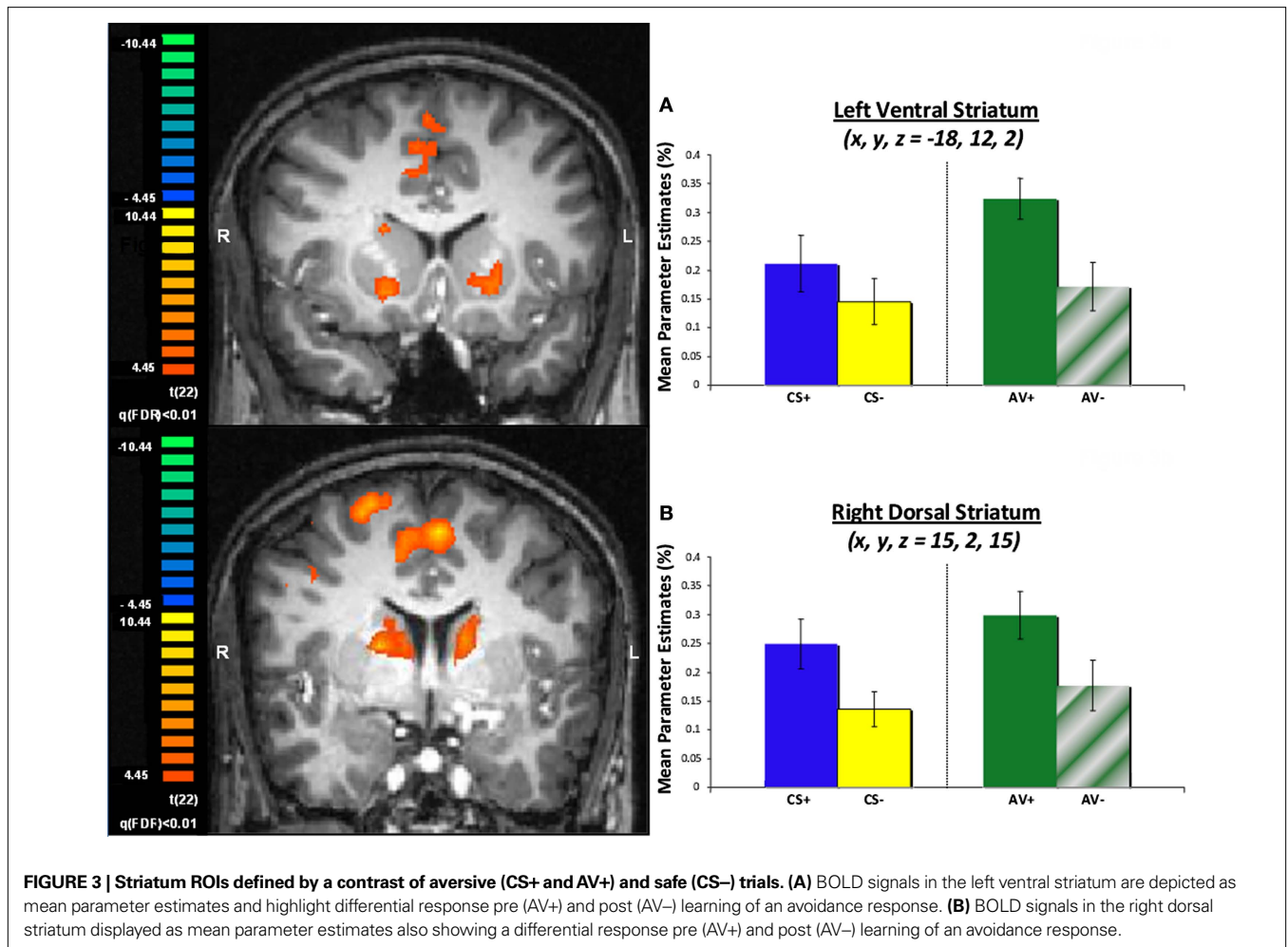


Table 1 | Contrast of aversive (CS+ and AV+) and Safe (CS-) trials at FDR <0.01 and contiguity threshold of 10 voxels.

Region of activation	Brodmann areas	Laterality	Talairach coordinates			#Voxels
			x	y	z	
Medial frontal gyrus	6	Left	0	-1	52	3965
Superior frontal gyrus	6	Right	28	-5	60	2275
Somatosensory cortex	1, 2, 3	Left	-39	-34	54	754
Superior parietal lobe	7	Left	-33	-51	50	330
Precentral gyrus	4	Left	-35	-23	56	1392
Precuneus	7	Left	-13	-70	48	672
Medial frontal gyrus	6, 8	Right	4	18	44	1433
Dorsolateral PFC	9	Right	43	4	36	445
Dorsolateral PFC	9	Left	-31	52	25	423
Inferior parietal lobe	40	Right	54	-47	30	750
Medial occipital gyrus	19	Left	-23	-86	19	552
Dorsal striatum		Right	15	2	15	2372
Dorsal striatum		Left	-13	-3	17	1008
Ventral striatum		Left	-18	12	2	492
Ventral striatum		Right	17	12	0	428
Inferior frontal gyrus	45	Left	-30	22	7	312
Lingual gyrus	18	Right	12	-62	6	380
Occipital lobe	17, 18		1	-80	-4	10996
Cerebellum		Left	-33	-52	-23	2297
Cerebellum		Right	31	-43	-21	2255



Activity within the amygdala ROI in the left hemisphere showed differential responses between CS+ and CS- trials [$t(22) = 2.28$, $p < 0.05$; **Figure 4**], consistent with previous literature and contrast used to define this ROI, while differential response between AV+ and AV- trials approached significance [$t(22) = 1.53$, $p = 0.14$].

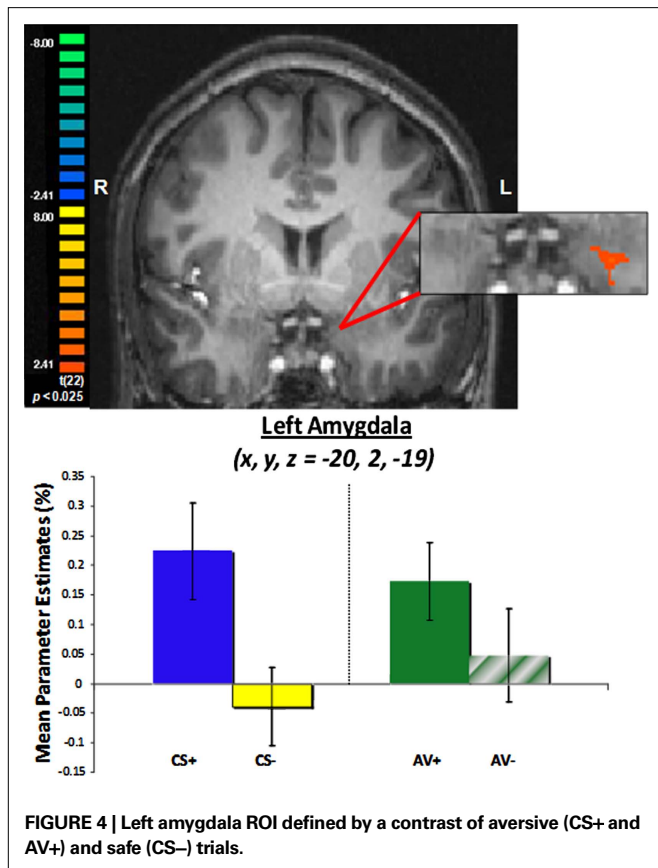
A measure of the magnitude of avoidance learning was calculated from mean parameter estimates (i.e., mean beta weights) with the goal of contrasting the *a priori* ROIs (i.e., amygdala and striatum) during task performance. Specifically, we used the difference between mean parameter estimates during AV+ and AV- trials, reflecting the difference between pre- and post-learning of an avoidance response. This measure of the magnitude of avoidance learning for the amygdala ROI was then correlated with the same measure for both left ventral striatum ($r = 0.51$, $p < 0.05$) and right dorsal striatum ($r = 0.54$, $p < 0.05$) ROIs previously described. To better quantify a potential interaction between the amygdala and striatum during avoidance learning, however, an exploratory connectivity analysis was performed where the time course of amygdala activation for each individual subject was used as a single predictor in a GLM. As hypothesized by animal models (LeDoux and Gorman, 2001), it was expected that the seed ROI, the amygdala activation pattern, would correlate with striatum activity during task performance. With the caveat that this analysis included the

entire task, and not selected types of trials (e.g., avoidance trials), activation in the striatum bilaterally was observed to correlate with the seed amygdala ROI. Specifically, ROIs in the left ($x, y, z = -7, 15, 5$; 1615 voxels) and right ($x, y, z = 7, 9, 3$; 903 voxels) ventral caudate nucleus were observed in this analysis (**Figure 5A**), with some degree of overlap with the striatum ROI previously defined by the general analysis (**Figure 5B**).

Finally, an exploratory analysis was performed comparing AV+ and AV- trials, which investigated brain regions distinctly associated with learning changes during avoidance trials. Corticostriatal circuits typically involved in reinforcement learning (for review see O'Doherty, 2004; Daw and Doya, 2006; Balleine et al., 2007) were identified in this contrast (**Table 2**). These included ROIs in the dorsal and ventral striatum, along with dorsal (BA 6) and ventromedial (BA 10/32) prefrontal regions. Interestingly, both striatum ROIs showed a pattern of response resembling learning signals, as only the AV+ trials, where learning could occur, elicited a strong BOLD signal as represented by higher mean parameter estimate values.

DISCUSSION

The goal of this study was to explore the neural circuitry underlying active coping of fear in humans using a variant of an aversive conditioning paradigm where conditioned fear could be



diminished by an instrumental action – an avoidance response. Participants first acquired a behavioral response to terminate delivery of a mild shock, and then continued to use this response to refrain from further aversive outcomes. Physiological arousal, as index by SCRs, supported the behavioral evidence of learning as arousal levels were decreased post-learning of an avoidance response. Additionally, instrumental behavior was faster during avoidance trials, compared to trials where a certain outcome was expected (i.e., non-contingent response), potentially indicating higher motivational levels when an opportunity to exert control over an emotional event is present. A contrast of aversive and safe trials identified *a priori* ROIs in both dorsal and ventral striatum along with amygdala, with BOLD signals within the striatum differing between pre- and post-learning of an avoidance response, a measure that correlated with BOLD signals in the amygdala. This was supported by a connectivity analysis using the amygdala as a seed ROI which found correlations with the striatum. As postulated by non-human animal models (Killcross et al., 1997; Everitt et al., 1999; LeDoux and Gorman, 2001; Cardinal et al., 2002), active coping of fear in humans may involve amygdala and striatal interactions as a means of diminishing conditioned fear and exerting control over emotional responses.

The involvement of the striatum in active avoidance has been previously observed in animal studies (Winocur and Mills, 1969; Allen and Davison, 1973; Neill et al., 1974; McCullough et al., 1993; Li et al., 2004). In the context of this human paradigm, the striatum was particularly involved in learning a behavioral action that allowed for control over conditioned fear, highlighting the role

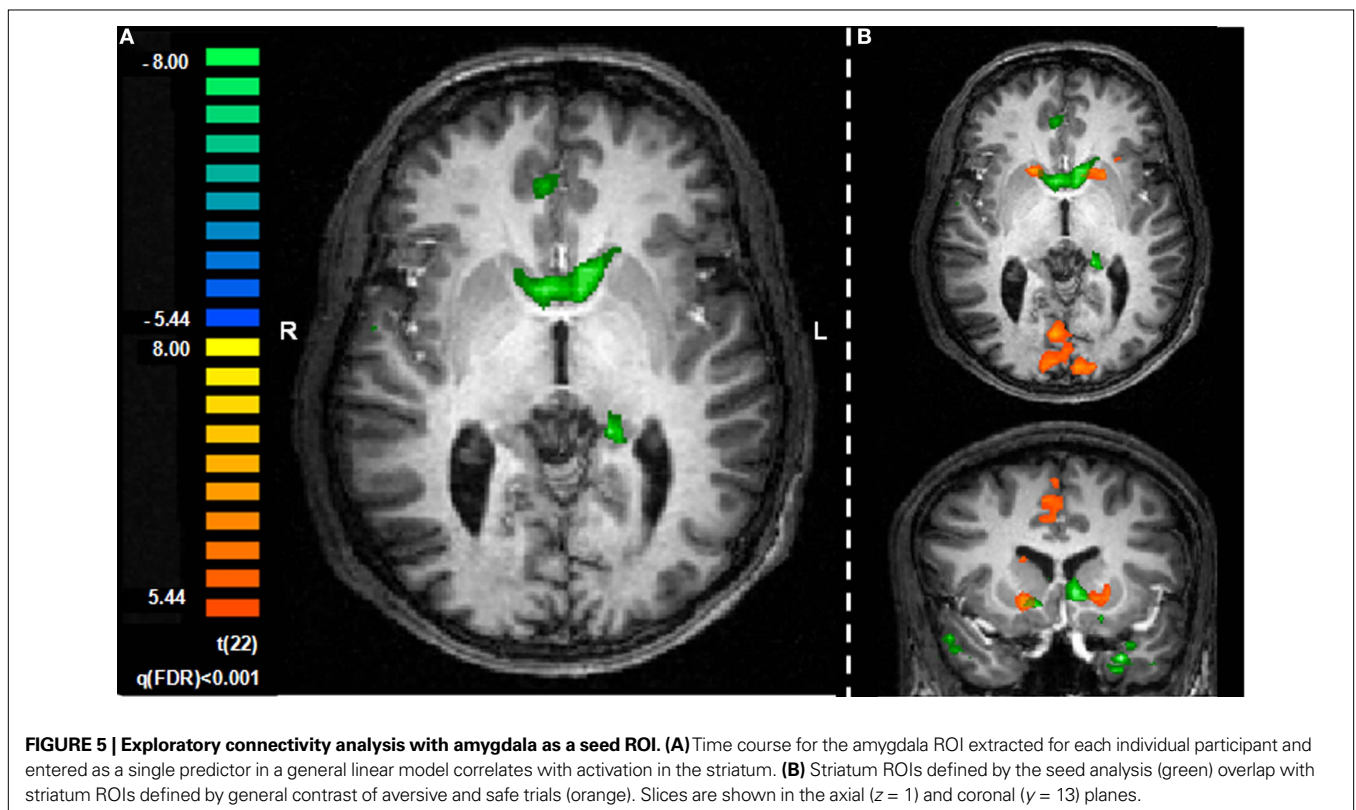


Table 2 | Contrast of AV+ and AV– trials at $p < 0.005$ and contiguity threshold of four voxels.

Region of activation	Brodmann areas	Laterality	Talairach coordinates			#Voxels
			x	y	z	
Dorsomedial frontal gyrus	6	Right	33	1	55	377
Dorsal striatum		Right	16	12	19	130
Ventral striatum		Left	–20	11	1	126
Frontal medial gyrus	10	Right	39	46	1	199
Ventromedial PFC	32/10	Right	25	45	–1	248
Ventromedial PFC	32/10	Right	11	37	–6	578

of the striatum in action-contingency during learning (O'Doherty et al., 2004; Tricomi et al., 2004), while extending it to aversive states. Despite its functional heterogeneity and connectivity to regions such as the amygdala, the human striatum is typically discussed in the context of reward processing (for review see Rangel et al., 2008), although evidence continues to suggest the striatum's involvement during affective learning irrespective of type of reinforcer (positive or negative). This paradigm provides additional support for the involvement of the human striatum in processing negative reinforcement.

The amygdala is a structure linked to aversive processes, particularly the acquisition of fears (for review see Phelps and LeDoux, 2005). Studies in animals also link the amygdala with certain forms of avoidance learning (e.g., Killcross et al., 1997; Machado et al., 2009), or escaping from fear (Amorapanth et al., 2000), leading to the hypothesis that amygdala–striatum interaction could underlie one's ability to actively cope with conditioned fear (LeDoux and Gorman, 2001). Given this a priori hypothesis, we used a lenient threshold previously used by other human fear conditioning studies (e.g., LaBar et al., 1998) to investigate the role of the amygdala in human avoidance learning. Although the results have to be carefully considered given the lenient threshold, we observed a correlation between the time course of amygdala activation during task performance and the striatum, supportive of a potential interaction between the two structures during avoidance learning.

Previous research investigating the regulation of fear in humans has examined passive extinction techniques (Milad and Quirk, 2002; Knight et al., 2004; Phelps et al., 2004) and the use of cognitive strategies (for review see Ochsner and Gross, 2005). In the current paper, we examine the role of active coping strategies, particularly taking an action to avoid a potential threat, to adaptively control fear. A common finding across the three types of techniques is that conditioned fear is diminished, as evidenced by a physiological correlate of fear (SCRs) and decreases in BOLD response in the amygdala. Interestingly, the left amygdala ROI identified in a previous cognitive regulation study of conditioned fear (Delgado et al., 2008b; $x, y, z = -20, 0, -20$) was quite similar to the left amygdala ROI identified in the current study using an avoidance paradigm ($x, y, z = -20, 2, -19$). One potential difference across the techniques, however, is the role of the striatum in the control of fears. Striatum activation has been reported in previous papers examining the control of fear through passive extinction and cognitive strategies techniques (Phelps et al., 2004; Delgado

et al., 2008b), although the particular contrast was a general effect of conditioning. The current findings suggest that the motivation to avoid a negative outcome with an instrumental response engages the striatum even more than just simple conditioning as evidenced by increased activation in the left striatum during AV+ trials, when subjects were learning the avoidance response, compared to CS+ trials when they were simply anticipating an aversive outcome, further highlighting the involvement of the striatum in learning via negative reinforcement.

The paradigm used for this experiment was adapted from previous animal (Amorapanth et al., 2000) and human (Phelps et al., 2004; Delgado et al., 2008b) studies of aversive conditioning. It is a simple task that has distinct advantages for studying a complex process such as avoidance learning. The inclusion of separate CS and response phase, for instance, allows probing of neural responses to the initial representation of the CS without the potential motor and motor preparation confounds associated with this instrumental procedure. This paradigm can also be adapted to study avoidance learning with secondary reinforcers (e.g., money; see Kim et al., 2006), comparisons between reinforcers of different valence (positive and negative reinforcement), and varying levels of probability or complexity of avoidance response (e.g., manipulation of effort required to successfully avoid negative outcome). This avoidance paradigm also has its limitations, however, such as the minimal amount of trials experienced by a participant per condition (6), which required a fixed reinforcement schedule. The task length was designed to limit the amount of shocks administered and, due to piloting, provide an ideal time window where participants felt that they could indeed be successful. It is also possible that some type of habituation can occur in this design as AV+ and AV– trials are separated in time within a scanning session. An argument against habituation being an explanation for the observed results, however, is that individuals who failed to learn the task did not show the differential responses in the striatum when comparing AV+ and AV– trials, which was characteristic of successful task learning (see Supplementary Materials).

In summary, we used a variant of a fear conditioning study where participants had a chance to avoid a negative outcome with an instrumental behavior. Consistent with animal models (e.g., LeDoux and Gorman, 2001), we found amygdala–striatal interactions in humans potentially underlying avoidance learning and exerting control over conditioned fears. Future studies will probe how human mechanisms of affective learning through negative

reinforcers (i.e., avoidance learning) compare to learning through positive reinforcers (i.e., approach learning) to further understand the range of involvement of regions such as the striatum in affective learning, behavioral control and decision-making.

ACKNOWLEDGMENTS

This work was supported by James S. McDonnell foundation grant to Elizabeth A. Phelps, the Beatrice and Samuel A. Seaver

Foundation and a National Institute on Drug Abuse grant to Mauricio R. Delgado (DA022998). We would like to acknowledge Mike Niznikiewicz for technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/behavioralneuroscience/paper/10.3389/neuro.08/033.2009/>

REFERENCES

- Allen, J. D., and Davison, C. S. (1973). Effects of caudate lesions on signaled and nonsignaled Sidman avoidance in the rat. *Behav. Biol.* 8, 239–250.
- Amorapanth, P., LeDoux, J. E., and Nader, K. (2000). Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nat. Neurosci.* 3, 74–79.
- Balleine, B. W., Delgado, M. R., and Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* 27, 8161–8165.
- Buchel, C., Morris, J., Dolan, R. J., and Friston, K. J. (1998). Brain systems mediating aversive conditioning: an event-related fMRI study. *Neuron* 20, 947–957.
- Cain, C., and LeDoux, J. E. (2008). Emotional processing and motivation: in search of brain mechanisms. In *Handbook of Approach and Avoidance Motivation*, A. Elliot, ed. (New York, Psychology Press), pp. 17–34.
- Cardinal, R. N., Parkinson, J. A., Hall, J., and Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* 26, 321–352.
- Daw, N. D., and Doya, K. (2006). The computational neurobiology of learning and reward. *Curr. Opin. Neurobiol.* 16, 199–204.
- Delgado, M. R. (2007). Reward-related responses in the human striatum. *Ann N. Y. Acad. Sci.* 1104, 70–88.
- Delgado, M. R., Li, J., Schiller, D., and Phelps, E. A. (2008a). The role of the striatum in aversive learning and aversive prediction errors. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 363, 3787–3800.
- Delgado, M. R., Nearing, K. I., LeDoux, J. E., and Phelps, E. A. (2008b). Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron* 59, 829–838.
- Delgado, M. R., Stenger, V. A., and Fiez, J. A. (2004). Motivation-dependent responses in the human caudate nucleus. *Cereb. Cortex* 14, 1022–1030.
- Everitt, B. J., Morris, K. A., O'Brien, A., and Robbins, T. W. (1991). The basolateral amygdala–ventral striatal system and conditioned place preference: further evidence of limbic–striatal interactions underlying reward-related processes. *Neuroscience* 42, 1–18.
- Everitt, B. J., Parkinson, J. A., Olmstead, M. C., Arroyo, M., Robledo, P., and Robbins, T. W. (1999). Associative processes in addiction and reward. The role of amygdala–ventral striatal sub-systems. *Ann. N. Y. Acad. Sci.* 877, 412–438.
- Jensen, J., McIntosh, A. R., Crawley, A. P., Mikulis, D. J., Remington, G., and Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron* 40, 1251–1257.
- Kalisch, R., Wiech, K., Critchley, H. D., Seymour, B., O'Doherty, J. P., Oakley, D. A., Allen, P., and Dolan, R. J. (2005). Anxiety reduction through detachment: subjective, physiological, and neural effects. *J. Cogn. Neurosci.* 17, 874–883.
- Killcross, S., Robbins, T. W., and Everitt, B. J. (1997). Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. *Nature* 388, 377–380.
- Kim, H., Shimojo, S., and O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol.* 4, e233. doi:10.1371/journal.pbio.0040233
- Knight, D. C., Smith, C. N., Cheng, D. T., Stein, E. A., and Helmstetter, F. J. (2004). Amygdala and hippocampal activity during acquisition and extinction of human fear conditioning. *Cogn. Affect. Behav. Neurosci.* 4, 317–325.
- Knutson, B., and Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Curr. Opin. Neurol.* 18, 411–417.
- LaBar, K. S., Gatenby, J. C., Gore, J. C., LeDoux, J. E., and Phelps, E. A. (1998). Human amygdala activation during conditioned fear acquisition and extinction: a mixed-trial fMRI study. *Neuron* 20, 937–945.
- LaBar, K. S., LeDoux, J. E., Spencer, D. D., and Phelps, E. A. (1995). Impaired fear conditioning following unilateral temporal lobectomy in humans. *J. Neurosci.* 15, 6846–6855.
- LeDoux, J. E., and Gorman, J. M. (2001). A call to action: overcoming anxiety through active coping. *Am. J. Psychiatry* 158, 1953–1955.
- Li, M., Parkes, J., Fletcher, P. J., and Kapur, S. (2004). Evaluation of the motor initiation hypothesis of APD-induced conditioned avoidance decreases. *Pharmacol. Biochem. Behav.* 78, 811–819.
- Machado, C. J., Kazama, A. M., and Bachevalier, J. (2009). Impact of amygdala, orbital frontal, or hippocampal lesions on threat avoidance and emotional reactivity in nonhuman primates. *Emotion* 9, 147–163.
- McCullough, L. D., Sokolowski, J. D., and Salamone, J. D. (1993). A neurochemical and behavioral investigation of the involvement of nucleus accumbens dopamine in instrumental avoidance. *Neuroscience* 52, 919–925.
- Menon, M., Jensen, J., Vitcu, I., Graff-Guerrero, A., Crawley, A., Smith, M. A., and Kapur, S. (2007). Temporal difference modeling of the blood-oxygen level dependent response during aversive conditioning in humans: effects of dopaminergic modulation. *Biol. Psychiatry* 62, 765–772.
- Milad, M. R., and Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* 420, 70–74.
- Mogenson, G. J., Jones, D. L., and Yim, C. Y. (1980). From motivation to action: functional interface between the limbic system and the motor system. *Prog. Neurobiol.* 14, 69–97.
- Montague, P. R., and Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron* 36, 265–284.
- Neill, D. B., Boggan, W. O., and Grossman, S. P. (1974). Impairment of avoidance performance by intrastratial administration of 6-hydroxydopamine. *Pharmacol. Biochem. Behav.* 2, 97–103.
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304, 452–454.
- O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* 14, 769–776.
- Ochsner, K. N., and Gross, J. J. (2005). The cognitive control of emotion. *Trends Cogn. Sci* 9, 242–249.
- Phelps, E. A., Delgado, M. R., Nearing, K. I., and LeDoux, J. E. (2004). Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* 43, 897–905.
- Phelps, E. A., and LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* 48, 175–187.
- Rangel, A., Camerer, C., and Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556.
- Robbins, T. W., Cador, M., Taylor, J. R., and Everitt, B. J. (1989). Limbic–striatal interactions in reward-related processes. *Neurosci. Biobehav. Rev.* 13, 155–162.
- Seymour, B., Daw, N., Dayan, P., Singer, T., and Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. *J. Neurosci.* 27, 4826–4831.
- Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., Friston, K. J., and Frackowiak, R. S. (2004). Temporal difference models describe higher-order learning in humans. *Nature* 429, 664–667.
- Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the*

- Human Brain: An Approach to Medical Cerebral Imaging. Stuttgart, NY, Medical Publishers.
- Tricomi, E. M., Delgado, M. R., and Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron* 41, 281–292.
- Winocur, G., and Mills, J. A. (1969). Effects of caudate lesions on avoidance behavior in rats. *J. Comp. Physiol. Psychol.* 68, 552–557.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 10 June 2009; paper pending published: 28 July 2009; accepted: 12 September 2009; published online: 01 October 2009.
- Citation: Delgado MR, Jou RL, LeDoux JE and Phelps EA (2009) Avoiding negative outcomes: tracking the mechanisms of avoidance learning in humans during fear conditioning. *Front. Behav. Neurosci.* 3:33. doi: 10.3389/neuro.08.033.2009
- Copyright © 2009 Delgado, Jou, LeDoux and Phelps. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.