



Neural systems underlying aversive conditioning in humans with primary and secondary reinforcers

Mauricio R. Delgado¹, Rita L. Jou² 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:

Paul Whalen, Dartmouth College, USA
Ben Seymour, Wellcome Trust Centre for Neuroimaging, UK

*Correspondence:

Elizabeth A. Phelps, Department of Psychology, Rutgers University, 6 Washington Place, Room 863, New York, NY 10003, USA.
e-mail: liz.phelps@nyu.edu

Money is a secondary reinforcer commonly used across a range of disciplines in experimental paradigms investigating reward learning and decision-making. The effectiveness of monetary reinforcers during aversive learning and associated neural basis, however, remains a topic of debate. Specifically, it is unclear if the initial acquisition of aversive representations of monetary losses depends on similar neural systems as more traditional aversive conditioning that involves primary reinforcers. This study contrasts the efficacy of a biologically defined primary reinforcer (shock) and a socially defined secondary reinforcer (money) during aversive learning and its associated neural circuitry. During a two-part experiment, participants first played a gambling game where wins and losses were based on performance to gain an experimental bank. Participants were then exposed to two separate aversive conditioning sessions. In one session, a primary reinforcer (mild shock) served as an unconditioned stimulus (US) and was paired with one of two colored squares, the conditioned stimuli (CS+ and CS-, respectively). In another session, a secondary reinforcer (loss of money) served as the US and was paired with one of two different CS. Skin conductance responses were greater for CS+ compared to CS- trials irrespective of type of reinforcer. Neuroimaging results revealed that the striatum, a region typically linked with reward-related processing, was found to be involved in the acquisition of aversive conditioned response irrespective of reinforcer type. In contrast, the amygdala was involved during aversive conditioning with primary reinforcers, as suggested by both an exploratory fMRI analysis and a follow-up case study with a patient with bilateral amygdala damage. Taken together, these results suggest that learning about potential monetary losses may depend on reinforcement learning related systems, rather than on typical structures involved in more biologically based fears.

Keywords: fear conditioning, striatum, amygdala, reinforcement, aversive learning, reward, insula

INTRODUCTION

Monetary rewards are a common reinforcer used in experimental paradigms across a range of disciplines, from behavioral economics to neuropsychological investigations of learning (e.g., Knutson et al., 2003; Delgado et al., 2006; Vohs et al., 2006). Money is a secondary reinforcer (i.e., reinforcers which acquire their properties through association with a primary reinforcer) that in many circumstances within human society could have similar or even stronger effects on behavior than more well characterized primary reinforcers (i.e., reinforcers that are innate to the organism and elicit a reaction) such as liquids and food. The use of monetary reinforcers are of particular interest in experiments that probe the neural correlates of learning and decision-making, since the value of money can be positive or negative depending on the context in which it is presented. Across such studies, the human striatum has been identified as a key region involved in reward-related processing that facilitates reward learning and goal-directed behaviors (Montague and Berns, 2002; Knutson and Cooper, 2005; Delgado, 2007; Rangel et al., 2008). Less is known about the effectiveness of money during aversive processing and its neural basis, particularly aversive conditioning in humans, which has mostly relied on primary reinforcers such as shock and found to be dependent on the

integrity of the amygdala (for review see Phelps and LeDoux, 2005). Although potential monetary losses can modulate decision-making under risk (Kahneman and Tversky, 1979), it is unclear if the initial acquisition of aversive representations of monetary losses depends on overlapping systems as more traditional aversive conditioning that involves primary reinforcers. The goal of this study is to provide a direct comparison between a biologically defined primary reinforcer (i.e., shock) and a socially defined secondary reinforcer (i.e., money) and their respective influences in the neural circuits and expression of aversive learning.

The human striatum has been linked to reward-related learning with both primary (e.g., juice; O'Doherty et al., 2004) and secondary (e.g., money; Kirsch et al., 2003) reward in several investigations where either one type of reinforcer or another are presented (for review see Knutson and Cooper, 2005; Delgado, 2007). More recently, blood oxygenation level dependent (BOLD) responses in the dorsal striatum have been shown to correlate with prediction errors in a task involving multiple types of reward presented within the same experiment (juice and money reinforcers; Valentin and O'Doherty, 2009). With respect to aversive conditioning, striatum BOLD signals have been found to correlate with learning in separate tasks that using either primary (shock) or secondary (money) reinforcers (For review

see Delgado et al., 2008). Thus, it is plausible that the human striatum may be involved in the acquisition of a conditioned response irrespective of type of reinforcer. The direct comparison of striatum signals within aversive conditioning with multiple aversive unconditioned stimuli, however, has not yet been investigated.

Aversive conditioning studies in humans typically use primary reinforcers and depend on the integrity of the amygdala (for review see Phelps and LeDoux, 2005). In contrast, the use of monetary reinforcers in aversive paradigms has yielded less consistent results. For instance, some fMRI studies have reported changes in amygdala activation in response to, or expectation of, monetary losses (e.g., Breiter et al., 2001; Yacubian et al., 2006; Smith et al., 2009), while others have failed to do so (e.g., Seymour et al., 2007; see Delgado et al., 2008 for review). These findings are in accordance with neuropsychological investigations of risky decision-making involving monetary losses which have also reached mixed results. In such studies, amygdala lesions have been shown to affect loss aversion (De Martino et al., 2010), and lead to a lack of anticipatory skin conductance responses (SCRs) in a risky gambling task (Bechara et al., 1999), while sparing biases to increase risk seeking behaviors when monetary gambles are framed as losses (Talmi et al., 2009). One way to potentially understand the mixed contributions of monetary reinforcers in aversive contexts is to probe the use of monetary loss as an unconditioned stimulus (US) during a purely aversive conditioning paradigm, with the goal of understanding if conditioning via monetary loss will depend on similar mechanisms used for the acquisition of fears derived from more biologically meaningful stimuli such as shocks.

We conducted an event-related fMRI study to investigate common and distinct neural substrates underlying aversive conditioning with primary and secondary reinforcers. In this experiment, participants were first instructed to play a gambling game where they could win or lose money based on their performance. The purpose of the game was to give participants a monetary endowment, ensuring that each participant had an experimental bank. Immediately following the gambling game, participants were then subjected to two separate aversive conditioning sessions. In one session, a primary reinforcer (i.e., mild shock to the wrist) served as an US and was paired with one of two colored squares, the conditioned stimuli (sCS+ and sCS-, respectively). In another session, a secondary reinforcer (i.e., loss of money, -\$6.00, which was extracted from their experimental bank) served as the US and was paired with one of two different colored squares. This design allowed for the extraction of a conditioned response for each type of reinforcer in its separate learning context, thus allowing for the isolation of the independent effect of each reinforcer on affective learning within an individual (Delgado et al., 2006).

MATERIALS AND METHODS

PARTICIPANTS

Thirty-two participants were enrolled in this study (19 F/13 M; mean age 22.81, SD = 3.58). Participants' inclusion in final data analysis for neuroimaging purposes was dependent on their behavioral performance *a priori*, that is, their ability to demonstrate successful conditioning with both primary and secondary reinforcers as assessed by SCR (see Physiological Set-up, Assessment, and Behavioral Analysis). Using this criteria, data from 6 participants

was not analyzed because of a lack of behavioral responses. An additional six participants only showed conditioning during the session with shock, while five participants only showed responses during the conditioning session with monetary loss. Thus, final analysis was performed on 15 participants (7 F/8 M; mean age 22.13, SD = 3.09) who showed evidence of affective learning during both primary and secondary aversive conditioning sessions. Participants responded to posted advertisement and all participants gave informed consent. The experiments were approved by the University Committee on Activities Involving Human Subjects.

PROCEDURE

The experiment consisted of three experimental blocks (Figure 1). First, participants were exposed to a gambling session (adapted from Delgado et al., 2000) in order to acquire a financial endowment, or an "experimental bank." Participants were then involved in two separate aversive conditioning sessions (adapted from Delgado et al., 2006) which were counterbalanced with respect to order of presentation across participants. In one block, referred to as the primary session due to the nature of the reinforcer, the US was a mild shock to the wrist and resembled traditional aversive conditioning human paradigms (e.g., Phelps et al., 2004). In another block, referred to as secondary session, a monetary loss served as the US and would be extracted from their experimental bank. In the gambling session, participants were told they were playing a "card-guessing" game, where the objective was to determine if the value of a given card was higher or lower than the number five. During each trial, a question mark was presented in the center of the "card," indicating that participants had 2 s to make a response. Using a MRI compatible response unit, participants made a 50/50 choice regarding the potential outcome of the trial. The outcome was either higher (6, 7, 8, 9) or lower (1, 2, 3, 4) than five. The outcome was then displayed for 500 ms, followed by a feedback arrow (which indicated positive or negative feedback) for another 500 ms and an inter-trial interval of 13 s before the onset of the next trial. A correct guess led to the display of a green upward arrow indicating a monetary reward of \$4.00 (reward trials), while an incorrect guess led to the display of a red downward

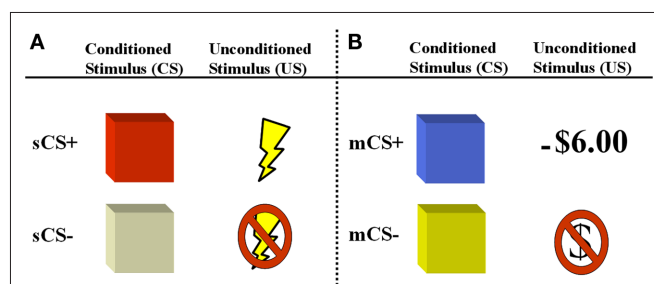


FIGURE 1 | Depiction of aversive conditioning components of experimental paradigm. Participants are presented with two counterbalanced aversive conditioning sessions following a gambling game where they earn a monetary endowment. **(A)** In the first session, the unconditioned stimulus is a mild electric shock (primary reinforcer) which is paired with a colored square (sCS+). **(B)** In the second session, the unconditioned stimulus is a monetary loss (-\$6.00), which is paired with a different colored square (mCS+) and detracted from the total sum earned during the gambling game.

arrow indication a monetary loss of $-\$2.00$ (punishment trials). Each trial was 16 s and participants played one block of the game containing 17 trials for each condition (reward and punishment). Unbeknownst to participants, the outcomes were predetermined ensuring a 50% reinforcement rate and a fixed profit across participants. Participants were initially told they were guaranteed \$25 for performance in the scanner and that anything they would earn in the game was theirs to keep. At the end of the gambling session, a screen appeared congratulating the participant for their total earnings of \$59.00 (\$25 guaranteed amount plus the additional sum of \$34 earned during the game) and informing them that the second part was about to start.

Following the gambling game, participants were exposed to two aversive conditioning sessions with either shock or monetary reinforcers (Figure 1). One session was referred to as primary session because it involved a primary reinforcer (shock). The other was referred to as secondary session because it involved a secondary reinforcer (money). In the primary session, participants were presented with two colored squares (e.g., red and opaque) which served as the CS. Both CS were presented for 6 s, followed by a 12-s inter-trial interval. The US was a mild shock to the wrist, which lasted 200 ms and co-terminated with the CS. In this partial reinforcement design, one colored square (e.g., red) was paired with the shock (CS+) on about 33% of the CS+ trials, while another colored square (e.g., opaque) was never paired with the US (CS-). Participants were instructed that they would see different colored squares and occasionally receive a mild shock. Participants were not told about the contingencies and had to demonstrate successful affective learning (as assessed by SCRs) to be included in the final analysis. There were 30 total trials broken down into 12 CS- trials and 18 CS+ trials, of which 6 were paired with the US.

The secondary session was similar to the primary one, except that the US was a monetary loss. During this session, participants were exposed to two different colored squares (e.g., blue and yellow) which served as the CS. One colored square (e.g., blue) was paired with the monetary loss (CS+) on about 33% of the CS+ trials, while another colored square (e.g., opaque) was never paired with the US (CS-). The monetary loss was depicted by the symbol $-\$6.00$ written in red font and projected inside the square for the last 500 ms. Participants were instructed that they would see different colored squares and occasionally an additional $-\$6.00$ sign indicating that \$6.00 were to be deducted from their "experimental bank" acquired in the gambling session. Participants were not told about the contingencies and had to demonstrate successful affective learning (as assessed by SCRs) to be included in the final analysis. There were 30 total trials broken down into 12 CS- trials and 18 CS+ trials, of which 6 were paired with the US. Finally, the monetary penalties accumulated in the aversive conditioning session resulted in a total loss of \$36.00. To ensure that each participant was paid \$60.00 in compensation following post-experimental questionnaires and debriefing, participants performed a final round of the gambling game (with similar structure to the first game).

Delivery of the US varied according to the type of aversive conditioning session. In the secondary session, the monetary loss was conveyed visually. In the primary session, the mild shock was administered through a stimulating bar electrode attached with a Velcro strap to the right wrist. A Grass Medical Instruments

stimulator charged by a stabilized current was used. The level of shock was set by the participants via a work up procedure that ensured the shocks were "uncomfortable," but not painful. Within this procedure, participants were first given a mild shock (10 V, 200 ms, 50 pulses/s) and gradually increased until the participant signaled so (maximum level of 60 V).

The order of the aversive conditioning sessions, as well as the color of the squares across all four CSs, was counterbalanced across participants. At the end of the experiment, participants filled out post-experimental questionnaires that assessed subjective feelings of intensity and valence toward all CSs. Specifically, participants were given a seven point likert scale and asked how intense their emotion or experience was upon seeing the particular colored square (seven being the most intense) and how bad a colored square was (seven being the worst). Participants were also given a questionnaire containing several gambles with the purpose of assessing individual differences in risk preferences (Holt and Laury, 2002).

PHYSIOLOGICAL SET-UP, ASSESSMENT, AND BEHAVIORAL ANALYSIS

Skin conductance responses were acquired from the participant's middle phalanges of the second and third fingers in the left hand via shielded Ag-AgCl electrodes which were grounded through an RF filter panel. Data acquisition was performed with a BIOPAC systems skin conductance module and AcqKnowledge software was used to analyze SCR waveforms. The level of SCR response was assessed as the base to peak difference for an increase in the 0.5 to 4.5-s window following the onset of a CS, the blue or yellow square (see LaBar et al., 1995). A minimum response criterion of $0.02 \mu\text{S}$ was used with lower responses scored as 0. Responses were square-root transformed prior to statistical analysis to reduce skewness (LaBar et al., 1998). Responses that were three SD from the individual participant's mean responses were excluded due to concerns of excessive motion. Acquired SCRs through the two aversive conditioning sessions were then averaged per participant, per type of trial (e.g., CS+, CS-). Trials in which the CS+ was paired with a shock or monetary loss were separated from analysis so only SCRs to the CS+ (without US) were included.

A repeated measures ANOVA with type of aversive conditioning session (primary or secondary reinforcer) and type of CS (CS+ and CS-) as within subjects factor was then conducted. Two-tailed paired *t*-tests were used to compare activity of CS+ versus CS- trials within session *post hoc* to demonstrate effective conditioning within a specific aversive conditioning session. Participants' inclusion in final data analysis was dependent on their behavioral performance, that is, their ability to demonstrate successful conditioning with both primary and secondary reinforcers as assessed by SCRs. More specifically, participants had to show a greater response for CS+ compared to CS- trials during both sessions. Fifteen participants met this criterion and were included in the final analysis.

Additional behavioral analysis was conducted by scoring the subjective ratings of intensity and valence across type of session and type of CS using a repeated measures ANOVA and *post hoc* two-tailed paired *t*-tests. Analysis of the gambling session was limited since (a) the main purpose of the gambling session was to allow the participant to earn an experimental bank and (b) results for the card-guessing game have been previously published with respect to

neuroimaging (for review see Delgado, 2007) and SCR (Delgado et al., 2006). Nevertheless, SCRs were collected and analyzed for both reward and punishment trials using one sample *t*-tests to examine participants' levels of engagement during the gambling session.

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 one-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 were 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 aversive learning functional data using a general linear model (GLM) on 15 participants. There were six different regressors, including four at the level of the CS that covered the primary session (sCS+ and sCS-) and the secondary session (mCS+ and mCS-) as well as two at US onset (shock or loss of money). The main statistical map of interest was a conjunction analysis that investigated voxels commonly recruited during aversive conditioning with primary and secondary reinforcers. The conjunction analysis tests the conjunction null hypothesis where the requirement is simply that all comparisons or included contrasts are individually significant. Specifically, this conjunction analysis contrasted all CS+ trials with CS- trials for both primary and secondary sessions separately and then produced a statistical parametric map (SPM) that represented commonly activated voxels between the two contrasts. This map was thresholded at $p < 0.005$ and used a cluster threshold with an extent of eight, suggesting that only clusters which are associated with a cluster level false positive rate of $\alpha = 0.05$ are sufficient to remain in the analysis (Forman et al., 1995; Goebel et al., 2006). Mean beta weights were extracted from whole ROIs identified in this contrast for *post hoc* analysis and graphing for visualization purposes. Differences within sessions were assessed by probing the interaction of CS (CS+ and CS-) and session (primary and secondary) using the same threshold criteria and correction method. Finally, an exploratory analysis was conducted to functionally identify an amygdala ROI using the contrast of sCS+ and sCS- during early acquisition of fear and an uncorrected threshold of $p < 0.01$.

RESULTS

PHYSIOLOGICAL ASSESSMENT OF GAMBLING SESSION

Skin conductance responses were acquired during the gambling session for reward and punishment trials to assess the overall level of engagement by participants in the card-guessing game.

There were no differences between reward ($M = 0.38$, $SE = 0.08$) and punishment ($M = 0.40$, $SE = 0.09$) trials during the session [$t(14) = 0.77$, $p = 0.45$]. Participants were motivated during performance in the gambling game, however, as assessed by one sample *t*-tests during both reward [$t(14) = 5.06$, $p < 0.0005$] and punishment [$t(14) = 4.68$, $p < 0.0005$] trials.

PHYSIOLOGICAL ASSESSMENT OF AVERSIVE CONDITIONING SESSIONS

A repeated measures ANOVA was conducted with the SCR data to measure the success of aversive conditioning with both shock and monetary reinforcers in a group of 15 participants that showed physiological responses during both aversive conditioning sessions (Figure 2A). A main effect of CS (CS+ and CS-) was observed, suggesting that participants were able to learn the contingencies irrespective of type of reinforcer used [$F(1, 14) = 34.35$, $p < 0.0001$]. This is supported by *post hoc t*-tests showing differential responses between CS+ and CS- trials during both primary [$t(14) = 3.88$, $p < 0.005$] and secondary [$t(14) = 4.83$, $p < 0.0005$] sessions. Given the nature of the US properties and delivery, a main effect of session (primary and secondary) was also observed [$F(1, 14) = 7.61$, $p < 0.05$]. Importantly, however, no interaction was apparent [$F(1, 14) = 0.18$, $p < 0.68$; Figure 2B], suggesting that the conditioned response was strong irrespective of type of session.

SUBJECTIVE RATINGS

Participants were administered likert scale questionnaires at the end of the experiment assessing their subjective perception of both the intensity and valence of the four CSs (sCS+, sCS-, mCS+, and mCS-). For intensity ratings, a repeated measures ANOVA revealed a main effect of CS [$F(1, 14) = 105.64$, $p < 0.0001$], with *post hoc t*-tests confirming differences during both primary [$t(14) = 9.13$, $p < 0.0005$] and secondary [$t(14) = 8.26$, $p < 0.0005$] sessions. No main effect of session [$F(1, 14) = 0.01$, $p = 0.95$], or interaction [$F(1, 14) = 3.06$, $p = 0.10$], were observed. For valence ratings, a repeated measures ANOVA revealed a main effect of CS [$F(1, 14) = 12.22$, $p < 0.005$], with *post hoc t*-tests showing differences during both the primary [$t(14) = 3.85$, $p < 0.005$] and secondary [$t(14) = 3.04$, $p < 0.01$] sessions. No main effect of session [$F(1, 14) = 0.22$, $p = 0.65$], or interactions [$F(1, 14) = 3.20$, $p = 0.10$], were observed.

NEUROIMAGING RESULTS: SIMILARITIES IN NEURAL CIRCUITRY

The main statistical map of interest was a conjunction analysis that investigated voxels commonly recruited during aversive conditioning with primary and secondary reinforcers. Specifically, this contrast looked for voxels activated by a CS+ – CS- contrast which overlapped across both types of sessions. This contrast led to the identification of several regions (Table 1), including the medial frontal gyrus (BA 6), anterior insula, and the striatum bilaterally showing greater responses during trials that predicted a potentially aversive outcome (CS+ trials). Of particular interest was the activation of the striatum, a region typically involved in reward-related processing, which was recruited during aversive learning with both primary and secondary reinforcers. Mean beta weights extracted from the striatum ROIs revealed no interactions between type of session (primary or secondary) and CS (CS+, CS-) in both the

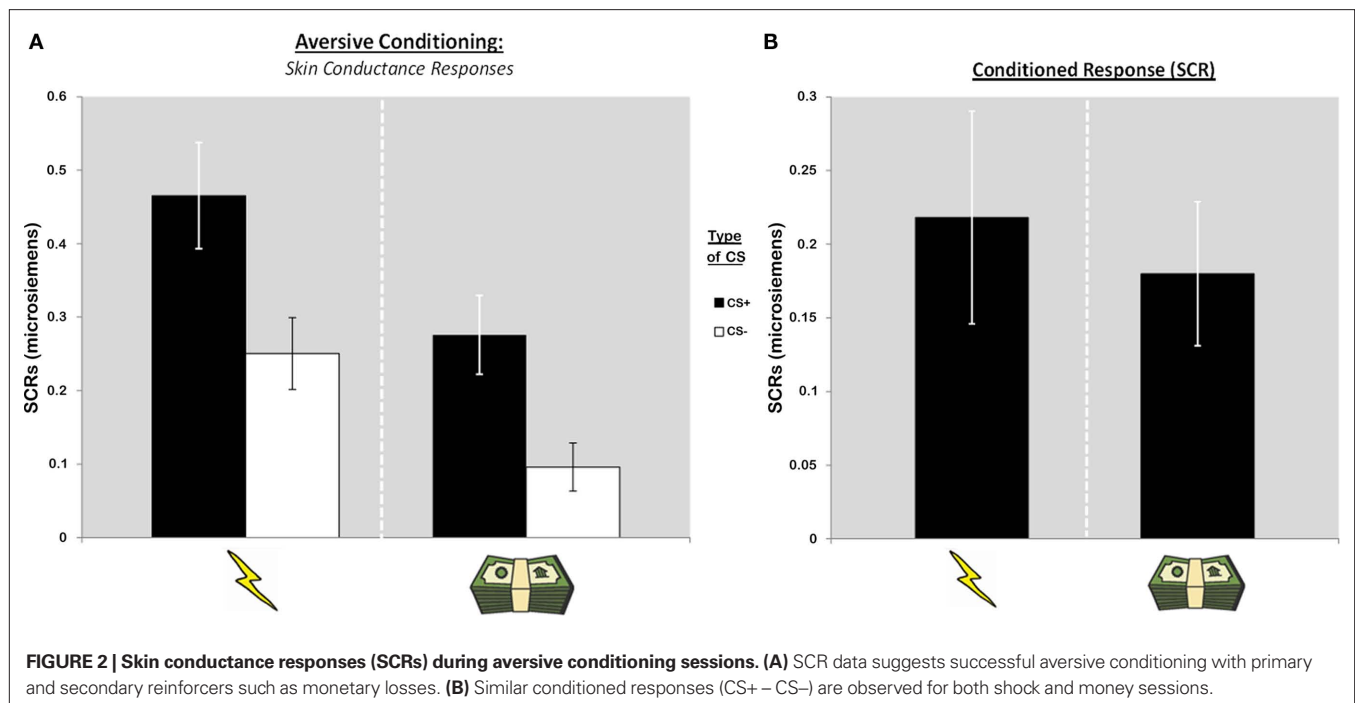


FIGURE 2 | Skin conductance responses (SCRs) during aversive conditioning sessions. (A) SCR data suggests successful aversive conditioning with primary and secondary reinforcers such as monetary losses. **(B)** Similar conditioned responses (CS+ – CS–) are observed for both shock and money sessions.

Table 1 | Conjunction analysis investigating voxels commonly recruited during aversive conditioning with primary and secondary reinforcers ($p < 0.005$).

Region of activation	Laterality	Talairach coordinates			Voxels	t-Stat
		x	y	z		
Paracentral lobule (BA 7)	R	7	–34	55	175	–3.51
Medial frontal gyrus (BA 6)	R	7	6	53	298	3.72
Medial frontal gyrus (BA 6)	L	–24	14	47	170	–3.45
Postcentral gyrus	R	39	–20	47	265	–3.54
Precentral gyrus (BA 4)	R	56	–9	26	761	–3.67
Cingulate gyrus (BA 23)	L	–1	–58	15	323	–3.48
Superior temporal gyrus (BA 42)	R	58	–10	10	579	–3.72
Insula	R	36	19	3	1157	3.65
Striatum	R	10	4	5	241	3.67
Striatum	L	–8	3	3	149	3.77

BA, Brodmann area; L, left; R, right.

left ventral striatum ROI [$F(1, 14) = 0.15, p = 0.7$] and the larger right striatum ROI [$F(1, 14) = 1.98, p = 0.18$] which extended from ventral to more dorsal medial striatum.

Interestingly, the differential response between mCS+ and mCS– mean beta weights, that is the conditioned response during the aversive conditioning session with secondary reinforcers (Figure 3), correlated with a measure of risk preference that was acquired outside the scanner (Holt and Laury, 2002). A Pearson's correlation suggested that the greater the conditioning response in the monetary session, the greater the risk aversion in the participant in the right striatum ROI ($r = 0.602, p < 0.05$) which also manifested as a trend approaching significance in the left striatum ROI ($r = 0.496, p = 0.07$). The same correlation for conditioned responses in the aversive conditioning session with primary

reinforcers was not observed in either ROI. No interactions or correlations with individual risk preferences were observed with the medial frontal gyrus (BA 6) and anterior insula ROIs also identified in this analysis.

NEUROIMAGING RESULTS: DIFFERENCES IN NEURAL CIRCUITRY

To examine differences in neural circuitry underlying aversive conditioning with primary and secondary reinforcers, we investigated voxels in the whole-brain that showed an interaction of CS (CS+ and CS–) and session (primary and secondary). This contrast yielded activity in regions such as the cingulate gyrus, anterior and posterior insula and the somatosensory cortex (Table 2). All regions identified by the interaction showed a greater response to sCS+ compared to mCS+.

Given the well characterized role of the amygdala in aversive conditioning with primary reinforcers, we conducted an additional, exploratory, analysis aimed at identifying a functional ROI in the amygdala. Specifically, we performed a contrast of sCS+ and sCS- trials during the early phases of learning (the first half of trials only) using a lenient threshold of $p < 0.01$ uncorrected and probed activity only in the amygdala (Figure 4). Parameter estimates extracted from this ROI revealed no effect of conditioning in the aversive conditioning session with secondary reinforcers [$t(14) = -0.62, p = 0.54$]. Although these results must be taken with caution due to the exploratory nature of this null result, it suggests that in this specific paradigm, the amygdala is not involved in acquiring a conditioned response to monetary losses.

NEUROPSYCHOLOGICAL CASE STUDY: BILATERAL AMYGDALA DAMAGE

A neuropsychological case study of a patient with bilateral amygdala damage was conducted to further investigate the involvement of the amygdala in aversive conditioning with loss of money as an US. Patient SP is a 62-year old, right-handed woman who underwent a temporal lobe resection to alleviate partial seizures originating in the right temporal lobe at age 48. Prior to the surgery, patient SP was also diagnosed with a lesion in her left amygdala (see Phelps et al., 1998 for a detailed description of SP). The patient shows normal general intelligence according to the Wechsler Adult Intelligence Scale (WAIS-R: Verbal IQ = 100; Performance IQ = 92; Full Scale IQ = 97), but characteristic of patients with amygdala lesions, she

Table 2 | Probing differences in neural circuitry underlying aversive conditioning with primary and secondary reinforcers with an interaction of CS (CS+ and CS-) and session (primary and secondary; $p < 0.005$).

Region of activation	Laterality	Talairach coordinates			Voxels	t-Stat
		x	y	z		
Postcentral gyrus (BA 7)	L	-2	-49	67	417	13.01
Precentral gyrus (BA 4)	L	-22	-21	59	249	20.46
Medial frontal gyrus (BA 6)	L	-3	-8	49	426	14.07
Cingulate gyrus (BA 24)	L	-3	3	38	718	14.37
Insula	L	-47	-29	20	356	14.23
Insula	L	-39	-6	4	1157	13.87
Insula	R	40	-1	-4	235	13.47
Uncus (BA 36)	R	15	-9	-27	222	14.67
Cerebellum	R	16	-46	-27	337	15.68

BA, Brodmann area; L, left; R, right.

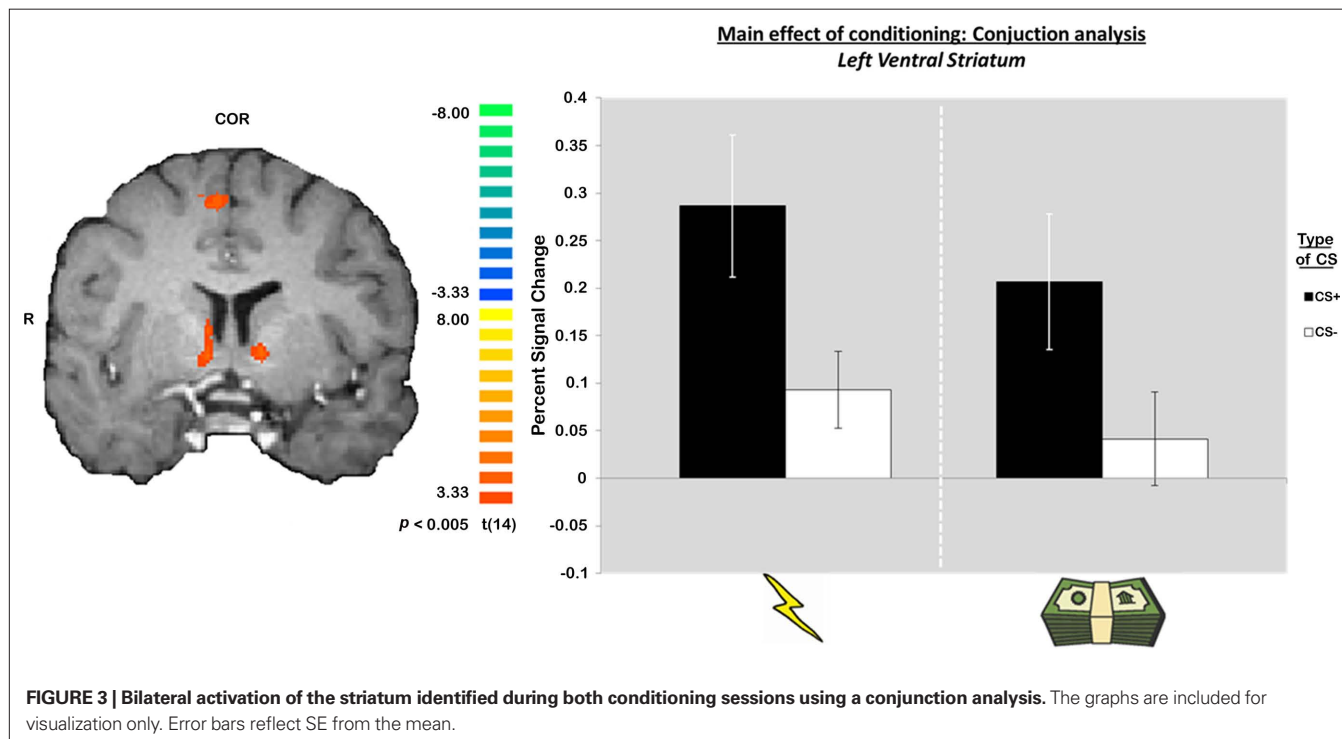
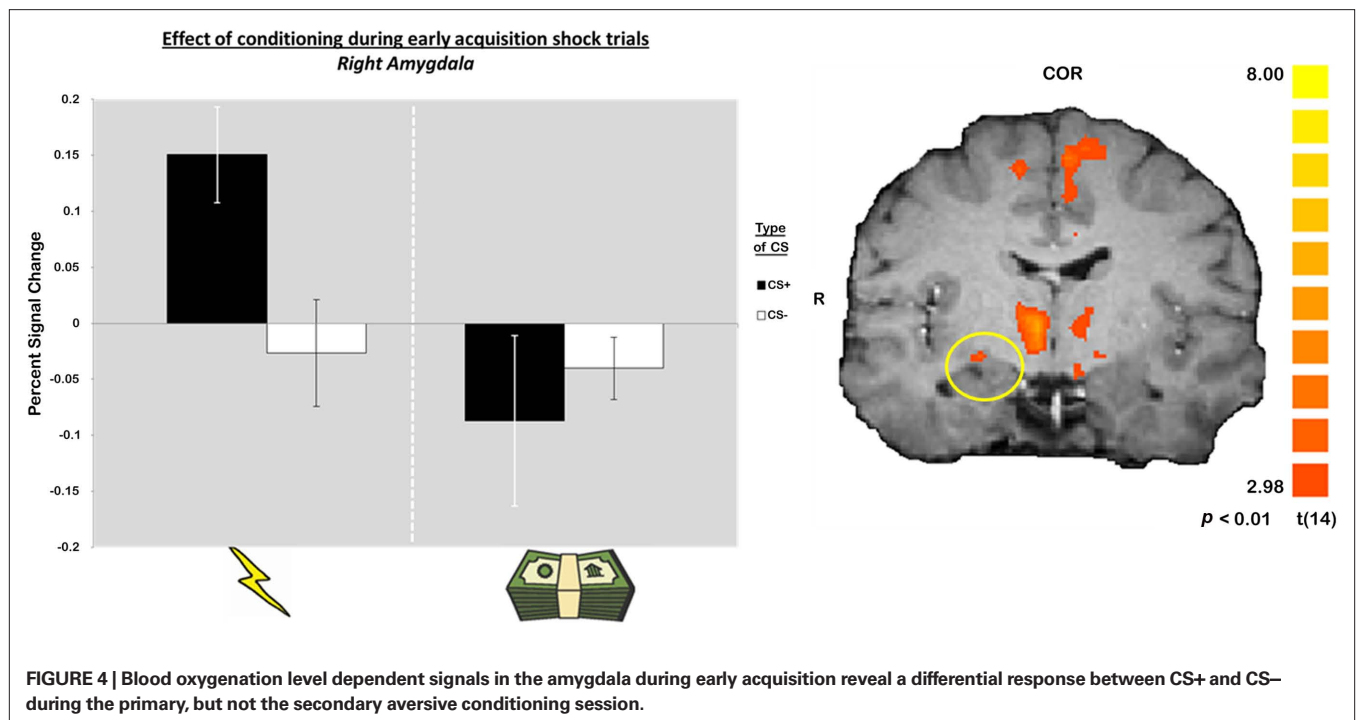


FIGURE 3 | Bilateral activation of the striatum identified during both conditioning sessions using a conjunction analysis. The graphs are included for visualization only. Error bars reflect SE from the mean.



demonstrates difficulties with measures of emotional processing including fear conditioning (e.g., Phelps et al., 1998; Anderson and Phelps, 2002).

Patient SP underwent the same procedure as previously described (Figure 1). First, she performed a gambling session to earn a monetary bank. She was then exposed to two aversive conditioning sessions with the monetary loss session being administered first. As in the fMRI experiment, SCRs were acquired continuously during the conditioning sessions as a measure of sympathetic arousal to the CS presented, while subjective ratings of intensity and valence were collected at the end of the paradigm using a Likert scale from 1 to 7. The level of SCR response was assessed as the base to peak difference for an increase in the 0.5 to 4.5-s window following the onset of a CS (see LaBar et al., 1995), with no minimum response criterion used and lack of responses being scored as 0. Responses were square-root transformed prior to statistical analysis to reduce skewness (LaBar et al., 1998).

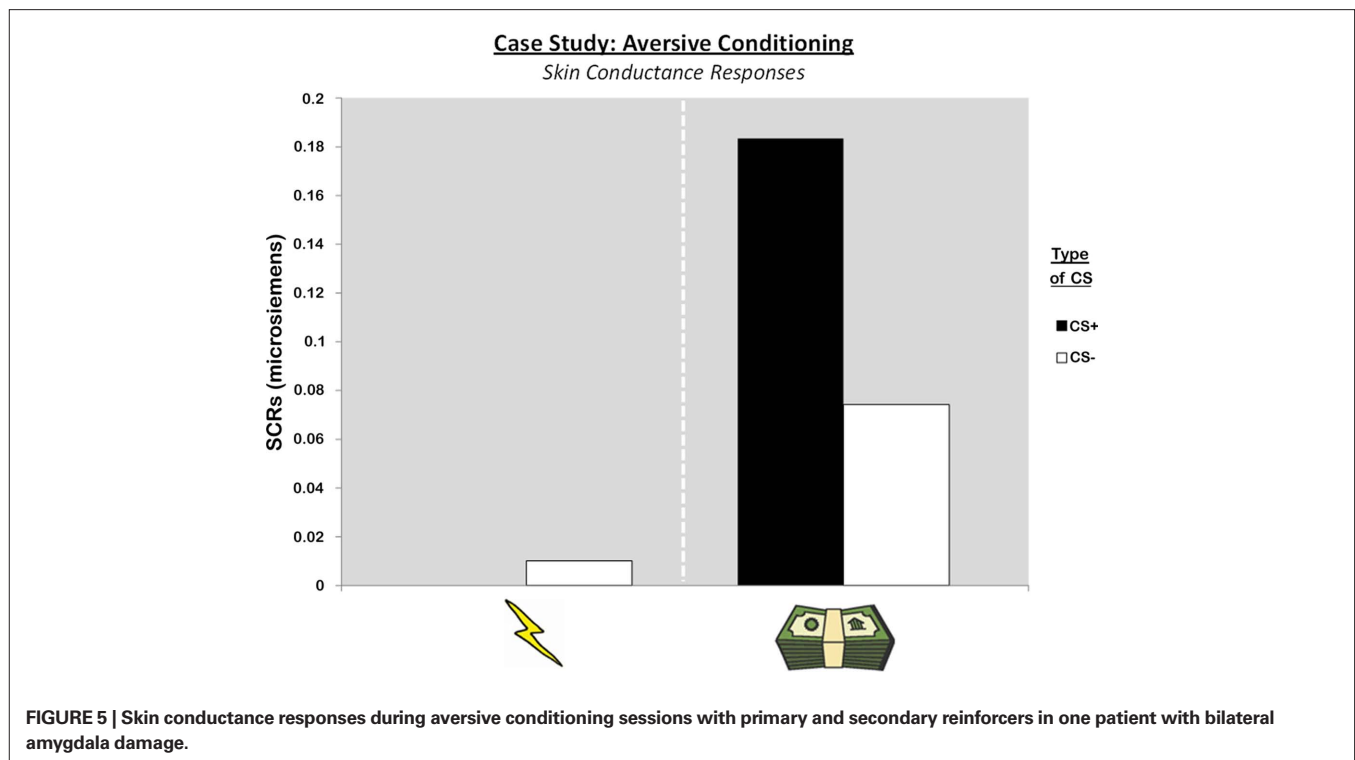
As in previous reports (Phelps et al., 1998), patient SP failed to show a conditioned response during aversive conditioning with a shock US. Her conditioned response (CS+ SCR minus CS- SCR) was slightly less than zero ($M = -0.01$) indicating similar responding to the two CS stimuli. In contrast, when a monetary US was used, SP showed evidence for a stronger SCR to the CS+ than the CS-, indicating the acquisition of a conditioned response ($M = 0.11$; Figure 5). Although it is not possible to conduct reliable tests of significance with this case study, one can contrast SP's data with the younger, neurologically intact participants of our fMRI study. Consistent with previous reports in patients with amygdala damage (Bechara et al., 1995; LaBar et al., 1995; Phelps et al., 1998), SP's conditioned response during aversive conditioning with shock was more than a SD lower than the fMRI participants ($M = 0.21$, $SD = 0.20$). However, SP's conditioned responses on

the monetary US session were well within one SD of the fMRI participants ($M = 0.18$, $SD = 0.14$). These results suggest SP shows impaired aversive conditioning with a shock US, but intact aversive conditioning with a monetary US in this experimental paradigm.

Patient SP was also administered a likert scale questionnaire (1–7) at the end of the experiment assessing her subjective perception of both the intensity and valence of the four CSs (sCS+, sCS-, mCS+, and mCS-). For intensity ratings, Patient SP rated both mCS+ (4) and mCS- (5) higher than sCS+ (2) and sCS- (1). For valence ratings, she rated both mCS- (7) and sCS- (7) higher than mCS+ (1) and sCS+ (1). These results suggest that she “liked” the CS that predicted a safe, rather than negative outcome, while experiencing greater intensity upon seeing CS in the money rather than the shock session.

DISCUSSION

The goal of this study was to provide a direct comparison between a biologically defined primary reinforcer (i.e., shock) and a learned secondary reinforcer (i.e., money) and their respective influences in the neural circuits and expression of learning about fears. Using a modified Pavlovian fear conditioning paradigm (Delgado et al., 2006), participants acquired the value of different CS during separate learning sessions that used a primary or secondary reinforcer which were subjectively perceived as equally intense. The use of money as an US during aversive conditioning led to the expression of a conditioned response, similar to responses elicited by shock, as measured by SCRs. Irrespective of the type of reinforcer used during aversive learning, the striatum was found to be commonly involved in the acquisition of such conditioned response, suggesting a general role for the striatum in affective learning. In contrast, the amygdala was found to be more involved in aversive conditioning with primary compared to secondary reinforcers both in an



exploratory fMRI analysis and a follow-up case study with a patient with bilateral amygdala damage, suggesting that learning to “fear” a potential monetary loss may not depend on typical structures involved in more biologically based fears.

The human striatum has been identified as a critical structure for reward-related processing (Montague and Berns, 2002; Knutson and Cooper, 2005; Delgado, 2007; Rangel et al., 2008), with BOLD signals correlating with both the anticipation (e.g., Knutson et al., 2003) and receipt (e.g., Delgado et al., 2000) of reward, which is often attributed to a general role in reward-related learning and decision-making (for review see Montague and Berns, 2002; O’Doherty, 2004; Rangel et al., 2008). Some observations of increases in BOLD signals to the anticipation of potentially aversive primary (Jensen et al., 2003) and secondary (Delgado et al., 2008) reinforcers have been observed, while a decrease in BOLD responses is sometimes reported at the receipt of monetary losses (Delgado et al., 2000) that resembles a prediction error signal (McClure et al., 2003; O’Doherty et al., 2003). Similar fMRI responses in the striatum have been observed when processing prediction errors during learning with primary (juice) and secondary (money) reinforcers in the same appetitive task (Valentin and O’Doherty, 2009). The current paper extends these results and demonstrates the involvement of the human striatum during aversive learning with both primary (shock) and secondary (money) reinforcers.

The potential role of the human striatum in aversive learning, a region typically associated with reward processing, is unclear. There is evidence of striatum signals correlating with prediction errors when the context is aversive using both primary (Seymour et al., 2004; Delgado et al., 2008) and secondary reinforcers (Seymour et al., 2007). It is possible that the striatum is involved in general affective learning irrespective of valence (appetitive, aversive) or

type of reinforcer (primary, secondary). It is also possible that distinct regions within the striatum code for such valence differences (see Seymour et al., 2007). Within this context, a potential role for the striatum in aversive learning may be to participate in a circuitry responsible for updating value representations in order to change learned fears (Schiller and Delgado, 2010) and actively cope with the aversive context (LeDoux and Gorman, 2001).

Compared to other studies investigating aversive learning with secondary reinforcers (Kim et al., 2006; Seymour et al., 2007; Schlund and Cataldo, 2010), one key feature of this paradigm is the use of an experimental bank prior to the aversive conditioning session (Delgado et al., 2006; Tom et al., 2007). This manipulation has the potential effect of increasing the significance of the monetary loss, in fact framing the loss as a real negative consequence since it is deducted from the participant’s own endowment. In contrast, in paradigms where participants do not have a sense of earning the monetary endowment, a loss may be experienced as a missed opportunity for a reward or the lesser of two outcomes. In support of this argument, striatum activity has been found to be stronger when participants earn the outcomes (Tricomi et al., 2004; Zink et al., 2004) and further modulated by changes in expected value in the context of a reference point (De Martino et al., 2009). This difference in the framing prior to the aversive conditioning session merits further investigation as a plausible mechanism responsible for differences in striatal responses to CS predicting monetary losses observed across different studies.

How money begins to acquire its conditioned reinforcer properties in humans could be akin to the process of second-order conditioning typically studied in non-human animals. In Pavlovian second-order conditioning, a CS acquires conditioned properties (either positive or negative) due to an association with a first-order

CS which elicits a conditioned response (Holland and Rescorla, 1975). The basolateral nucleus of the amygdala has been identified as a critical structure for the acquisition of a second-order conditioned response (Hatfield et al., 1996; Gewirtz and Davis, 1997). However, this role appears limited to the acquisition of information about the motivational value of the first-order CS rather than the maintenance or expression of information already learned. This is illustrated by lesions of the basolateral nucleus of the amygdala after first-order, but before second-order training having no effect on the expression of second-order conditioned behaviors (Setlow et al., 2002). In the current experiment, money itself may be the first-order conditioned reinforcer, while the aversive conditioning session with money as the US would be an example of second-order conditioning. Since the acquisition of motivational information about the first-order CS, money, is well established, the second-order expression may not be dependent on the integrity of the amygdala in humans.

Several investigations of the human amygdala support the assertion of a vast literature in non-human animals linking the amygdala with Pavlovian fear conditioning when primary reinforcers such as shock are used (for review see Phelps and LeDoux, 2005; Hartley and Phelps, 2010). Although the amygdala has been linked with attaching value to purely social reinforcers such as faces (Davis et al., 2010), there is more uncertainty with respect to the role of the human amygdala and the processing of secondary reinforcers such as money. While some neuroimaging studies find changes in BOLD signals in the amygdala correlating with the expectation or actual receipt of monetary losses (e.g., Breiter et al., 2001; Yacubian et al., 2006; Smith et al., 2009), or to cues that predict monetary losses in an instrumental context where they are avoidable (Schlund and Cataldo, 2010), there have also been null findings (see Delgado et al., 2008 for review), particularly when attempting to isolate regions of the brain involved in loss aversion (Tom et al., 2007). This inconsistency extends to neuropsychological investigations, as patients with bilateral amygdala damage have been shown to be sensitive to monetary loss aversion (De Martino et al., 2010), in contrast to neuroimaging findings (Tom et al., 2007), with these results demonstrating an overall deficit in amygdala patients in decisions under risk (Bechara et al., 1999; Brand et al., 2007). The current study adds

to this literature by suggesting that in a Pavlovian aversive conditioning session analogous to second-order conditioning, the expression of a conditioned response does not appear to depend on the amygdala.

A few important features of the Pavlovian aversive conditioning paradigm with monetary reinforcers may contribute to the discrepancy in findings with respect to the amygdala results. First, unlike tasks that probe loss aversion via active decision-making processes (Bechara et al., 1999; Brand et al., 2007; De Martino et al., 2010), this paradigm involves passive learning of associations which are less salient in human experiments. Second, although this paradigm aimed to equate the intensity of both primary and secondary reinforcers used in this task (as supported by subjective ratings), this was done at a group, not individual level as the monetary loss was uniform across the study. Third, the differences in delivery of the reinforcer could have accounted for much of the observed changes in the amygdala. That is, shock is immediately delivered and perceived through somatosensory pathways. In contrast, monetary loss did not physically take place till the end of the experiment and was conveyed visually. Perhaps the shock could have been accumulated in the experiment creating a “bank of shocks” to be delivered later. However, the intention of the experiment was to try to replicate previous instances of fear conditioning in humans. In addition, it is unclear how delayed reinforcement would affect the acquisition of an aversive response with either type of reinforcer. Finally, it is worth noting that the amygdala ROI was defined in an exploratory fMRI analysis and the neuropsychological investigation involved only one participant. More work is needed to fully understand the role of the amygdala in aversive conditioning with monetary losses, this study suggests important differences in the neural circuitry involved in the acquisition of fear from a biologically defined primary reinforcer (i.e., shock) and a socially defined secondary reinforcer (i.e., money).

ACKNOWLEDGMENTS

This study was funded by a Seaver Foundation grant to NYU's Center for Brain Imaging and James S. McDonnell Foundation grant to Elizabeth A. Phelps. During manuscript preparation, Mauricio R. Delgado was supported by NIDA grant DA027764 and Elizabeth A. Phelps was supported by NIMH grant MH80756.

REFERENCES

- Anderson, A. K., and Phelps, E. A. (2002). Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions. *J. Cogn. Neurosci.* 14, 709–720.
- Bechara, A., Damasio, H., Damasio, A. R., and Lee, G. P. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J. Neurosci.* 19, 5473–5481.
- Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C., and Damasio, A. R. (1995). Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. *Science* 269, 1115–1118.
- Brand, M., Grabenhorst, F., Starcke, K., Vandekerckhove, M. M., and Markowitsch, H. J. (2007). Role of the amygdala in decisions under ambiguity and decisions under risk: evidence from patients with Urbach-Wiethe disease. *Neuropsychologia* 45, 1305–1317.
- Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., and Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30, 619–639.
- Davis, F. C., Johnstone, T., Mazzulla, E. C., Oler, J. A., and Whalen, P. J. (2010). Regional response differences across the human amygdaloid complex during social conditioning. *Cereb. Cortex* 20, 612–621.
- De Martino, B., Camerer, C. F., and Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3788–3792.
- De Martino, B., Kumaran, D., Holt, B., and Dolan, R. J. (2009). The neurobiology of reference-dependent value computation. *J. Neurosci.* 29, 3833–3842.
- Delgado, M. R. (2007). Reward-related responses in the human striatum. *Ann. N. Y. Acad. Sci.* 1104, 70–88.
- Delgado, M. R., Labouliere, C. D., and Phelps, E. A. (2006). Fear of losing money? Aversive conditioning with secondary reinforcers. *Soc. Cogn. Affect. Neurosci.* 1, 250–259.
- Delgado, M. R., Li, J., Schiller, D., and Phelps, E. A. (2008). 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., Nystrom, L. E., Fissell, C., Noll, D. C., and Fiez, J. A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.* 84, 3072–3077.
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., and Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn. Reson. Med.* 33, 636–647.
- Gewirtz, J. C., and Davis, M. (1997). Second-order fear conditioning

- prevented by blocking NMDA receptors in amygdala. *Nature* 388, 471–474.
- Goebel, R., Esposito, F., and Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum. Brain Mapp.* 27, 392–401.
- Hartley, C. A., and Phelps, E. A. (2010). Changing fear: the neurocircuitry of emotion regulation. *Neuropsychopharmacology* 35, 136–146.
- Hatfield, T., Han, J. S., Conley, M., Gallagher, M., and Holland, P. (1996). Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *J. Neurosci.* 16, 5256–5265.
- Holland, P. C., and Rescorla, R. A. (1975). Second-order conditioning with food unconditioned stimulus. *J. Comp. Physiol. Psychol.* 88, 459–467.
- Holt, C. A., and Laury, S. K. (2002). Risk aversion and incentive effects. *Am. Econ. Rev.* 92, 1644–1655.
- 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.
- Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–292.
- 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
- Kirsch, P., Schienle, A., Stark, R., Sammer, G., Blecker, C., Walter, B., Ott, U., Burkart, J., and Vaitl, D. (2003). Anticipation of reward in a nonaversive differential conditioning paradigm and the brain reward system: an event-related fMRI study. *Neuroimage* 20, 1086–1095.
- Knutson, B., and Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Curr. Opin. Neurol.* 18, 411–417.
- Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., and Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *Neuroimage* 18, 263–272.
- 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.
- McClure, S. M., Berns, G. S., and Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38, 339–346.
- Montague, P. R., and Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron* 36, 265–284.
- 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.
- O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., and Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* 38, 329–337.
- 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., LaBar, K. S., Anderson, A. K., O'Connor, K. J., Fulbright, R. K., and Spencer, D. D. (1998). Specifying the contributions of the human amygdala to emotional memory: a case study. *Neurocase* 4, 527–540.
- 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.
- Schiller, D., and Delgado, M. R. (2010). Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends Cogn. Sci.* 14, 268–276.
- Schlund, M. W., and Cataldo, M. F. (2010). Amygdala involvement in human avoidance, escape and approach behavior. *Neuroimage* 53, 769–776.
- Setlow, B., Holland, P. C., and Gallagher, M. (2002). Disconnection of the basolateral amygdala complex and nucleus accumbens impairs appetitive pavlovian second-order conditioned responses. *Behav. Neurosci.* 116, 267–275.
- 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.
- Smith, B. W., Mitchell, D. G., Hardin, M. G., Jazbec, S., Fridberg, D., Blair, R. J., and Ernst, M. (2009). Neural substrates of reward magnitude, probability, and risk during a wheel of fortune decision-making task. *Neuroimage* 44, 600–609.
- Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. New York: Thieme Medical Publishers, Inc.
- Talmi, D., Dayan, P., Kiebel, S. J., Frith, C. D., and Dolan, R. J. (2009). How humans integrate the prospects of pain and reward during choice. *J. Neurosci.* 29, 14617–14626.
- Tom, S. M., Fox, C. R., Trepel, C., and Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science* 315, 515–518.
- Tricomi, E. M., Delgado, M. R., and Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron* 41, 281–292.
- Valentin, V. V., and O'Doherty, J. P. (2009). Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. *J. Neurophysiol.* 102, 3384–3391.
- Vohs, K. D., Mead, N. L., and Goode, M. R. (2006). The psychological consequences of money. *Science* 314, 1154–1156.
- Yacubian, J., Glascher, J., Schroeder, K., Sommer, T., Braus, D. F., and Buchel, C. (2006). Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *J. Neurosci.* 26, 9530–9537.
- Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., and Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron* 42, 509–517.

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: 29 December 2010; accepted: 03 May 2011; published online: 20 May 2011.
 Citation: Delgado MR, Jou RL and Phelps EA (2011) Neural systems underlying aversive conditioning in humans with primary and secondary reinforcers. *Front. Neurosci.* 5:71. doi: 10.3389/fnins.2011.00071
 This article was submitted to *Frontiers in Decision Neuroscience, a specialty of Frontiers in Neuroscience*.
 Copyright © 2011 Delgado, Jou and Phelps. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.