Neuron, Vol. 43, 897–905, September 16, 2004, Copyright ©2004 by Cell Press

Extinction Learning in Humans: Role of the Amygdala and vmPFC

Elizabeth A. Phelps,[1,*] Mauricio R. Delgado,[1] Katherine I. Nearing,[1] and Joseph E. LeDoux[2]
[1]Department of Psychology and
[2]Center for Neural Science
New York University
New York, New York 10003

Summary

Understanding how fears are acquired is an important step in translating basic research to the treatment of fear-related disorders. However, understanding how learned fears are diminished may be even more valuable. We explored the neural mechanisms of fear extinction in humans. Studies of extinction in nonhuman animals have focused on two interconnected brain regions: the amygdala and the ventral medial prefrontal cortex (vmPFC). Consistent with animal models suggesting that the amygdala is important for both the acquisition and extinction of conditioned fear, amygdala activation was correlated across subjects with the conditioned response in both acquisition and early extinction. Activation in the vmPFC (subgenual anterior cingulate) was primarily linked to the expression of fear learning during a delayed test of extinction, as might have been expected from studies demonstrating this region is critical for the retention of extinction. These results provide evidence that the mechanisms of extinction learning may be preserved across species.

Introduction

Investigations of the neural systems of fear learning have examined classical fear conditioning as a model paradigm. Using this paradigm, researchers have been able to map the pathways of fear learning from stimulus input to response output (LeDoux, 2002). Studies exploring fear conditioning in humans have largely supported these findings from nonhuman animals (Phelps, 2004). However, understanding how fears are acquired will only partially help in extending these animal models to the treatment of fear-related disorders. Perhaps even more important is determining how these learned fears are diminished or inhibited. Although there are clear links between the neural systems of fear acquisition and expression across species, there is less understanding of the mechanisms of fear extinction. The present study attempts to expand our understanding of the neural mechanisms of fear inhibition across species by using fMRI to examine fear extinction in humans.

In a typical fear-conditioning paradigm, a neutral event such as a tone (the conditioned stimulus, or CS) is paired with an aversive event, such as a shock (the unconditioned stimulus, or US). After several pairings of the tone and shock, the presentation of the tone itself leads to a fear response (the conditioned response, or CR). Extinction occurs when a CS is presented alone, without the US, for a number of trials and eventually the CR is diminished or eliminated. Behavioral studies of extinction suggest that it is not a process of "unlearning" but rather is a process of new learning of fear inhibition. This view of extinction as an active learning process is supported by studies showing that after extinction the CR can return in a number of situations, such as the passage of time (spontaneous recovery), the presentation of the US alone (reinstatement), or if the animal is placed in the context of initial learning (renewal; see Bouton, 2002, for a review). Although animal models of the mechanisms of fear acquisition have been investigated over the last several decades, studies of the mechanisms of extinction learning have only recently started to emerge. Research examining the neural systems of fear extinction in nonhuman animals have focused on two interconnected brain regions: the ventral medial prefrontal cortex (vmPFC) and the amygdala.

The vmPFC was first implicated in fear extinction when Morgan et al. (1993) demonstrated that lesions to this region led to an impairment in extinction. Further lesion studies (Quirk et al., 2000) demonstrated that damage to this region did not result in an impairment of extinction learning overall but rather an impairment in the retention of extinction learning over subsequent days. For example, Quirk et al. reported that lesions of the infralimbic cortex in the vmPFC did not impair extinction learning that occurred immediately following acquisition. However, when the rats were tested a day later, the rats with lesions to the vmPFC demonstrated extensive spontaneous recovery, performing similarly to rats that had no extinction training at all. These results indicate an impairment in the *retention* or recall of extinction learning (but see also Gewirtz et al., 1997). More recently, electrophysiological recordings have helped clarify the role of the vmPFC in fear conditioning and extinction. Milad and Quirk (2002) found that vmPFC neurons respond to a tone CS only during a delayed test of extinction. No effects were found for fear acquisition or within session extinction (but see also Garcia et al., 1999). Furthermore, when the presentation of the tone CS is paired with stimulation to the vmPFC in rats that have not been given extinction training, the expression of conditioned fear is diminished, suggesting a role for the vmPFC in the inhibition of the CR (Milad and Quirk, 2002).

Two recent studies have suggested that the vmPFC may affect fear inhibition by influencing specific subregions within the amygdala. The central nucleus (CE) of the amygdala receives input from the lateral nucleus (LA) and projects to a number of brain regions involved in the physiological expression of conditioned fear. Stimulation of the vmPFC changed the response rate of CE output neurons in response to input from the LA, suggesting that the vmPFC modulates the CE primarily and the expression of the CR (Quirk et al., 2003). Further, stimulation of the vmPFC diminished responsiveness in the basolateral nucleus (BLA) of the amygdala to a tone

*Correspondence: liz.phelps@nyu.edu

CS as well as unconditioned tones (Rosenkranz et al., 2003).

Recent investigations into the precise role of the amygdala in fear extinction have focused on pharmacological manipulations targeting specific neurotransmitter systems. Building on the finding that NMDA receptors are important in the acquisition of fear conditioning, the role of NMDA receptors in extinction has been examined. Falls et al. (1992) administered intraamygdala injections of an NMDA antagonist, AP5, prior to extinction training and found a dose-dependent impairment in the expression of extinction a day later. In addition, an intraamygdala injection of an NMDA agonist, D-cycloserine, enhanced extinction learning (Walker et al., 2002). Although other pharmacological manipulations (e.g., GABA, dopamine, acetylcholine antagonists) have also been shown to influence extinction learning, most of these were delivered systemically, precluding the conclusion that the effects involve the amygdala (see Myers and Davis, 2002, for a review).

The recent findings examining extinction in nonhuman animals indicate that both the amygdala and vmPFC are important regions to investigate further in our efforts to understand the neural mechanisms of extinction. In humans, an fMRI study examined amygdala activation during both acquisition and extinction of conditioned fear and found an amygdala response during both stages of learning (LaBar et al., 1998). However, in this study, rapid extinction occurred (within two trials), perhaps because there was 100% reinforcement during training. More recently, brain imaging studies have started to examine the mechanisms of fear inhibition without conditioning. These studies present negative (fearful) and neutral scenes and ask subjects to attempt to reduce their fear responses to these scenes by focusing on positive or nonemotional aspects of the scene (Oschner et al., 2002; Schaefer, et al., 2002). These studies have reported diminished negative affect during the reappraisal of negative scenes as well as diminished amygdala activation. In addition, the Ochsner et al. study found that activation in a right lateral PFC region was correlated with reappraisal success and a diminished amygdala response. This lateral PFC region may play a role in working memory, executive processing, or the active maintenance of online information (Smith and Jonides, 1999). Although there are no direct projections between this lateral PFC region and the amygdala, this region does project to medial PFC regions that are more directly connected with the amygdala (Amaral, 2002; Groenewegen et al., 1997; Mcdonald et al., 1996).

These early studies in humans are suggestive of roles for the amygdala and PFC in the inhibition of fear, but at the present time, the links between these paradigms and those used with nonhuman animals in fear extinction are not clear. In the present study, we attempt to examine more carefully the relation between the neural mechanisms of fear inhibition in humans and other animals by using fMRI to assess the involvement of the amygdala and medial PFC (mPFC) during extinction learning in humans. In order to assess extinction learning over time, we used a partial reinforcement, simple discrimination fear-conditioning paradigm. A partial reinforcement paradigm was used to slow extinction learning, which occurs rapidly in humans with 100% reinforcement (LaBar et al., 1998). In this paradigm, the CS^+ and CS^- were colored squares (blue and yellow) and the US was a mild shock to the wrist. During acquisition, there were 15 unreinforced trials of each type (CS^+ and CS^-) and another eight CS^+ trials that coterminated with a shock to the wrist. There were two extinction sessions. Day 1 extinction immediately followed acquisition with 15 unreinforced trials of each type. In order to assess the retention of extinction learning, subjects were brought back 24 hr later for day 2 extinction, which consisted of 17 unreinforced presentations of each trial type.

Results

Physiological Assessment of Fear Conditioning and Extinction

Subjects who failed to show acquisition of the CR as indicated by a differential skin conductance response to the CS^+ relative to the CS^- were eliminated from further analysis because it would not be possible to assess extinction learning. All trials on which a shock was delivered were excluded from analysis, since we were only interested in learned responses to the CS. On day 1, the first CS^+ and CS^- trial of each phase was excluded from analysis, since learning had not yet occurred on the first trial of extinction, and in acquisition, although the subjects were instructed of the CS-US contingency, the CS was not paired with shock until the first trial. On day 2, the first three CS^+ and CS^- trials were excluded from the analyses to eliminate responses related to a transient spontaneous recovery in some subjects (see Figure 1). Fourteen trials per phase were included in the final analyses.

In order to assess learning over trials, we defined the first half of the remaining trials of each phase as *early* trials and the last half as *late* trials. Not surprisingly, there was a significantly greater differential SCR response to the CS^+ versus CS^- during acquisition [$t(10) = 6.41$, $p < 0.001$]. There was also a significant differential SCR response during extinction for both the within session or day 1 extinction trials [$t(10) = 6.65$, $p < 0.001$] as well as next day or day 2 extinction trials [$t(10) = 3.85$, $p < 0.01$]. Importantly, the strength of the CR diminished over extinction trials. A comparison of the CR in acquisition versus late trials on day 1 extinction indicated a marginally significant decrease or extinction [$t(10) = 2.08$, $p < 0.07$]. In early extinction day 2, some subjects showed a brief increase of the CR that could have been due to spontaneous recovery, reinstatement, or renewal. However, as extinction continued on day 2 there was a significant decline in the CR from early to late trials on day 2 [$t(10) = 3.06$, $p < 0.05$] as well as significant extinction of the CR from acquisition to the late trials of day 2 [$t(10) = 2.91$, $p < 0.05$].

Group Analysis of fMRI Data

The analysis of imaging data excluded the same trials as the analyses of the physiological data (see section on "Physiological Assessment of Fear Conditioning and Extinction" above). Within the mPFC, there were three regions that showed a significant differential blood oxygenation level-dependent (BOLD) response to the CS^+ versus CS^- (see Figure 2). These regions appeared in

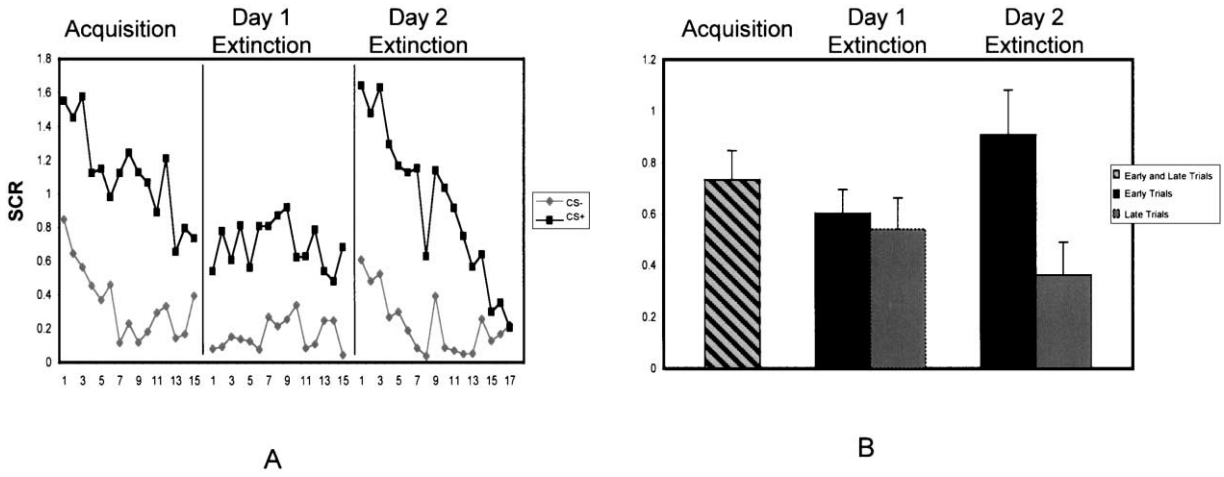

Figure 1. SCR during Acquisition, Day 1 Extinction, and Day 2 Extinction

(A) Square root transformed SCR responses to the CS^+ and CS^- for all trials across acquisition, day 1 extinction, and day 2 extinction.
(B) Conditioned response (mean differential SCR response to the CS^+ minus CS^-) during acquisition, early and late day 1 extinction, and early and late day 2 extinction. Error bars indicate standard error.

all three stages of the study (acquisition, day 1 extinction, day 2 extinction), although the magnitude of the response varied. Two of these regions were in the anterior cingulate, bilaterally. One was a more dorsal region ($x = 0$, $y = 15$, $z = 31$; Talairach and Tournoux, 1988; Brodmann's area 32) that showed a greater BOLD response to the CS^+ relative to the CS^-. The other was in the ventral, subgenual anterior cingulate region (0, 35, −8; BA 32, extending to BA 24 and 25) and showed a greater BOLD response to the CS^- relative to the CS^+. A third vmPFC activation was adjacent to subgenual anterior cingulate in the medial frontal gyrus (0, 54, 6; BA 10) and also showed a greater response to the CS^- relative to the CS^+ (see Figure 2). At the threshold used for group analysis ($p < 0.01$, Bonferoni corrected), there was no significant activation in the amygdala. However, due to our a priori hypothesis of amygdala involvement, the threshold was lowered ($p < 0.005$, uncorrected). There was a relatively greater BOLD response to the CS^+ than CS^- during acquisition (26, −2, −9). This pattern was reversed during day 1 extinction with a greater BOLD response to the CS^- relative to the CS^+ (15, −3, −13 and −17, −3, −13). There was no significant differential amygdala BOLD response during day 2 extinction. In addition to the regions of activation in the areas most analogous to those highlighted in the nonhuman animal research, there were several additional areas of group activation. These activation responses varied in direction in that sometimes the CS^+ BOLD response was greater than the CS^- and other times it was the opposite (see Table 1).

Region of Interest Analysis

Research with nonhuman animals examining the neural systems of extinction have highlighted the roles of the mPFC and amygdala. Because of this, our region of interest (ROI) analyses focused on these areas. There were three ROIs selected within the mPFC which were centered on the Talairach coordinates of the peak group activation response within this region. An additional ROI

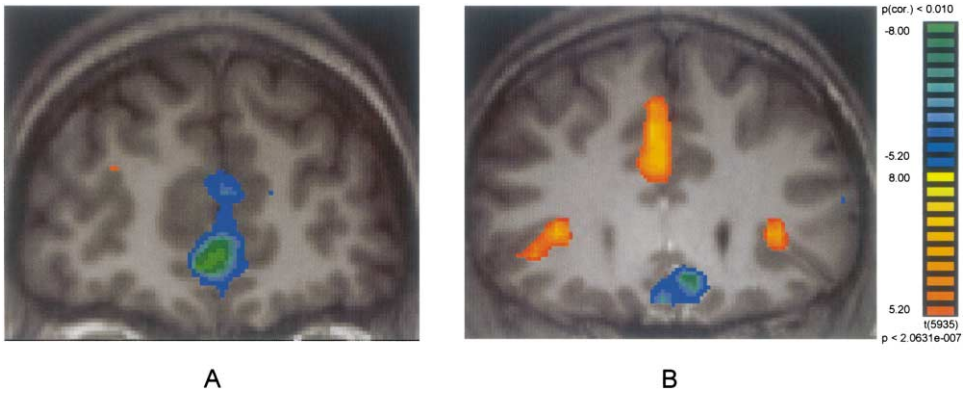

Figure 2. Significant Regions of Activation in the Medial Prefrontal Cortex

(A) The medial frontal gyrus.
(B) The dorsal anterior cingulate (top) and subgenual anterior cingulate (bottom).

Table 1. Group Activation for Acquisition, Extinction Day 1, and Extinction Day 2, $p < 0.01$, Corrected

Acquisition

Area of Activation	Activation	Talairach Coordinates	# of Voxels	Avg t Value
Dorsal cingulate	$CS^+ > CS^-$	0, 14, 33	5641	6.59
Subgenual cingulate	$CS^- > CS^+$	−2, 36, −9	2035	−5.78
Medial frontal gyrus	$CS^- > CS^+$	0, 55, 7	463	−5.66
Insular cortex, left	$CS^+ > CS^-$	−30, 22, 7	5400	6.48
Insular cortex, right	$CS^+ > CS^-$	45, 12, 6	9741	6.86
Posterior cingulate, left	$CS^- > CS^+$	−9, −58, 15	405	−5.46
Posterior cingulate, right	$CS^- > CS^+$	14, −53, 15	121	−5.53
Superior occipital gyrus, left	$CS^- > CS^+$	−44, −71, 30	247	−5.82
Superior occipital gyrus, right	$CS^- > CS^+$	34, −76, 33	134	−5.57
IPL, left	$CS^+ > CS^-$	−53, −39, 27	195	5.53
IPL, right	$CS^+ > CS^-$	58, −39, 28	2210	5.20
Pre-central gyrus, right	$CS^- > CS^+$	43, −7, 48	103	−5.22
Caudate, left	$CS^+ > CS^-$	−8, 2, 9	90	5.49
Caudate, right	$CS^+ > CS^-$	8, 3, 9	552	5.89

Extinction, Day 1

Area of Activation	Activation	Talairach Coordinates	# of Voxels	Avg t Value
Dorsal cingulate	$CS^+ > CS^-$	4, 33, 33	496	5.59
Subgenual cingulate	$CS^- > CS^+$	−2, 38, −3	865	−5.64
Medial frontal gyrus	$CS^- > CS^+$	0, 55, 6	1377	−6.00
Amygdala, right	$CS^- > CS^+$	20, −6, −15	323	−5.78
Insular cortex, left	$CS^+ > CS^-$	−35, 14, 3	2041	5.99
Insular cortex, right	$CS^+ > CS^-$	31, 20, 5	5288	6.52
Posterior cingulate, left	$CS^- > CS^+$	15, −58, 13	5407	−6.62
Posterior cingulate, right	$CS^- > CS^+$	−8, −56, 14	2875	−6.21
Superior occipital gyrus, left	$CS^- > CS^+$	−41, −74, 29	2562	−5.67
Superior occipital gyrus, right	$CS^- > CS^+$	45, −74, 29	1810	−5.86
Cuneus, right	$CS^- > CS^+$	12, −86, 36	5289	−6.13
Cuneus, left	$CS^- > CS^+$	−6, −86, 34	4650	−5.99
Post-central gyrus, left	$CS^- > CS^+$	−52, −22, 44	864	−5.87
Pre-central gyrus, right	$CS^- > CS^+$	60, −10, 27	711	−5.79
Lingual gyrus, right	$CS^- > CS^+$	15, −63, −5	1022	−5.75
Hippocampus, left	$CS^- > CS^+$	−22, −16, −13	419	−6.07
Hippocampus, right	$CS^- > CS^+$	21, −14, −10	127	−5.48
Parahippocampal gyrus, left	$CS^- > CS^+$	−29, −37, −9	3168	−6.29
Parahippocampal gyrus, right	$CS^- > CS^+$	20, −34, −9	1200	−6.09
Caudate, right	$CS^+ > CS^-$	7, 2, 7	307	5.71

Extinction, Day 2

Area of Activation	Activation	Talairach Coordinates	# of Voxels	Avg t Value
Dorsal cingulate	$CS^+ > CS^-$	0, 19, 30	6796	6.47
Subgenual cingulate	$CS^- > CS^+$	−4, 31, −6	938	−5.99
Medial frontal gyrus	$CS^- > CS^+$	−4, 57, 10	23	−5.34
Insular cortex, left	$CS^+ > CS^-$	−28, 21, 3	4210	6.60
Insular cortex, right	$CS^+ > CS^-$	30, 20, 3	11022	7.76
Posterior cingulate, left	$CS^- > CS^+$	−11, −47, 5	1293	−5.64
Posterior cingulate, right	$CS^- > CS^+$	9, −52, 6	671	−5.62
Superior occipital gyrus, Left	$CS^- > CS^+$	−38, −73, 29	4039	−5.83
Superior occipital gyrus, right	$CS^- > CS^+$	36, −81, 21	181	−5.48
Post-central gyrus, left	$CS^- > CS^+$	−50, −23, 46	751	−5.97
Pre-central gyrus, left	$CS^- > CS^+$	−53, −9, 30	150	−5.55
Pre-central gyrus, right	$CS^- > CS^+$	58, −13, 34	884	−5.84
Parahippocampal gyrus, left	$CS^- > CS^+$	−21, −36, −9	47	−5.47
Parahippocampal gyrus, right	$CS^- > CS^+$	29, −29, −12	6	−5.38
Caudate, right	$CS^+ > CS^-$	9, 4, 6	998	6.16

of the amygdala was hand drawn based on individual subjects' anatomy. For each ROI, we calculated β values for both the CS^+ versus CS^- for each subject within each phase (acquisition, day 1, and day 2 extinction) of the experiment. In order to examine extinction learning over trials, we also divided each phase into early and late components, as we had for the physiological responses. We then correlated across subjects the difference in the β values for the CS^+ and CS^- with the differential SCR response to the CS^+ versus CS^- in order to determine how activity in these regions may be related to the expression of the CR.

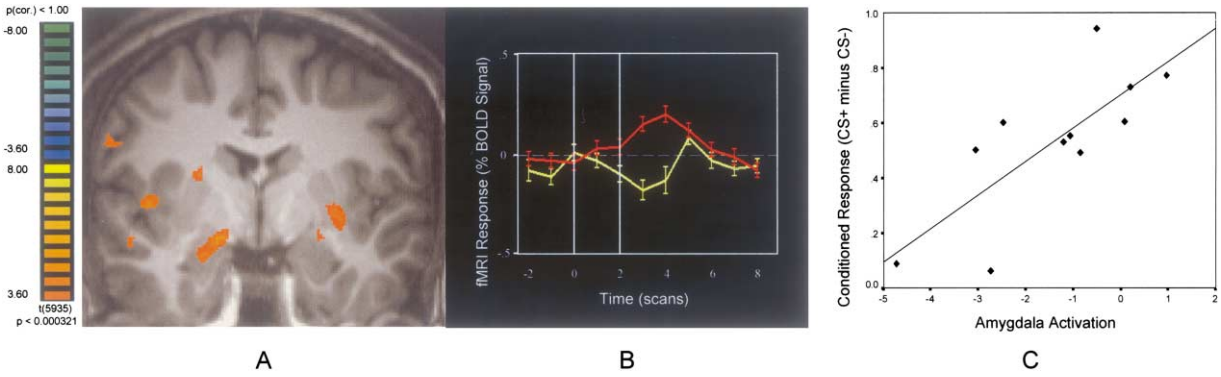

Figure 3. Amygdala Activation and Correlation with SCR

(A) Amygdala activation to the CS^+ during acquisition versus day 1 extinction.

(B) Mean time course of the amygdala response to CS^+ during acquisition (red line) and day 1 extinction (yellow line).

(C) Correlation between amygdala activation (β value for CS^+ minus CS^-) and the CR (SCR for the CS^+ minus CS^-) during day 1 extinction, indicating that greater differential amygdala response predicts greater extinction.

For the amygdala, the response shifted from acquisition to extinction, so that there was an increase in the BOLD response to CS^+ during acquisition and a decrease in the BOLD response to the CS^+ during day 1 extinction (see Figures 3A and 3B). By day 2 extinction, the amygdala response diminished. Although we observed responses in the right and left amygdala that had the same general pattern, activation was always greater in the right amygdala, and correlations with behavior were only significant for the right amygdala. During acquisition, there was a positive correlation between the strength of the right amygdala response to the CS^+ relative to the CS^- and the strength of the CR, but this was only significant in early acquisition ($r = 0.643$, $p < 0.05$), consistent with previous reports (LaBar et al., 1998). On day 1 extinction, even though the direction of the amygdala response changed (greater to CS^- than CS^+), a greater differential amygdala response was also correlated with the CR. This time, however, the greater differential amygdala response predicted less of a CR ($r = 0.771$, $p < 0.01$), suggesting that this shift in the amygdala response may be related to extinction learning and early extinction success (see Figure 3B). On day 2 extinction, there was no correlation between the amygdala and the magnitude of the CR.

Within the mPFC, the dorsal anterior cingulate region showed an increase in BOLD to the CS^+ versus CS^-, but there was no correlation with the CR during acquisition or extinction. The response observed in the ventral subgenual anterior cingulate and medial frontal gyrus was primarily expressed as a decrease to the CS^+. Although this was assessed relative to the CS^-, an analysis of the response patterns indicates this differential response was primarily driven by a depression in BOLD to the CS^+. For example, in the subgenual anterior cingulate ROI, an examination of the β values found that they were significantly less than zero for the CS^+ in all phases of the experiment [acquisition, $t(10) = -4.08$; extinction day 1, $t(10) = -4.96$; extinction day 2, $t(10) = -4.18$, $p < 0.01$), indicating a decrease in BOLD to the CS^+, whereas the response to the CS^- was not significantly different than zero. This depression diminished as extinction progressed each day (see Figure 4A).

When the magnitude of response in the two ROIs in the ventral mPFC was correlated with the expression of the CR during extinction, the only region that emerged as predictive of the CR was the subgenual anterior cingulate. There was no correlation between the strength of the CR and magnitude of the subgenual anterior cingulate response during acquisition or day 1 extinction. On day 2 extinction, activity at this region correlated with relative extinction success. Extinction success was determined by examining the change in the CR from early to late extinction trials. For the early trials of day 2 extinction, the response in the subgenual anterior cingulate was correlated with day 1 extinction success ($r = 0.748$, $p < 0.01$). That is, those subjects who showed more extinction on day 1 showed less of a subgenual anterior cingulate depression at the beginning of day 2 (see Figure 4B). This response was primarily driven by a correlation between the strength of the CR late on day 1 extinction and subgenual anterior cingulate response early on day 2 of extinction ($r = -0.701$, $p < 0.01$). This finding is consistent with the suggestion that the vmPFC may play a role in the retention of extinction learning. In addition, activity in this region on the later trials of day 2 extinction showed a marginally significant correlation with extinction success on that same day ($r = 0.590$, $p < 0.06$). These results suggest that the subgenual cingulate may be particularly involved in extinction processes after initial extinction learning, during retention.

Given that the only mPFC region that predicted extinction was the subgenual anterior cingulate, this region was the only one explored for its relation to the amygdala response. There was no correlation between the subgenual anterior cingulate response and the amygdala response during acquisition or day 1 extinction. On day 2 extinction, there was a significant correlation between the early responses in the subgenual anterior cingulate and the strength of amygdala activation ($r = 0.796$, $p < 0.01$), consistent with the hypothesis that this vmPFC region may be linked to a diminished amygdala response at retention.

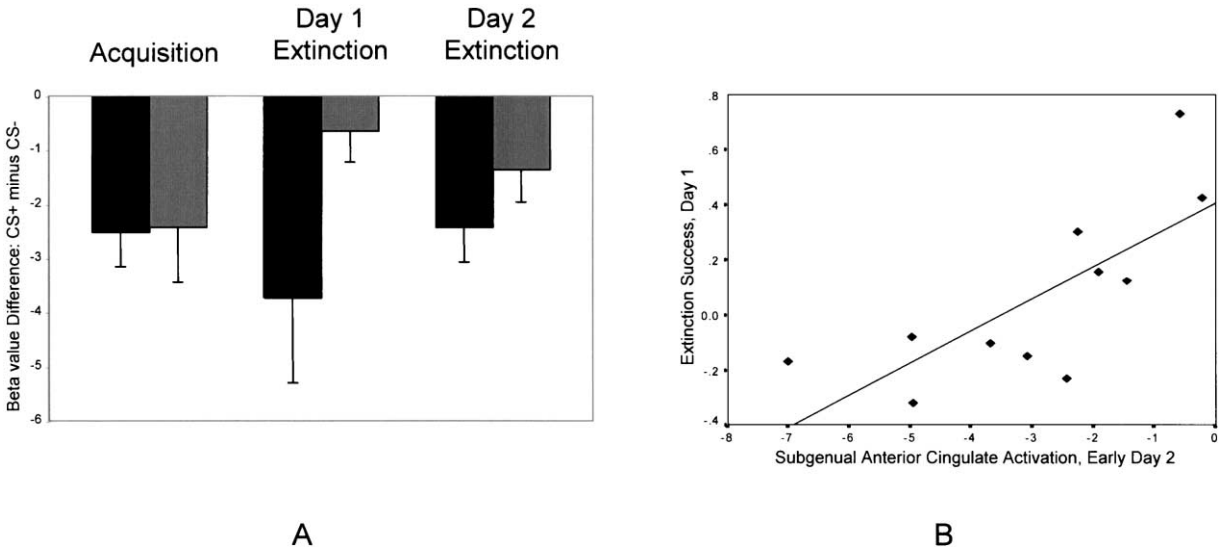

Figure 4. Subgenual Anterior Cingulate Activation and Correlation with SCR

(A) Mean differential β values (CS^+ minus CS^-) for the subgenual anterior cingulate ROI during early and late acquisition, day 1 extinction, and day 2 extinction. Error bars represent standard error.

(B) Correlation between subgenual anterior cingulate activation (β value for CS^+ minus CS^-) at the beginning of day 2 extinction and day 1 extinction success (CR for early day 1 extinction trials minus late trials—higher numbers indicate greater extinction).

In addition to the a priori ROIs within the mPFC and amygdala, we conducted an exploratory correlation analysis on the relation between the CR and BOLD responses in others brain regions that showed differential responses to the CS^+ versus CS^- throughout the three phases of the study. Three regions were identified: the caudate nucleus, the posterior cingulate, and the insular cortex. Within these regions, ROIs were selected using the same method used for the mPFC, that is, centering the ROI on the peak of activation (caudate, 9, 2, 9; insular cortex, 45, 11, 6; posterior cingulate, −7, −56, 13). The response in the caudate was significantly correlated with the CR during early acquisition ($r = 0.776$, $p < 0.01$). The posterior cingulate response was significantly correlated with the CR during day 1 extinction ($r = 0.658$, $p < 0.05$). No other significant correlations were observed.

Discussion

Much like the research on the acquisition of conditioned fear in humans, the present results on extinction learning largely support what has been learned from research with nonhuman animals. Consistent with the nonhuman animal studies suggesting that the amygdala is important for both the acquisition and extinction of conditioned fear (Myers and Davis, 2002), we found that amygdala activation predicted the CR in both acquisition and early extinction. The vmPFC (subgenual anterior cingulate), although active throughout all stages of learning, seems to be primarily linked to the expression of the conditioned response during the retention of extinction learning, as might have been expected from studies demonstrating that damage to this region in rats only impairs the retention or recall of extinction (Quirk et al., 2000). These initial results on extinction learning are encouraging in our efforts to form links across species, although there are some important differences in the paradigm and pattern of results.

Even though our goal was to create an extinction learning paradigm that was analogous to those typically used with nonhuman animals, there were some differences that may have influenced the findings. In the present study, we used a partial reinforcement paradigm in an effort to slow extinction learning in humans. It is unclear how changing from a 100% to a partial reinforcement paradigm alters the mechanisms of fear extinction. The present study also differed from the standard animal paradigm in that we used a discrimination procedure with a CS^- that could act as a baseline to the CS^+. Although this type of paradigm is ideal for fMRI analysis, this discrimination could change the nature of the task since both the CS^+ and CS^- provide information about the presentation of the US. Finally, the present study differed from those used with nonhuman animals in that subjects had verbal instruction of the CS-US contingency prior to fear conditioning. Although fear learning through instruction has been shown to be dependent on the amygdala for expression (Funayama et al., 2001), the combination of learning through instruction and conditioning may have altered the results, particularly the early amygdala response and the early expression of the CR during acquisition. In spite of these differences in the paradigms across species, our findings in humans are largely consistent with those from nonhuman animals.

The amygdala response observed in the present study was characterized by an increase in activation to the CS^+ relative to the CS^- during acquisition that was correlated with the strength of the CR, replicating our previous results (LaBar et al., 1998). However, during extinction learning, we observed a reversal of the response in

the amygdala, so that there was a significantly greater response to the CS^- relative to the CS^+ on day 1 extinction, with little overall response on day 2 extinction. The differential amygdala response on day 1 extinction correlated with extinction learning, that is, relatively greater response to the CS^- relative to the CS^+ predicted less of a CR. This reversal of the amygdala response in early extinction was unexpected, especially given that the only previous report of extinction learning in humans (LaBar et al., 1998) showed the opposite pattern. However, as mentioned above, this earlier study used a 100% reinforcement that led to very rapid extinction, so it is difficult to know if this extinction response was reflecting the end of acquisition, uncertainty, or extinction learning. In the present study, we slowed extinction to explore the mechanisms of extinction learning. The results showed a shift in the amygdala response during day 1 extinction that is correlated with a diminished CR, suggesting that the amygdala is actively coding the predictive value of the CS^+ and altering its response when new information is available during within session extinction learning as well as acquisition.

Although three ROIs were identified in the mPFC, only one of these, the ventral, subgenual anterior cingulate, was related to extinction behavior. This region is similar to what other investigators have called the ventral mPFC (Kim et al., 2003) or the subgenual PFC (Drevets et al., 1997). It is suggested (Kim et al., 2003) that this region in humans may be most analogous to the infralimbic area in rats that has been implicated in fear extinction (Morgan et al., 1993; Quirk et al., 2000). Even though it is not possible to precisely determine the homology across species, both its anatomical location and previous results suggest that this subgenual anterior cingulate region is a reasonable candidate.

The concept of infralimbic cortex was originally developed in the rabbit and cat as the cortex that is located under the tip of the callosum and is located in an analogous position to Brodmann's area 25 (the subcallosal cortex), with area 32 being the equivalent of the prelimbic cortex (Rose and Woosley, 1948). In monkeys, the prelimbic/infralimbic cortex has been suggested to encompass regions rostral and ventral to the corpus callosum, including BA areas 24, 32, and 25 (Amaral, 1992). In rats, the infralimbic cortex have been shown to share the same type of amygdala afferents as area 25 in monkeys (Vogt and Pandya, 1987; Porrino et al., 1981; Amaral and Price, 1984). In humans, although precise homology is difficult to determine, functional similarity across species may help support the notion that there are homologous regions. A previous study by Kim et al. (2003) showed an inverse correlation between activation of this subgenual anterior cingulate region and the amygdala response to faces with emotional expressions, consistent with connectivity suggested between the amygdala and infralimbic regions in monkeys (Amaral, 1992), as well as extinction studies in rats (Milad and Quirk, 2002).

In the present study, activation of this subgenual anterior cingulate region was correlated with the expression of the CR and the amygdala response primarily during the retention phase of extinction (day 2). Those subjects who showed more extinction learning on day 1 showed less of activation in this region at the beginning of day 2. A similar pattern was observed as extinction progressed on day 2. On day 1 of extinction, there was no relation between activation of this region and the expression of the CR or the amygdala. These results support the interpretation that responses in this region may be related to the retention of extinction learning specifically.

Overall, activation of the subgenual anterior cingulate in all phases of the study was characterized primarily as a depression in the BOLD response to the CS^+ that diminished as extinction learning progressed. Although a decrease in BOLD can be somewhat difficult to interpret (Gusnard and Raichle, 2001), there is some suggestion that a decrease in BOLD may reflect a reduction in neuronal activity (Shmuel et al., 2002; Zenger-Landolt and Heeger, 2003). The finding that this vmPFC region responds with a decrease in signal to the CS^+ throughout all phases of learning is inconsistent with the findings of Milad and Quirk (2002) using electrophysiology in rats, which showed an increase in the firing rate in vmPFC neurons to a CS only during the retention of extinction. However, the pattern of BOLD response in the present study is similar to a finding by Garcia et al. (1999) reporting a reduction in the spontaneous firing rate of neurons in this vmPFC region during fear conditioning. This study examined responses in mice during a conditioned inhibition paradigm. When the CS was presented alone, fully predicting the US, there was a reduction in the spontaneous firing rate. However, when a second stimulus (the conditioned inhibitor) preceded the CS, indicating that a shock would not be delivered, this reduction in the firing rate to the CS was diminished. In other words, as the CS became less predictive of the US, there was less of a reduction. This pattern mirrors that observed in the present study with fMRI. There was a depression of the BOLD response in the subgenual anterior cingulate that was diminished as extinction progressed and the CS^+ no longer predicted the occurrence of the US. Like the Garcia et al. (1999) study and unlike Milad and Quirk (2002), the present study examined responses to two stimuli: one that predicted the shock (the CS^+) and another that did not (CS^-). It is unclear if the addition of a second stimulus changed the nature of the response in this vmPFC region or if the inconsistency in the pattern of response between the present study and the Milad and Quirk (2002) findings is due to cross species differences. However, similar to the Milad and Quirk (2002) results, responses in this vmPFC region were predictive of extinction learning after a retention interval.

Understanding how fears are acquired is an important step in our ability to translate basic research to the treatment of fear-related disorders. Understanding how learned fears are diminished may be even more valuable. In the present study, we explored the links between what is known about the mechanisms of fear extinction from research with nonhuman animals and human function. The fMRI results support animal models suggesting that the amygdala may play an important role in extinction learning as well as acquisition and that vmPFC may be particularly involved in the retention of extinction learning. The present results provide a demonstration that the mechanisms of extinction learning may be preserved across species.

Experimental Procedures

Subjects

Eighteen right-handed subjects, 18 to 25 years of age, were recruited through posted advertisements. Six of these subjects were eliminated after day 1 due to a lack of an SCR response (nonresponders, $n = 3$) or a failure to show acquisition of the conditioned response ($n = 3$). One subject was eliminated due to an error in the parameters used for image acquisition. The remaining eleven subjects (5 male, 6 female) completed the study. All subjects gave informed consent and were paid for their participation.

Conditioning Paradigm and Psychophysiological Assessment

A simple discrimination, partial reinforcement paradigm was used. The CSs were colored squares (blue and yellow) and the US was a mild shock to the wrist. All CSs were presented for 4 s, with a 12 s ITI. One of the colored squares was designated as the CS^+ (paired with US) and the other the CS^- (never paired with shock). Subjects were instructed of these contingencies prior to the start of the conditioning paradigm. There were three phases to the study. During acquisition, there were 15 presentations of the CS^+ and CS^- and an additional eight presentations of the CS^+ that coterminated with the presentation of the US. Day 1 extinction immediately followed acquisition and consisted of 15 unreinforced presentations of the CS^+ and the 15 presentations of the CS^-. Approximately 24 hr after the first session, subjects participated in day 2 extinction, which consisted of 17 unreinforced presentations of the CS^+ and 17 presentations of the CS^-. Prior to day 2 extinction, subjects were told that the procedure would be similar to day 1, but shorter. The order of trial type presentation was designated by two pseudorandom orders, which were counterbalanced across subjects.

Mild shocks were delivered through a stimulating bar electrode attached with a velcro strap to the right wrist. A Grass Instruments stimulator was used with cable leads that were magnetically shielded and grounded through an RF filter. The subjects were asked to set the level of shock themselves using a work up procedure prior to scanning both days. In this procedure, the subject was first given a mild shock (200 ms duration, 50 pulses/s) which was gradually increased to the maximum level the subject indicated was "uncomfortable, but not painful" (the maximum shock given will be 50 V). Skin conductance was assessed with shielded Ag-AgCl electrodes attached to the middle phalanges of the second and third fingers of the left hand using BIOPAC systems skin conductance module. The electrode cables were grounded through an RF filter panel. Offline data analysis of SCR waveforms was conducted using AcqKnowledge software. The level of SCR response was assessed as the base to peak difference for the largest deflection in the 0.5 to 4.5 s window following stimulus onset (see LaBar et al., 1995). SCR was only analyzed on trials that did not coterminate with presentation of the US.

Imaging and Analysis Parameters

The study was conducted at the NYU Center for Brain Imaging using a 3T Siemens Allegra scanner and a Siemens head coil. The scanning session began with MPRage anatomical scans to obtain a 3D volume for slice selection. This was followed by 3 mm thick axial slices to obtain anatomical slices in the same plane as the functional data acquisition. Functional scans used a gradient echo sequence, TR = 2 s, TE = 20, flip angle = 90, FOV = 192, 3 mm slice thickness. A total of 35 axial slices were sampled for whole brain coverage. The in-plane resolution was 3 mm × 3 mm. Functional image acquisition was divided into four runs on day 1 (two for acquisition and two for extinction) and two runs on day 2. Between runs there was a break of approximately 15–30 s.

Imaging data were analyzed using Brain Voyager software (2000, version 4.9). The data were temporally and spatially smoothed (4 mm FWHM) and motion corrected. We set a threshold of 2 mm of movement to eliminate subjects due to excessive movement. None of the subjects exceeded this threshold. An overall group analysis was conducted for each phase of learning (acquisition, day 1 extinction, and day 2 extinction). Individual data were transformed into Talairach space for group analysis. Data for individual trial types were convolved with the canonical hemodynamic response using a general linear model. The group activation maps comparing BOLD responses to the CS^+ relative to the CS^- were thresholded at $p < 0.01$, Bonferoni corrected. Consistent with previous studies, a lower threshold was used for the amygdala due to the a priori hypotheses for response ($p < 0.005$, uncorrected). An overall group analysis was conducted for each phase of learning (acquisition, day 1 extinction, and day 2 extinction).

Because the research with nonhuman animals highlighted the amygdala and mPFC as regions important in fear extinction, an ROI analysis on individual subjects was conducted for the amygdala and regions of activation that emerged within the mPFC in the overall group analysis. The amygdala ROI was hand drawn on the anatomical images for each subject. The ROIs were drawn separately for the right and left amygdala. There were three significant mPFC ROIs defined in the group analysis. For each of these, the Talairach coordinates for the peak of activation was determined (see Table 1). This served as the center of a 6 mm^3 cube that defined the ROI for individual subjects. For each subject and each ROI, the β values for the CS^+ and CS^- were calculated for all phases of the experiment. In addition, separate analyses were conducted for the early and late trials of each phase in order examine the change in responses in these regions as learning progressed. Early and late trials were defined as they were for the physiological measurement (see Results).

Acknowledgments

The authors would like to thank Azurii Collier, Patrick Hof, Ben Holmes, Souheil Inati, Greg Quirk, Brett Sedgewick, Kristen Stedenfeld, George Tourtellot, and Paul Whalen for assistance with this project. This research was supported by the National Institutes of Health, P50 MH8911 to J.E.L. and MH62104 to E.A.P. The authors would also like to acknowledge the support of the Beatrice and Samuel A. Seaver Foundation.

Received: March 12, 2004
Revised: July 12, 2004
Accepted: August 30, 2004
Published September 15, 2004

Selected Reading

Amaral, D.G. (1992). Anatomical organization of the primate amygdaloid complex. In The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction, J.P. Aggleton, ed. (New York: Wiley-Liss), pp. 1–66.

Amaral, D.G. (2002). The primate amygdala and the neurobiology of social behavior: implications for understanding social anxiety. Biol. Psychiatry *51*, 11–17.

Amaral, D.G., and Price, J.L. (1984). Amygdalo-cortical projections in the monkey (Macaca fascicularis). J. Comp. Neurol. *230*, 465–496.

Bouton, M.E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. Biol. Psychiatry *52*, 976–986.

Drevets, W.C., Price, J.L., Simpson, J.R., Jr., Todd, R.D., Reich, T., Vannier, M., and Raichle, M.E. (1997). Subgenual prefrontal cortex abnormalities in mood disorders. Nature *386*, 824–827.

Falls, W.A., Miserendino, M.J., and Davis, M. (1992). Extinction of fear-potentiated startle: Blockade by infusion of an NMDA antagonist into the amygdala. J. Neurosci. *12*, 854–863.

Funayama, E.S., Grillon, C.G., Davis, M., and Phelps, E.A. (2001). A double dissociation in the affective modulation of startle in humans: Effects of unilateral temporal lobectomy. J. Cogn. Neurosci. *13*, 721–729.

Garcia, R., Vouimba, R.M., Baudry, M., and Thompson, R.F. (1999). The amygdala modulates prefrontal cortex activity relative to conditioned fear. Nature *402*, 294–296.

Gewirtz, J.C., Falls, W.A., and Davis, M. (1997). Normal conditioned inhibition and extinction of freezing and fear-potentiated startle following electrolytic lesions of medical prefrontal cortex in rats. Behav. Neurosci. *111*, 712–726.

Groenewegen, H.J., Wright, C.I., and Uylings, H.B. (1997). The ana-

tomical relationships of the prefrontal cortex with limbic structures and the basal ganglia. J. Psychopharmacol. *11*, 99–106.

Gusnard, D.A., and Raichle, M.E. (2001). Searching for a baseline: functional imaging and the resting human brain. Nat. Rev. Neurosci. *10*, 685–694.

Kim, H., Somerville, L.H., Johnstone, T., Alexander, A., and Whalen, P.J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. Neuroreport *14*, 2317–2322.

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.

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.

LeDoux, J.E. (2002). Synaptic Self—How Our Brains Become Who We Are (New York: Viking).

McDonald, A.J., Mascagni, F., and Guo, L. (1996). Projections of the medial and lateral prefrontal cortices to the amygdala: a Phaseolus vulgaris leucoagglutinin study in the rat. Neuroscience *71*, 55–75.

Milad, M.R., and Quirk, G.J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. Nature *420*, 70–74.

Morgan, M.A., Romanski, L.M., and LeDoux, J.E. (1993). Extinction of emotional learning: contribution of medial prefrontal cortex. Neurosci. Lett. *163*, 109–113.

Myers, K.M., and Davis, M. (2002). Behavioral and neural analysis of extinction. Neuron *36*, 567–584.

Phelps, E.A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. Curr. Opin. Neurobiol. *14*, 198–202.

Porrino, L.J., Crane, A.M., and Goldman-Rakic, P.S. (1981). Direct and indirect pathways from the amygdala to the frontal lobe in rhesus monkey. J. Comp. Neurol. *230*, 465–469.

Oschner, K.N., Bunge, S.A., Gross, J.J., and Gabrielli, J.D.E. (2002). Rethinking feelings: An fMRI study of the cognitive regulation of emotion. J. Cogn. Neurosci. *14*, 1215–1229.

Quirk, G.J., Russo, G.K., Barron, J.L., and Lebron, K. (2000). The role of ventromedial prefrontal cortex in the recovery of extinguished fear. J. Neurosci. *20*, 6225–6231.

Quirk, G.J., Likhtik, E., Pelletier, J.G., and Pare, D. (2003). Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. J. Neurosci. *23*, 8800–8807.

Rose, J.E., and Woosley, C.N. (1948). Structure and relations of limbic cortex and anterior thalamus nuclei in rabbit and cat. J. Comp. Neurol. *89*, 279–347.

Rosenkranz, J.A., Moore, H., and Grace, A.A. (2003). The prefrontal cortex regulates lateral amygdala neuronal plasticity and responses to previously conditioned stimuli. J. Neurosci. *23*, 11054–11064.

Schaefer, S.M., Jackson, D.C., Davidson, R.J., Kimberg, D.Y., and Thompson-Schill, S.L. (2002). Modulation of amygdalar activity by the conscious regulation of negative emotion. J. Cogn. Neurosci. *14*, 913–921.

Shmuel, A., Yacoub, E., Pfeuffer, J., Van de Moortele, P.F., Andriany, G., Hu, X., and Ugurbil, K. (2002). Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. Neuron *36*, 1195–1210.

Smith, E.E., and Jonides, J. (1999). Storage and executive processes in the frontal lobes. Science *283*, 1657–1661.

Talairach, J., and Tournoux, P. (1988). Co-Planar Stereotaxic Atlas of the Human Brain (New York: Thieme Medical Publishers).

Vogt, B.A., and Pandya, D.N. (1987). Cingulate cortex of the rhesus monkey I: Cytoarchitecture and thalamic afferents. J. Comp. Neurol. *262*, 256–270.

Walker, D.L., Ressler, K.J., Lu, K.T., and Davis, M. (2002). Facilitation of conditioned fear extinction by systemic administration or intra-amygdala infusions of D-cycloserine as assessed with fear-potentiated startle in rats. J. Neurosci. *22*, 2343–2351.

Zenger-Landolt, B., and Heeger, D.J. (2003). Response suppression in V1 agrees with psychophysics of surround masking. J. Neurosci. *23*, 6884–6893.