Published in final edited form as:

Trends Cogn Sci. 2010 June ; 14(6): . doi:10.1016/j.tics.2010.04.002.

Overlapping neural systems mediating extinction, reversal and regulation of fear

Daniela Schiller^{1,2} and Mauricio R. Delgado³

¹Center for Neural Science, New York University, New York, NY

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

³Department of Psychology, Rutgers University, Newark, NJ

Abstract

Learned fear is a process allowing quick detection of associations between cues in the environment and prediction of imminent threat ahead of time. Adaptive function in a changing environment, however, requires organisms to quickly update this learning and have the ability to hinder fear responses when predictions are no longer correct. Here we focus on three strategies that can modify conditioned fear, namely extinction, reversal, and regulation of fear, and review their underlying neural mechanisms. By directly comparing neuroimaging data from three separate studies that employ each strategy, we highlight overlapping brain structures that comprise a general circuitry in the human brain which potentially enables the flexible control of emotions such as fear, regardless of the particular task demands.

Changing fear

Fear learning allows an organism to use cues in the environment in order to predict upcoming aversive events. This is an efficient, rapid and persistent learning process where even after one learning trial, humans and animals are capable of accurately predicting danger and forming long lasting fear memories ¹. From an evolutionary perspective, this is adaptive in minimizing exposure to the source of threat, promoting ways of escape and avoidance, and saving the need to relearn. Ever-changing environments, however, introduce another challenge: the ability to flexibly readjust fear learning such that it appropriately tracks the ongoing change in circumstances (e.g., a stimulus might cease to signal danger while another becomes threatening).

Here, we provide an overview of the neural mechanisms underlying the ability to flexibly change fear. In particular, we focus on three representative ways to modify fear learning: 1) Extinction – a process by which learned fear responses are no longer expressed after repeated exposure to the conditioned stimuli with no aversive consequences ²; 2) Reversal – a procedure where fear responses are switched between two stimuli following a reversal of reinforcement contingencies ^{3, 4}; 3) Regulation – a technique involving a cognitive re-evaluation of the conditioned stimulus to attenuate a conditioned response ⁵ (Fig. 1). We first review what is currently known about the neural mechanisms underlying these different approaches to changing fear. Then, we directly compare three data sets collected independently with the paradigms described above and investigate the potential overlap between neural structures involved in adapting to changes in fear across the separate paradigms. We posit that the observed overlapping regions comprise a general circuitry in

the human brain that enables the flexible control of emotions such as fear, regardless of the particular task demands.

Extinction, reversal, and regulation of fear

One way to model fear learning in the laboratory is by Pavlovian fear conditioning wherein a neutral sensory stimulus (the conditioned stimulus; CS) such as a shape or a tone is presented in close temporal contiguity with an aversive stimulus (the unconditioned stimulus; US), such as an electric shock⁴. Consequently, organisms learn to fear the previously neutral stimulus because it is now predictive of the shock. Studies in humans commonly use a discrimination variant of this protocol where two different natural stimuli are presented, but only one is associated with the aversive outcome (CS+), while the other one serves to provide a baseline for comparison⁶ (CS-). Across species, a common finding is that the integrity of the amygdala is critical for the acquisition and expression of conditioned fear^{4, 6-12}. While neuroimaging and neuropsychological studies have supported a role for the human amygdala in emotional processing^{6, 11, 12}, animal studies have further detailed the contribution of specific amygdala subregions^{4, 7-9, 13, 14}.

Based on the understanding of how fear conditioning is attained and expressed in the brain, research has begun to elucidate the neural processes required to eliminate or modify these learned fear responses^{2, 15-19}. Three representative ways to modify fear learning are extinction², reversal^{3, 4}, and regulation of fear⁵ (Fig. 1). These paradigms differ in two key aspects. The first is the strategy to change fear, where an organism either forms a new representation that competes for expression with the initial learned fear (extinction and reversal), or uses cognitive control to change the representation of fear inherent in a stimulus (emotion regulation). The second is the presence of fear during the learning process. Reversal and regulation are similar in this sense as both are acquired and maintained in the presence of fear. In extinction, however, there is an overall reduction in fear as the threatening stimulus is removed (See Supplementary Online Material for examination of overlap based in these two key aspects; Table S1). The difference between extinction and reversal is particularly interesting because the causal inference in either case may differ, as well as what is learned about the environment. In the first case, the environment is safe and predictable due to extinction, whereas in the latter case, danger is continuously present but its predictability could dynamically shift between stimuli.

In light of these differences and commonalities it is interesting to explore whether a joint mechanism underlies the ability to change fear regardless of the particular strategy employed and what unique mechanisms are called upon due to specific task demands. In the next sections, we review findings from studies in humans using functional magnetic resonance imaging (fMRI) where brain activation is indexed by blood-oxygen-level-dependent (BOLD) responses. To directly pinpoint commonalities in the underlying neural mechanisms, we reanalyzed three previously reported data sets and extracted regions of overlap. This allowed us to gauge the extent to which different fear modulation strategies share a common neural circuitry specialized for changing fear. The index of fear in the three data sets we used was the skin conductance response (SCR). A widespread neural circuitry shows correlated activity with SCR during fear learning (Box 2; Table S2). For our reanalysis, however, we focused on regions that show correlated activity with the SCR measure but are also typically involved in studies of affective learning and value representation: namely the striatum and the vmPFC (Box 3; Table S3)²⁰⁻²³.

Fear extinction

Extinction occurs when the CS is repeatedly presented without the US, leading to a gradual lessening in the conditioned fear response¹. Extinction is considered a learning process,

forming a novel association between the CS and no-US that competes for expression with the initial CS-US association to take control over behavior^{1, 2, 24}. This view of extinction is based on findings that conditioned fear to the CS can return under certain conditions, suggesting the original CS-US association was still intact only not expressed^{24, 25}. The important parameters in determining the dominant association are the context of learning and passage of time². If after extinction, for example, an animal undergoes a stressful exposure (such as receiving unsignaled USs) in the same context of learning, the fear memory may be reinstated. Also, if an animal acquires fear in context A and extinguishes it in context B, fear response to the CS may be renewed in a context that is different than B^{2, 24}. Finally, fear response to the CS can spontaneously recover with the passage of time²⁶. These factors also affect reacquisition of conditioned fear when using the same extinguished stimuli²⁴. Reinstatement, renewal, spontaneous recovery and reacquisition, are therefore the major assays to gauge whether a memory is merely suppressed or permanently erased^{24, 26}.

Given that the memory is evidently not erased, a large body of animal research has investigated where it is maintained, how it is recalled, and how the competing association exerts its inhibitory effects^{2, 15-19, 27}. Building on the detailed knowledge of the neural mechanisms supporting acquisition of fear^{4, 7-9, 13, 14}, studies of extinction learning reveal a critical role of the medial prefrontal cortex (mPFC) and its interactions with the amygdala^{10, 12, 15-19, 27}. One suggested model is that during fear conditioning multimodal sensory inputs signaling the neutral (CS) and the aversive (US) stimuli converge onto neurons in the lateral amygdala (LA). The flow of information is either through thalamocortico-amygdala pathways, or direct thalamo-amygdala pathways. The CS-US convergence leads to the long term potentiation of CS input synapses, such that when the CS later occurs on its own, these inputs are sufficient to drive LA outputs and trigger the fear response^{4, 7-9, 13, 14}. The major output structure of the amygdala is the central nucleus (CE). Projections from the CE to the hypothalamus and brainstem mediate the fear response comprising of behavioral and physiological reactions including freezing, change in heart rate and blood pressure, and release of stress hormones^{4, 7-9, 13, 14}. Within the amygdala, information is relayed serially from LA directly to CE or via the basal nucleus (the basal and lateral nuclei together are referred to as the basolateral amygdala or BLA)^{4, 7-9, 13, 14}, but there is also evidence for parallel processing in BLA and CE²⁸⁻³².

Once the fear response is triggered, its maintenance is potentially mediated by a dorsal part of the mPFC called the prelimbic cortex³³. An adjacent region, the infralimbic cortex, is required for the reduction of fear seen following extinction training^{18, 19, 34, 35}. Neurons in this region terminate on an intermediate mass of inhibitory cells within the amygdala, called the intercalated cells, located on the border between BLA and CE²⁷. These cells exert inhibitory control of CE output by integrating excitatory inputs from BLA and mPFC, both of which undergo plasticity during extinction consolidation^{13, 27, 35}. Retrieval of extinction memory might involve potentiated inhibitory circuits in BLA or increased mPFC output to amygdala^{13, 34}. Inputs to the mPFC from various regions, including the hippocampus, cortical regions, and the thalamus, also contribute to the modulation of this inhibitory process^{13, 18, 19}. This simplified description is one possible model and it applies mostly to auditory fear conditioning and extinction. Learning about other modalities (such as visual or gustatory) or about context might involve other systems including the perirhinal and visual cortex, insula and hippocampus^{2, 36, 37}.

In the human brain, the ventral portion of the medial prefrontal cortex (vmPFC), located below and anterior to the genu of the corpus callosum, is the putative homologue of the infralimbic PFC in non-human primates and rodents^{38, 39}. Human fMRI experiments confirm the functional similarities across species using fear conditioning and extinction

paradigms^{6, 10-12}. Specifically, amygdala BOLD signals were shown to increase during fear conditioning and early extinction, and decrease as extinction training progressed and as a function of extinction retrieval⁴⁰⁻⁴³. In contrast, BOLD signals in the vmPFC were shown to increase during extinction training and recall^{38, 40, 43, 44}, with signals during recall correlating with the success of extinction learning⁴³. The recall of extinguished memories was context-dependent, as previously shown in rats and humans^{24, 45-48}, and co-activated the hippocampus^{38, 44}. The amount of recall further correlated with vmPFC thickness⁴⁷. Finally, consistent with the view that PTSD might involve deficient extinction processes⁴⁹⁻⁵³, PTSD patients typically show vmPFC hypofunction and reduced volume, along with increased amygdala activation and hippocampal abnormalities⁴⁹⁻⁵³.

Fear reversal

In vast contrast to the rapidly growing knowledge about the neural mechanisms of fear extinction, very little is known about the neural processes mediating reversal of Pavlovian fear conditioning. This is surprising given the close relationship between the two paradigms. In both cases, the initial CS-US association is suppressed by new learning introduced in a subsequent phase^{54, 55}. A typical reversal procedure starts with the acquisition phase where two stimuli are presented, one is associated with the US (CS+) and the other is not (CS-). This is followed by reversal wherein the CS+ is no longer associated with the US (in essence undergoing extinction, becoming 'new CS-') while the CS- is now paired with the US ('new CS+'). A recent study examined the neural processes underlying reversal of conditioned fear in the human brain using fMRI⁵⁶. Throughout the task, the amygdala and the striatum tracked the stimuli that predicted the shock by showing increased BOLD responses to the CS+ (during acquisition) and the 'new CS+' (after reversal). In contrast, the vmPFC, which projects to both amygdala and striatum^{57, 58}, tracked those stimuli that were not paired with the shock (CS- and 'new CS-'). Moreover, responses in the vmPFC were stronger to the 'new CS-' compared to the CS- suggesting it might uniquely signal 'safety' or positive value for stimuli that were previously associated with an aversive US.

Another study of Pavlovian fear reversal in humans⁵⁹ found different results. This study reported increased vmPFC activation in response to the CS+ compared to CS- during acquisition, followed by a reversal of these responses. However, this pattern of responding is atypical of the vmPFC in aversive manipulations. This region typically shows a decrease in response to aversive outcomes and an increase in response to positive outcomes⁶⁰⁻⁶². An increase in vmPFC responses have even been observed following successful instrumental avoidance of an aversive outcome⁶³. A possible explanation for this discrepancy might be that this study used an indirect, task-irrelevant, instrumental measure of fear reactions (reaction time) as opposed to other studies assessing physiological changes (such as skin conductance response or fear potentiated startle) that typically correspond to changes in emotional states⁶⁴.

Although very little is known about reversal of Pavlovian fear conditioning, the neural mechanisms underlying the reversal of instrumental responses driven by aversive or appetitive outcomes have been more thoroughly investigated, with such research implicating the lateral region of the ventral PFC as a key structure^{59, 62, 65-68}. Increased activation in this region has also been associated with punishment, reward omission and with a response switch^{62, 69}. It is possible that aversive instrumental and Pavlovian reversal might be dissociated in the lateral and medial regions of the ventral PFC, respectively. The former may mediate inhibition of instrumental responses whereas the latter may mediate inhibition of physiological fear reactions. However, there are other fundamental differences between these studies. For example, here the reversal was between aversive and neutral associations, whereas previous studies shifted between appetitive and aversive associations. Those studies also use serial reversals, which might engage higher-order rule learning and different

temporal integration⁷⁰. Thus, additional studies are required to elucidate the differential contribution of these two regions to reversal learning.

Regulation of fear

Understanding the neurobiology of how fears can be changed and adapted has traditionally relied on a rich animal literature and the use of classical models of learning. An alternative for humans for controlling fears, however, may come from their distinct ability to use higher order cognitive strategies to regulate emotional responses. The application of cognitive strategies typically involve changing the way one thinks about a situation or a stimulus in order to alter one's emotional reaction to it and can also vary with respect to the time of application⁵. For instance, antecedent-focused emotion regulation strategies can act early in the emotion generation process to attenuate experienced emotion, compared to more response-focused strategies (e.g., suppression) which focus on the response to the negative outcome itself⁷¹. The most frequent approach involves antecedent-focused emotion regulation strategies, and ranges from general cognitive strategies aimed at diverting attention from the aversive stimulus (e.g., thinking of something calming rather than the source of anguish) to more focused reevaluations of stimuli into less negative contexts (e.g., reinterpreting the image of a screaming woman as an actor playing a scene), a strategy commonly known as reappraisal⁷¹.

The successful use of emotion regulation strategies have been shown to reduce the experience of negative emotion when viewing negatively valenced pictures⁵. In such studies, the use of reappraisal while viewing a negative stimulus is contrasted with a control condition such as attending to one's natural emotions. Trials where emotion regulation is applied are characterized by increases in BOLD signals in various cortical regions such as the dorsolateral prefrontal cortex (dlPFC), a region commonly found in studies of executive processes and cognitive control⁷², coupled with decreases in BOLD signals in the amygdala. Previous emotion regulation studies have used a wide range of stimuli that depict a strong negative emotional content (e.g., pictures, movie clips, narratives) along with different types of negative emotions (e.g., sadness, disgust, pain) to support the main observation of top down modulation of emotional responses by cognitive strategies⁷³⁻⁸⁰. Although there are slight differences in the specific areas of prefrontal cortex recruited during emotion regulation across studies, these discrepancies are likely due to variations in the regulation technique, type of emotion elicited and affective stimuli used⁵.

More recently, the efficacy of cognitive strategies have been probed with relation to conditioned fear, using a paradigm and dependent measure typical of studies of diminishing conditioned fear such as extinction⁸¹. Participants were exposed to a CS+ (paired with a shock) and a CS-. Prior to CS presentation, an instructional cue prompted participants to either attend or regulate the upcoming CS⁷⁸. During attend participants focused on their natural feelings (e.g., "I may get a shock"), while during regulate participants used an imagery technique (e.g., "think of soothing scene from nature"). Emotional responses decreased during CS+ trials when regulation was used, suggesting that cognitive strategies can provide an efficient way to actively cope with conditioned fear. The use of cognitive strategies also led to increases in dlPFC and vmPFC, while attenuating BOLD signals to the CS+ in the amygdala. The pattern of activation in the vmPFC correlated with both the amygdala and dlPFC, suggesting a potential pathway through which cognitive strategies could influence conditioned fear. Specifically, these results suggest that higher order cognitive processes, potentially mediated by the dlPFC, may take advantage of mechanisms involved in passive extinction of fears, such as the vmPFC^{4, 6, 11, 15, 43, 47}, to exert an effect on subcortical regions involved in producing an emotional response.

A general neural mechanism for changing fear

In this review, we discuss recent efforts aimed at understanding the neural mechanisms underlying our ability to control our fears by focusing on three distinct strategies: Classic extinction, reversal learning and emotion regulation. A common pattern across the three paradigms is the recruitment of overlapping regions, including the amygdala, the striatum and the vmPFC during the initial acquisition and eventual modulation of the fear response. While the amygdala is often observed during studies of aversive conditioning, the striatum and the vmPFC are more typically associated with studies of positive reinforcement and affective learning²⁰⁻²³, where predictions about values of conditioned stimuli are acquired and updated dynamically (See Box 3). In the context of the aversive learning paradigms, activity in the amygdala and the striatum tracked the strength of the conditioned fear signal, with BOLD signals observed during the expression of a conditioned fear decreasing as learned fear changes. The vmPFC showed decreased levels of BOLD responses during fear acquisition, which increased as the conditioned stimuli become extinguished, reversed or regulated with cognitive strategies. This pattern was apparent during a reanalysis of the three data sets and examination of BOLD responses from the conjunction between the tasks in two specific regions of interest, the striatum and the vmPFC. This is a powerful demonstration of the consistency in activation patterns across this potential network involved in controlling fear irrespective of the particular strategy used to change fear.

In addition to identifying overlap across tasks, examining the differences could reveal how the system is adjusted according to particular task demands. Extinction studies were some of the first to reveal that vmPFC responses are related to the attenuation of conditioned fear responses^{10, 12, 15-19, 27}. Recent evidence from the reversal paradigm⁵⁶ suggested that vmPFC responses were stronger to a 'new CS-' during reversal (used to be a CS+) compared to a 'naïve' CS- during acquisition. These results suggest that the vmPFC does not encode overall reduction in fear, but rather a specific value signal or a selective safety signal related to the omission of the aversive US. While the same information is processed during extinction, the reversal data suggests that vmPFC responses would scale differently to various stimuli in the environment depending on their positive or safety properties. Another difference between the tasks was the unique activation of the dlPFC in emotion regulation⁸¹ but not extinction or reversal (Tables S1, S3). Emotion regulation involves cognitive re-evaluation⁵ whereas extinction and reversal are based on the learning of a new competing association^{1, 4}. The dlPFC is not directly connected with the amygdala but it might exert indirect effects via connections with the vmPFC^{81, 88}. It is possible that through these connections, the fear modulation system is susceptible to top down modulation from the dlPFC when cognitive regulation strategies are employed.

In the striatum, the pattern of responses mirrored the vmPFC. That is, increases of BOLD signal during the acquisition of a conditioned fear response that decreased after application of extinction, reversal or regulation. The human striatum, a region often associated with appetitive conditioning and positive reinforcers^{21, 23, 86, 87}, has also been shown to be involved in human aversive conditioning⁸⁵, suggesting a general role for the striatum in affective learning irrespective of the emotional context (positive or negative). Recent rodent⁸⁹⁻⁹¹ and human⁹² studies postulate the striatum's role in aversive learning to involve interactions with the amygdala that will lead to an active response to the conditioned fear. However, the level of specificity between nuclei within the amygdala and regions of the striatum are currently limited in human studies. The use of high resolution imaging in the future could enhance this discussion, further investigating the interaction between the amygdala and striatum during both affective learning and the acquisition of an adaptive response to cope with learned fears.

Concluding remarks

In this review, we outline a potential neural circuit in the human brain that may underlie the successful adaptation to a fearful environment. Irrespective of the particular strategy involved in modulating fear responses, the amygdala, the striatum and the vmPFC were found to identify stimuli in the environment that are predictive of danger, while also adjusting their responses when predictions change. While the particular computation carried out by each component of this circuitry, along with what determines the transition between fear and non-fear states remains to be resolved (see Box 3 for outstanding questions), the implication of this collection of studies is that changing learned fear relies on a common neural mechanism, despite the type of strategies, that essentially allows for the flexible control of emotions. Whether such flexibility could be applied in either direction is currently unclear (see Box 3 for outstanding questions). Nevertheless, the existing literature allows for speculation about the role of each structure during aversive learning, with the initial motivational value being calculated in the amygdala, but being maintained and updated in the striatum and the vmPFC. The intra-connectivity between these structures would then subserve different functions, including inhibitory control over fear responses via vmPFC-amygdala connections, and output to motor systems via amygdala-striatum connections to initiate instrumental responses to cope with conditioned fear.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to acknowledge Ifat Levy for advice on the reanalysis included in this review and comments on earlier versions of this manuscript. We also thank Elizabeth Phelps, Joseph LeDoux and Joshua Johansen for discussions, and the anonymous reviewers for their constructive comments. During manuscript preparation, MRD was supported by NIDA grant (RO1 DA027764), and DS was supported by NIH R21 grant (MH072279) to Elizabeth Phelps.

References

1. Pearce JM, Hall G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev.* 1980; 87:532–552. [PubMed: 7443916]
2. Bouton ME, et al. Contextual and temporal modulation of extinction: behavioral and biological mechanisms. *Biol Psychiatry.* 2006; 60:352–360. [PubMed: 16616731]
3. Kehagia AA, et al. Learning and cognitive flexibility: frontostriatal function and monoaminergic modulation. *Curr Opin Neurobiol.* 2010 DOI 10.1016/j.conb.2010.1001.1007.
4. LeDoux JE. Emotion circuits in the brain. *Annu Rev Neurosci.* 2000; 23:155–184. [PubMed: 10845062]
5. Ochsner KN, Gross JJ. The cognitive control of emotion. *Trends Cogn Sci.* 2005; 9:242–249. [PubMed: 15866151]
6. Phelps EA, LeDoux JE. Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron.* 2005; 48:175–187. [PubMed: 16242399]
7. Davis M, Shi C. The amygdala. *Curr Biol.* 2000; 10:R131. [PubMed: 10704422]
8. Fanselow MS, Poulos AM. The neuroscience of mammalian associative learning. *Annu Rev Psychol.* 2005; 56:207–234. [PubMed: 15709934]
9. Maren S. Neurobiology of Pavlovian fear conditioning. *Annu Rev Neurosci.* 2001; 24:897–931. [PubMed: 11520922]
10. Milad MR, et al. Fear extinction in rats: implications for human brain imaging and anxiety disorders. *Biol Psychol.* 2006; 73:61–71. [PubMed: 16476517]
11. Phelps, EA.; Whelan, PJ. *The Human Amygdala.* Guilford Press; 2009.

12. Sehlmeier C, et al. Human fear conditioning and extinction in neuroimaging: a systematic review. *PLoS One*. 2009; 4:e5865. [PubMed: 19517024]
13. Ehrlich I, et al. Amygdala inhibitory circuits and the control of fear memory. *Neuron*. 2009; 62:757–771. [PubMed: 19555645]
14. LeDoux, JE.; Schiller, D. *The Human Amygdala*. Gilford Press; 2009. *The Human Amygdala: Insights from Other Animals*; p. 43-60.
15. Milad MR, Quirk GJ. Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*. 2002; 420:70–74. [PubMed: 12422216]
16. Myers KM, Davis M. Mechanisms of fear extinction. *Mol Psychiatry*. 2007; 12:120–150. [PubMed: 17160066]
17. Pare D, et al. New vistas on amygdala networks in conditioned fear. *J Neurophysiol*. 2004; 92:1–9. [PubMed: 15212433]
18. Quirk GJ, Mueller D. Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology*. 2008; 33:56–72. [PubMed: 17882236]
19. Sotres-Bayon F, et al. Brain mechanisms of fear extinction: historical perspectives on the contribution of prefrontal cortex. *Biol Psychiatry*. 2006; 60:329–336. [PubMed: 16412988]
20. Balleine BW, O'Doherty JP. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*. 2010; 35:48–69. [PubMed: 19776734]
21. Hare TA, et al. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J Neurosci*. 2008; 28:5623–5630. [PubMed: 18509023]
22. Knutson B, Cooper JC. Functional magnetic resonance imaging of reward prediction. *Curr Opin Neurol*. 2005; 18:411–417. [PubMed: 16003117]
23. Rangel A, et al. A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci*. 2008; 9:545–556. [PubMed: 18545266]
24. Bouton ME. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol Psychiatry*. 2002; 52:976–986. [PubMed: 12437938]
25. Rescorla RA. Retraining of extinguished Pavlovian stimuli. *J Exp Psychol Anim Behav Process*. 2001; 27:115–124. [PubMed: 11296487]
26. Rescorla RA. Spontaneous recovery. *Learn Mem*. 2004; 11:501–509. [PubMed: 15466300]
27. Likhtik E, et al. Amygdala intercalated neurons are required for expression of fear extinction. *Nature*. 2008; 454:642–645. [PubMed: 18615014]
28. Balleine BW, Killcross S. Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci*. 2006; 29:272–279. [PubMed: 16545468]
29. Cardinal RN, et al. Effects of selective excitotoxic lesions of the nucleus accumbens core, anterior cingulate cortex, and central nucleus of the amygdala on autoshaping performance in rats. *Behav Neurosci*. 2002; 116:553–567. [PubMed: 12148923]
30. Sah P, et al. The amygdaloid complex: anatomy and physiology. *Physiol Rev*. 2003; 83:803–834. [PubMed: 12843409]
31. Samson RD, et al. Synaptic plasticity in the central nucleus of the amygdala. *Rev Neurosci*. 2005; 16:287–302. [PubMed: 16519006]
32. Wilensky AE, et al. Rethinking the fear circuit: the central nucleus of the amygdala is required for the acquisition, consolidation, and expression of Pavlovian fear conditioning. *J Neurosci*. 2006; 26:12387–12396. [PubMed: 17135400]
33. Burgos-Robles A, et al. Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *J Neurosci*. 2009; 29:8474–8482. [PubMed: 19571138]
34. Quirk GJ, et al. Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. *J Neurosci*. 2003; 23:8800–8807. [PubMed: 14507980]
35. Amano T, et al. Synaptic correlates of fear extinction in the amygdala. *Nat Neurosci*. 2010; 13:489–494. [PubMed: 20208529]
36. Pelletier JG, et al. Interaction between amygdala and neocortical inputs in the perirhinal cortex. *J Neurophysiol*. 2005; 94:1837–1848. [PubMed: 16105956]

37. Shi C, Davis M. Visual pathways involved in fear conditioning measured with fear-potentiated startle: behavioral and anatomic studies. *J Neurosci.* 2001; 21:9844–9855. [PubMed: 11739592]
38. Milad MR, et al. Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol Psychiatry.* 2007; 62:446–454. [PubMed: 17217927]
39. Ongur D, et al. Architectonic subdivision of the human orbital and medial prefrontal cortex. *J Comp Neurol.* 2003; 460:425–449. [PubMed: 12692859]
40. Gottfried JA, Dolan RJ. Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nat Neurosci.* 2004; 7:1144–1152. [PubMed: 15361879]
41. Knight DC, et al. Amygdala and hippocampal activity during acquisition and extinction of human fear conditioning. *Cogn Affect Behav Neurosci.* 2004; 4:317–325. [PubMed: 15535167]
42. LaBar KS, Disterhoft JF. Conditioning, awareness, and the hippocampus. *Hippocampus.* 1998; 8:620–626. [PubMed: 9882019]
43. Phelps EA, et al. Extinction learning in humans: role of the amygdala and vmPFC. *Neuron.* 2004; 43:897–905. [PubMed: 15363399]
44. Kalisch R, et al. Context-dependent human extinction memory is mediated by a ventromedial prefrontal and hippocampal network. *J Neurosci.* 2006; 26:9503–9511. [PubMed: 16971534]
45. Alvarez RP, et al. Contextual-specificity of short-delay extinction in humans: renewal of fear-potentiated startle in a virtual environment. *Learn Mem.* 2007; 14:247–253. [PubMed: 17412963]
46. LaBar KS, Phelps EA. Reinstatement of conditioned fear in humans is context dependent and impaired in amnesia. *Behav Neurosci.* 2005; 119:677–686. [PubMed: 15998188]
47. Milad MR, et al. Thickness of ventromedial prefrontal cortex in humans is correlated with extinction memory. *Proc Natl Acad Sci U S A.* 2005; 102:10706–10711. [PubMed: 16024728]
48. Schiller D, et al. Evidence for recovery of fear following immediate extinction in rats and humans. *Learn Mem.* 2008; 15:394–402. [PubMed: 18509113]
49. Bremner JD, et al. Structural and functional plasticity of the human brain in posttraumatic stress disorder. *Prog Brain Res.* 2008; 167:171–186. [PubMed: 18037014]
50. Liberzon I, Martis B. Neuroimaging studies of emotional responses in PTSD. *Ann N Y Acad Sci.* 2006; 1071:87–109. [PubMed: 16891565]
51. Milad MR, et al. Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biol Psychiatry.* 2009; 66:1075–1082. [PubMed: 19748076]
52. Rauch SL, et al. Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research—past, present, and future. *Biol Psychiatry.* 2006; 60:376–382. [PubMed: 16919525]
53. Shin LM, et al. Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. *Ann N Y Acad Sci.* 2006; 1071:67–79. [PubMed: 16891563]
54. Bouton ME. Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychol Bull.* 1993; 114:80–99. [PubMed: 8346330]
55. Brooks DC, Bouton ME. A retrieval cue for extinction attenuates spontaneous recovery. *J Exp Psychol Anim Behav Process.* 1993; 19:77–89. [PubMed: 8418218]
56. Schiller D, et al. From fear to safety and back: reversal of fear in the human brain. *J Neurosci.* 2008; 28:11517–11525. [PubMed: 18987188]
57. Haber SN, Knutson B. The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology.* 2010; 35:4–26. [PubMed: 19812543]
58. Peters J, et al. Extinction circuits for fear and addiction overlap in prefrontal cortex. *Learn Mem.* 2009; 16:279–288. [PubMed: 19380710]
59. Morris JS, Dolan RJ. Dissociable amygdala and orbitofrontal responses during reversal fear conditioning. *Neuroimage.* 2004; 22:372–380. [PubMed: 15110029]
60. Gottfried JA, et al. Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J Neurosci.* 2002; 22:10829–10837. [PubMed: 12486176]
61. Hampton AN, et al. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci.* 2006; 26:8360–8367. [PubMed: 16899731]

62. O'Doherty J, et al. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci.* 2001; 4:95–102. [PubMed: 11135651]
63. Kim H, et al. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol.* 2006; 4:e233. [PubMed: 16802856]
64. Critchley HD. Electrodermal responses: what happens in the brain. *Neuroscientist.* 2002; 8:132–142. [PubMed: 11954558]
65. Cools R, et al. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J Neurosci.* 2002; 22:4563–4567. [PubMed: 12040063]
66. Evers EA, et al. Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. *Neuropsychopharmacology.* 2005; 30:1138–1147. [PubMed: 15689962]
67. Rolls ET. The functions of the orbitofrontal cortex. *Brain Cogn.* 2004; 55:11–29. [PubMed: 15134840]
68. Schoenbaum G, et al. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron.* 2003; 39:855–867. [PubMed: 12948451]
69. Schoenbaum G, et al. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat Neurosci.* 1998; 1:155–159. [PubMed: 10195132]
70. Glascher J, Buchel C. Formal learning theory dissociates brain regions with different temporal integration. *Neuron.* 2005; 47:295–306. [PubMed: 16039570]
71. Gross JJ. Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology.* 2002; 39:281–291. [PubMed: 12212647]
72. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci.* 2001; 24:167–202. [PubMed: 11283309]
73. Goldin PR, et al. The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biol Psychiatry.* 2008; 63:577–586. [PubMed: 17888411]
74. Harenski CL, Hamann S. Neural correlates of regulating negative emotions related to moral violations. *Neuroimage.* 2006; 30:313–324. [PubMed: 16249098]
75. Kalisch R, et al. Anxiety reduction through detachment: subjective, physiological, and neural effects. *J Cogn Neurosci.* 2005; 17:874–883. [PubMed: 15969906]
76. Kim SH, Hamann S. Neural correlates of positive and negative emotion regulation. *J Cogn Neurosci.* 2007; 19:776–798. [PubMed: 17488204]
77. Levesque J, et al. Neural circuitry underlying voluntary suppression of sadness. *Biol Psychiatry.* 2003; 53:502–510. [PubMed: 12644355]
78. Ochsner KN, et al. Rethinking feelings: an fMRI study of the cognitive regulation of emotion. *J Cogn Neurosci.* 2002; 14:1215–1229. [PubMed: 12495527]
79. Phan KL, et al. Neural substrates for voluntary suppression of negative affect: a functional magnetic resonance imaging study. *Biol Psychiatry.* 2005; 57:210–219. [PubMed: 15691521]
80. Urry HL, et al. Amygdala and ventromedial prefrontal cortex are inversely coupled during regulation of negative affect and predict the diurnal pattern of cortisol secretion among older adults. *J Neurosci.* 2006; 26:4415–4425. [PubMed: 16624961]
81. Delgado MR, et al. Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron.* 2008; 59:829–838. [PubMed: 18786365]
82. Boucsein, W. *Electrodermal Activity.* Plenum Press; 1992.
83. Bechara A, et al. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J Neurosci.* 1999; 19:5473–5481. [PubMed: 10377356]
84. Cheng DT, et al. Human amygdala activity during the expression of fear responses. *Behav Neurosci.* 2006; 120:1187–1195. [PubMed: 17201461]
85. Delgado MR, et al. The role of the striatum in aversive learning and aversive prediction errors. *Philos Trans R Soc Lond B Biol Sci.* 2008; 363:3787–3800. [PubMed: 18829426]
86. Kable JW, Glimcher PW. The neurobiology of decision: consensus and controversy. *Neuron.* 2009; 63:733–745. [PubMed: 19778504]

87. O'Doherty JP. Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr Opin Neurobiol.* 2004; 14:769–776. [PubMed: 15582382]
88. Quirk GJ, Beer JS. Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. *Curr Opin Neurobiol.* 2006; 16:723–727. [PubMed: 17084617]
89. Amorapanth P, et al. Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nat Neurosci.* 2000; 3:74–79. [PubMed: 10607398]
90. Cain CK, LeDoux JE. Escape from fear: a detailed behavioral analysis of two atypical responses reinforced by CS termination. *J Exp Psychol Anim Behav Process.* 2007; 33:451–463. [PubMed: 17924792]
91. Killcross S, et al. Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. *Nature.* 1997; 388:377–380. [PubMed: 9237754]
92. Delgado MR, et al. Avoiding negative outcomes: tracking the mechanisms of avoidance learning in humans during fear conditioning. *Front Behav Neurosci.* 2009; 3:33. [PubMed: 19847311]
93. Zink CF, et al. Human striatal responses to monetary reward depend on saliency. *Neuron.* 2004; 42:509–517. [PubMed: 15134646]
94. Tom SM, et al. The neural basis of loss aversion in decision-making under risk. *Science.* 2007; 315:515–518. [PubMed: 17255512]
95. Delgado MR, et al. Tracking the hemodynamic responses to reward and punishment in the striatum. *J Neurophysiol.* 2000; 84:3072–3077. [PubMed: 11110834]
96. Berns GS, et al. Predictability modulates human brain response to reward. *J Neurosci.* 2001; 21:2793–2798. [PubMed: 11306631]
97. Seymour B, et al. Differential encoding of losses and gains in the human striatum. *J Neurosci.* 2007; 27:4826–4831. [PubMed: 17475790]
98. McClure SM, et al. Neural correlates of behavioral preference for culturally familiar drinks. *Neuron.* 2004; 44:379–387. [PubMed: 15473974]
99. Gottfried JA, et al. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science.* 2003; 301:1104–1107. [PubMed: 12934011]
100. Herry C, et al. Switching on and off fear by distinct neuronal circuits. *Nature.* 2008; 454:600–606. [PubMed: 18615015]
101. Staudinger MR, et al. Cognitive reappraisal modulates expected value and prediction error encoding in the ventral striatum. *Neuroimage.* 2009; 47:713–721. [PubMed: 19442745]
102. Delgado MR, et al. Regulating the expectation of reward via cognitive strategies. *Nat Neurosci.* 2008; 11:880–881. [PubMed: 18587392]

Box 1**The relationship between brain activity and physiological index of conditioned fear**

Skin conductance response refers to phasic changes in electrical conductance of the skin resulting from neural activity of the sympathetic axis of the autonomic nervous system⁶⁴. Sweat glands are innervated by afferent neurons from the sympathetic axis, and applying a current to the skin and gauging changes in conductance can reveal their activity. SCR is therefore a sensitive measure indexing emotional responses associated with autonomic arousal⁸². The neural mechanisms mediating SCR include regions with autoregulatory function such as the hypothalamus and brainstem modulating SCR via homeostatic control of sympathetic arousal, as well as regions that exert higher-level control. For example, the amygdala and the vmPFC are associated with SCR induced by motivational processes such as stimulus-outcome associations and anticipatory behavior⁸³. The insula and anterior cingulate cortex are involved in integrating autonomic bodily states with behavior, and the parietal cortex is associated with attention induced changes in SCR⁶⁴.

There is evidence that SCR correlates with BOLD signals in the amygdala during fear expression⁸⁴, the vmPFC during extinction⁴³, and the dlPFC during regulation⁸¹. To probe the potential network that tracks the dynamics of the conditioned fear response in the human brain we used data from a previous reversal study⁵⁶. Specifically, SCR from each and every subject throughout acquisition and reversal was used as a regressor for brain activation (indexed by BOLD response; FDR correction for multiple comparisons set at the level of 0.05). To create the SCR regressor we computed a single SCR for each CS event and then convolved it with a hemodynamic response function. This analysis reveals a network of regions (Table S2) tracking the CS+ throughout the task (i.e., positively correlated with SCR), including the striatum, the insula and the dorsal anterior cingulate cortex (Fig. 2*a*); as well as regions negatively correlated with SCR, including the vmPFC and the posterior cingulate cortex (Fig. 2*b*).

Because different regions have distinct contributions to the modulation of SCR, understanding the relationship between SCR and regional neural activity is critical for the interpretation of fMRI studies. Within this network showing correlated activity with SCR during reversal of conditioned fear, we were interested in further examining the particular contribution of the striatum and the vmPFC, both implicated in the representation and update of value signals²⁰⁻²³ (see Box 2).

Box 2**Direct examination of overlapping neural systems underlying changing conditioned fear**

To directly compare the pattern of responses between the three fear modulation strategies, we extracted BOLD responses from the overlapping regions across the three paradigms (Fig. 3). Consistent with the abundant evidence for the important role of the amygdala in fear acquisition^{4, 6-14}, the three studies reported increased amygdala BOLD responses to the CS+ during acquisition or expression of fear and a reduction of these responses when the modulation strategy of extinction, reversal or regulation was applied^{43, 56, 81}. Here, we focused on two other regions of interest: the striatum and the vmPFC. The striatum receives projections from the amygdala⁵⁷ and has been previously linked with aversive learning in both human and non-human animals (see review⁸⁵). The relationship between the vmPFC and amygdala has been extensively investigated in extinction^{10, 12, 15-19, 27}, but more recently was the a focus of research on fear regulation and reversal^{56, 81}. Both regions have also been associated with positive reinforcement^{21, 23, 85-87}, suggesting an important role for processing of motivationally significant stimuli irrespective of valence^{23, 85}.

All three tasks were based on a discrimination fear conditioning paradigm with partial reinforcement (Fig. 1). The details of each procedure can be found in the original reports from which we took the data sets of extinction⁴³, reversal⁵⁶, and regulation⁸¹ of conditioned fear. For each task, we constructed statistical activation maps based on a contrast of all events versus fixation (FDR correction for multiple comparisons set at level of 0.05). This allowed us to probe regions engaged in the task without an *a priori* hypothesis. The activation maps were overlaid to outline the conjunction between the tasks in the regions of the striatum and the vmPFC (Fig. 3a; see Table S3 for complete list of regions). BOLD responses for each stimulus in each phase within each task were extracted from the entire conjunction region of the striatum (Fig. 3b, top panel; $x=11$, $y=4$, $z=9$, right side, 859mm³ voxels) and the vmPFC (Fig. 3b, bottom panel; $x=0$, $y=40$, $z=-3$, 2083mm³ voxels). The acquisition phase and fear modulation phase (extinction, reversal and regulation) are presented in gray and purple bars, respectively (Fig. 3c). The Y-axis represents the differential BOLD signal (CS+ minus CS-). Within each task, the differential scores varied significantly between the acquisition and modulation phases for all comparisons (two-tailed *t*-tests, $p < 0.05$), with the exception of vmPFC responses in the regulation task (showing a consistent trend). These results reveal striking similarities across regions during the three modulation strategies. The striatum showed increased activation to the fear predictive stimulus (CS+) in the acquisition phase. These responses decreased when this stimulus was extinguished or regulated, and switched following reversal of fear. In contrast, the vmPFC showed decreased activation to the fear predictive stimulus, and these responses increased with extinction or regulation, while switching following reversal of fear.

Box 3**Outstanding questions****Representation of value in the striatum**

The term valuation loosely refers to a process where values are assigned to stimuli or actions that guide the computation of decisions²³. Such values can be positive, as the case of a reward, or negative, as is the context of fear. In a conditioning experiment, valuation may occur during the initial stages of acquisition, when a conditioned stimulus acquires a positive or negative valence, while changes in conditioned fear may result due to a change in the initial prediction of the value of the stimulus. Evidence from aversive and appetitive tasks examining the role of striatum in the representation of value has been difficult to reconcile. One argument is that the striatum responds to salient events⁹³, or even primarily to rewarding stimuli²². However, studies using secondary reinforcers such as money often report decreases in striatum activity during either anticipation⁹⁴ or receipt⁹⁵ of negative outcomes. Another possibility is that the striatum is involved in affective learning, irrespective of reinforcer valence, and sensitive to the predictability of contingencies^{96,97}. Future studies may look to modulate not only the valence of a reinforcer (appetitive or aversive) but also the type (primary, secondary) or schedule (probabilistic or deterministic) of reinforcer to understand the role of the striatum in the representation of value.

Reconciling the role of the vmPFC in fear and reward learning

Activation patterns in the vmPFC typically track reward value^{21, 23, 60-62, 87}, often correlating with behavioral preferences⁹⁸. Interestingly, during the aversive learning paradigms described above, where the representation of fear changes from threat to non-threat, the vmPFC shows an increasing response as the representation of fear is diminished. This evidence leads to the suggestion that the vmPFC tracks changes in the representation of value as it becomes positive, exemplified by extinction and reversal learning studies where a change in contingencies to a more positive state leads to greater engagement of the vmPFC^{38, 40, 43, 44, 56}, along with other examples from devaluation of a conditioned stimulus showing decreases in BOLD signals in both amygdala and vmPFC⁹⁹.

The transition between fearful and non-fearful states

Studies to date have elucidated the neural processes occurring during the different phases of fear learning including acquisition, expression, and modulation of conditioned fear. However, an intriguing question is what mechanism determines the transition between these phases and the extent to which each state would be expressed. Two recent animal studies suggest that specific brain regions are involved in triggering the transition or regulating the balance between fearful and non-fearful states. Using fear acquisition and extinction protocols in rats, it was suggested that the expression of each state might depend on the balance between two adjacent regions in the medial PFC, the prelimbic and the infralimbic PFC³³, or that the transition between the states might be regulated by separate populations of neurons in the basal amygdala¹⁰⁰.

The direction of the emotional change

The amygdala, striatum and vmPFC were identified in this review as structures that flexibly adjust their responses when predictions of aversive outcomes change. One question of interest is whether this can occur irrespective of the direction of the emotional change, for example, controlling the expectation of rewards. Consistent with this idea, it has been shown that the use of cognitive strategies is effective in reducing

physiological responses (i.e., SCRs) and BOLD signals associated with the expectation of rewards (e.g., striatum), while engaging more prefrontal regions (e.g., dlPFC and vmPFC), suggesting that these structures play a more general role in emotional flexibility^{101, 102}. Further research will elucidate the distinct and common contributions of this circuitry across appetitive and aversive domains when learning is being constantly updated.

Figure 1. Schematic of the experimental procedures

The three tasks were based on a discrimination fear conditioning paradigm with partial reinforcement. The aversive outcome was a mild electric shock to the wrist (US, unconditioned stimulus). The conditioned stimuli were colored squares (in extinction and regulation) or faces (in reversal). For discrimination, one specific stimulus (e.g., a yellow square) was designated as the conditioned stimulus (CS+) and was paired with the shock on about 30% of the trials, whereas the other stimulus (e.g., a blue square) was never paired with the shock (CS-). In extinction, the conditioning session was followed by two extinction sessions (one immediately after and the other 24 hours later) consisted of repeated non-reinforced presentations of the CS+ and CS-. In reversal, the conditioning session was immediately followed by an identical conditioning session only with reversed reinforcement contingencies, such that the stimuli designated as CS+ and CS- flipped roles. In regulation, the conditioning trials were interleaved with the regulation trials. Before each trial, subjects were instructed to either attend (“Try to focus on your natural feelings”) or to regulate (“Try to think of something calming in nature”). The index of fear was skin conductance responses (SCR) detected by two electrodes attached to the first and second fingers. In all tasks, the stimuli were presented for 4 sec and the inter-trial-interval was 12 sec. The US lasted 200 msec co-terminating with the conditioned stimulus. Each trial type was typically presented 12-16 times.

Figure. 2.
Brain regions showing correlation between BOLD signals and SCR during reversal of conditioned fear (Placed in Box 1)

Figure. 3. Overlapping regions in the striatum and vmPFC show consistent activation patterns across three different fear modulation strategies (Placed in Box 2)