Review

Overlapping neural systems mediating extinction, reversal and regulation of fear

Daniela Schiller[1,2] and Mauricio R. Delgado[3]

[1] Center for Neural Science, New York University, New York, NY 10003, USA
[2] Department of Psychology, New York University, New York, NY 10003, USA
[3] Department of Psychology, Rutgers University, Newark, NJ 07102, USA

Learned fear is a process allowing quick detection of associations between cues in the environment and prediction of imminent threat. 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. This circuitry potentially enables the flexible control of fear, regardless of the particular task demands.

Changing learned fear

Fear learning allows an organism to use cues in the environment 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 would appropriately track 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 learned fear. In particular, we focus on three representative ways to modify fear learning: (i) extinction – a process by which learned fear responses are no longer expressed after repeated exposure to the conditioned stimulus with no aversive consequences [2]; (ii) reversal – a procedure in which fear responses are switched between two stimuli following a reversal of reinforcement contingencies [3,4]; (iii) regulation – a technique involving a cognitive re-evaluation of the conditioned stimulus to attenuate a conditioned response [5] (Figure 1). We first review what is currently known about the neural mechanisms underlying these different approaches to changing learned fear. Then, we directly compare three data sets collected independently with the paradigms described above. We investigate the potential overlap between neural structures involved in adapting to changes in learned 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 fear, irrespective 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 shock [4]. 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+), whereas the other one (CS−) serves to provide a baseline for comparison [6]. A common finding across species is that the integrity of the amygdala is crucial for the acquisition and expression of conditioned fear [4,6–12]. Neuroimaging and neuropsychological studies have supported a role for the human amygdala in emotional processing [6,11,12], whereas 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 [2,10,15–19]. Three representative ways to modify fear learning are extinction [2], reversal [3,4], and regulation of fear [5] (Figure 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 modulation process. Reversal and regulation are similar in this sense because both are

Corresponding author: Schiller, D. (schiller@cns.nyu.edu).

1364-6613/$ – see front matter doi:10.1016/j.tics.2010.04.002 Trends in Cognitive Sciences 14 (2010) 268–276

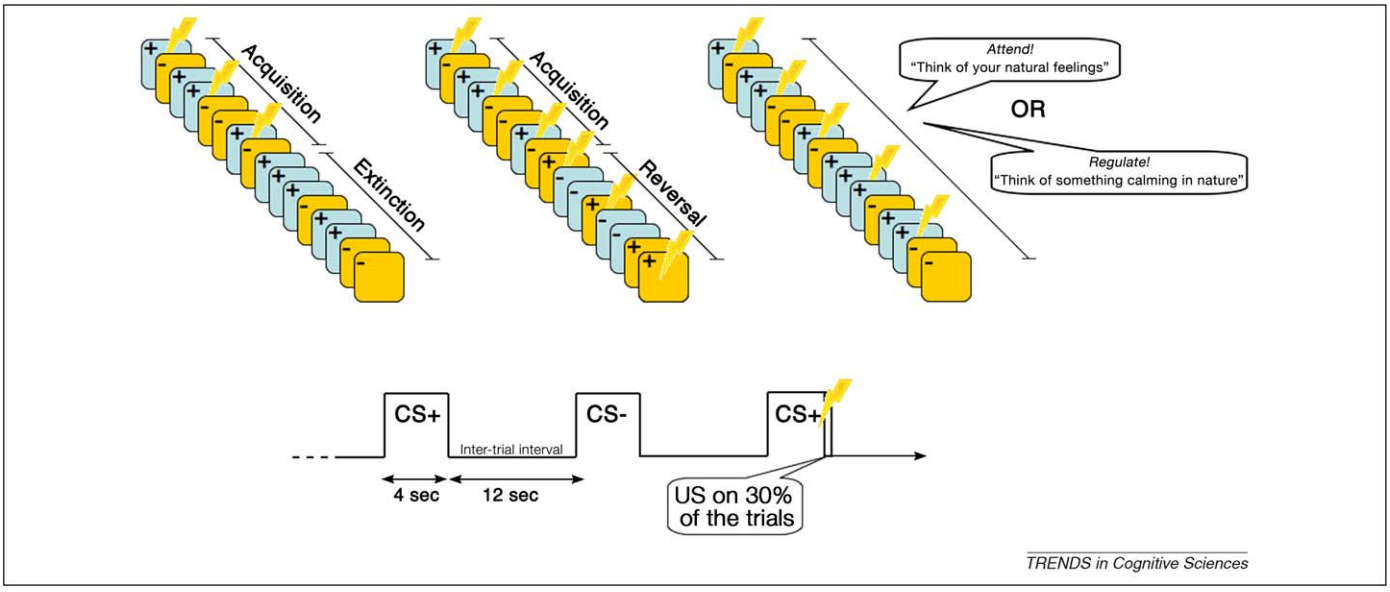

Figure 1. Schematic of the experimental procedures. The three tasks were based on a discrimination fear-conditioning paradigm with partial reinforcement. The aversive outcome was a mild electric shock to the wrist (US, unconditioned stimulus). The conditioned stimuli were colored squares (in extinction and regulation) or angry 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 an extinction session that consisted of repeated non-reinforced presentations of the CS+ and CS−. In reversal, the conditioning session was immediately followed by a similar 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 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 between 12 and 16 times.

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 an examination of overlap based on these two key aspects; Table S1). The difference between extinction and reversal is particularly interesting because the causal inference in either case can 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 learned fear. The index of fear learning in the three data sets we used was the skin conductance response (SCR). A widespread neural circuitry showed correlated activity with SCR during fear learning (Box 1; 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 ventral portion of the medial prefrontal cortex (vmPFC) [20–23].

Fear extinction

Extinction occurs when the CS is repeatedly presented without the US, leading to a gradual lessening in the conditioned fear response [1]. 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, indicating that the original CS–US association was still intact only not expressed [24,25]. Some of the important parameters in determining the dominant association are the context of learning and passage of time [2]. 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 could be reinstated. Also, if an animal acquires fear in context A and extinguishes it in context B, fear response to the CS could be renewed in a context that is different from B [2,24]. Finally, fear response to the CS can spontaneously recover with the passage of time [26]. These factors also affect reacquisition of conditioned fear when using the same extinguished stimuli [24]. Reinstatement, renewal, spontaneous recovery and reacquisition, are therefore the major assays to gauge whether a memory is merely suppressed or permanently erased [2,24,26].

Given that the memory is evidently not erased, a large body of animal research has investigated where is it maintained, how is it recalled, and how the competing association exerts its inhibitory effects [2,15–19,27]. Building on the detailed knowledge of the neural mechanisms supporting

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

Skin conductance response (SCR) refers to phasic changes in electrical conductance of the skin resulting from neural activity of the sympathetic axis of the autonomic nervous system [64]. 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 arousal [64,82]. 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 behavior [83]. 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 SCR [64].

There is evidence that SCR correlates with BOLD signals in the amygdala during fear expression [84], the vmPFC during extinction [43], and the dlPFC during regulation [81]. To probe the potential network in the human brain that tracks the dynamics of the conditioned fear response as assessed by SCR we used from a previous study on reversal of fear [56]. 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 (Figure Ia). Regions negatively correlated with SCR included the vmPFC and the posterior cingulate cortex (Figure Ib).

Because different regions have distinct contributions to the modulation of SCR, understanding the relationship between SCR and regional neural activity is crucial 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 signals [20–23] (see Box 2).

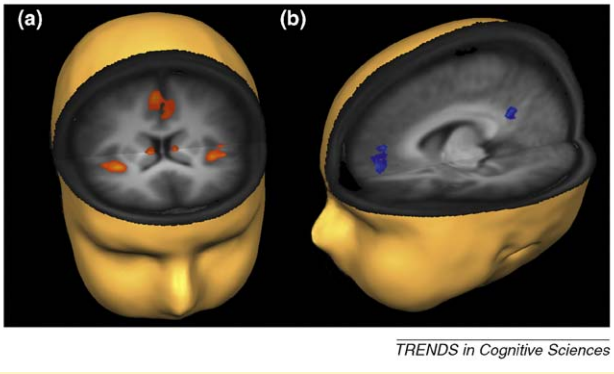

Figure I. Brain regions showing correlation between BOLD signals and SCR during reversal of conditioned fear

acquisition of learned fear [4,7–9,13,14], studies of extinction learning reveal a crucial role of the medial prefrontal cortex (mPFC) and its interactions with the amygdala [10,12,15–19,27]. One proposed 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 thalamo-cortico-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 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 CE [28–32].

Once the fear response is triggered, its maintenance is potentially mediated by a dorsal part of the mPFC called the prelimbic cortex [33]. 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 CE [17]. 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 through 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 vmPFC, located below and anterior to the genu of the corpus callosum, is the putative homolog 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 retrieval [40–43]. By 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 learning [43]. The recall of extinguished memories was context-dependent, as previously shown in rats and humans [24,45–48], and co-activated the hippocampus [38,41,44]. The amount of recall further correlated with vmPFC thickness [47]. Finally, consistent with the view that post-traumatic stress disorder (PTSD) might involve deficient extinction processes [49–53], PTSD patients typically show vmPFC hypofunction and reduced volume, along with increased amygdala activation and hippocampal abnormalities [49–53].

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 in which 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 fMRI [56]. 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). By 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−. This suggests that the vmPFC might uniquely signal 'safety' or positive value for stimuli that were previously associated with an aversive US.

Another study of Pavlovian fear reversal in humans [59] found different results. This study reported increased vmPFC activation in response to the CS+ compared with 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 outcomes [60–62]. An increase in vmPFC responses have even been observed following successful instrumental avoidance of an aversive outcome [63]. 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 that assessed physiological changes (such as SCR or fear potentiated startle) that typically correspond to changes in emotional states [64].

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 might mediate inhibition of instrumental responses whereas the latter might 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 used serial reversals, which might engage higher-order rule learning and different temporal integration [70]. 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, might come from their distinct ability to use higher-order cognitive strategies to regulate emotional responses. The application of cognitive strategies typically involves changing the way one thinks about a situation or a stimulus in order to alter one's emotional reaction to it and such strategies can also vary with respect to the time of application [5]. For instance, antecedent-focused emotion regulation strategies can act early in the emotion generation process to attenuate experienced emotion, compared with more response-focused strategies (e.g. suppression) that focus on the response to the negative outcome itself [71]. 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 re-evaluations 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 reappraisal [71].

The successful use of emotion regulation strategies has been shown to reduce the experience of negative emotion when viewing negatively valenced pictures [5]. 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 in which 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 control [72], 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 strategies [73–80]. Although there are slight differences in the specific areas of prefrontal cortex recruited during emotion regulation across studies, these discrepancies are probably due to variations in the regulation technique, type of emotion elicited and affective stimuli used [5].

More recently, the efficacy of cognitive strategies has been probed with relation to conditioned fear, using a paradigm and dependent measure typical of studies of diminishing conditioned fear such as extinction [81]. Participants were exposed to a CS+ (paired with a shock) and a CS−. Before CS presentation, an instructional cue prompted participants to either attend to or regulate the upcoming CS [78]. During 'attend' the participants focused on their natural feelings (e.g. "I may get a shock"), whereas during 'regulate' participants used an imagery technique

(e.g. "think of soothing scene from nature"). Emotional responses (assessed by SCR) decreased during CS+ trials when regulation was used, indicating that cognitive strategies can provide an efficient way to actively cope with conditioned fear. The use of cognitive strategies also led to increased BOLD signals in dlPFC and vmPFC, while attenuating BOLD signals 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 indicate that higher-order cognitive processes, potentially mediated by the dlPFC, can take advantage of mechanisms involved in passive extinction of fears, such as the vmPFC [4,6,11,12,15,43,47], to exert an effect on subcortical regions involved in producing an emotional response.

A general neural mechanism for changing learned 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.

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

To directly compare the pattern of responses among the three fear modulation strategies, we extracted BOLD responses from the overlapping regions across the three paradigms (Figure I). 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 amygdala [57] and has been previously linked with aversive learning in both human and non-human animals (see review [85]). The relationship between the vmPFC and amygdala has been extensively investigated in extinction [10,12,15,19,27] but was more recently the focus of research on fear regulation and reversal [56,81]. Both regions have also been associated with positive reinforcement [21,23,85–87], indicating 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 (Figure 1). The details of each procedure can be found in the original reports from which we took the data sets of extinction [43], reversal [56], and regulation [81] 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 (Figure Ia; see Table S3 for complete list of regions). The activation maps were overlaid to outline the conjunction between the tasks in the regions of the striatum and the vmPFC (Figure Ib). BOLD responses for each stimulus in each phase within each task were extracted from the entire conjunction region of the striatum (Figure Ib, top panel; x=11, y=4, z=9, right side, 859 mm^3 voxels) and the vmPFC (Figure Ib, bottom panel; x=0, y=40, z=−3, 2083 mm^3 voxels). The acquisition phase and fear modulation phase (extinction, reversal and regulation) are presented in gray and purple bars, respectively (Figure Ic). 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 to the CS− following reversal of fear. By contrast, the vmPFC showed decreased activation to the fear predictive stimulus, and these responses increased with extinction or regulation, and switched to the CS− following reversal of fear.

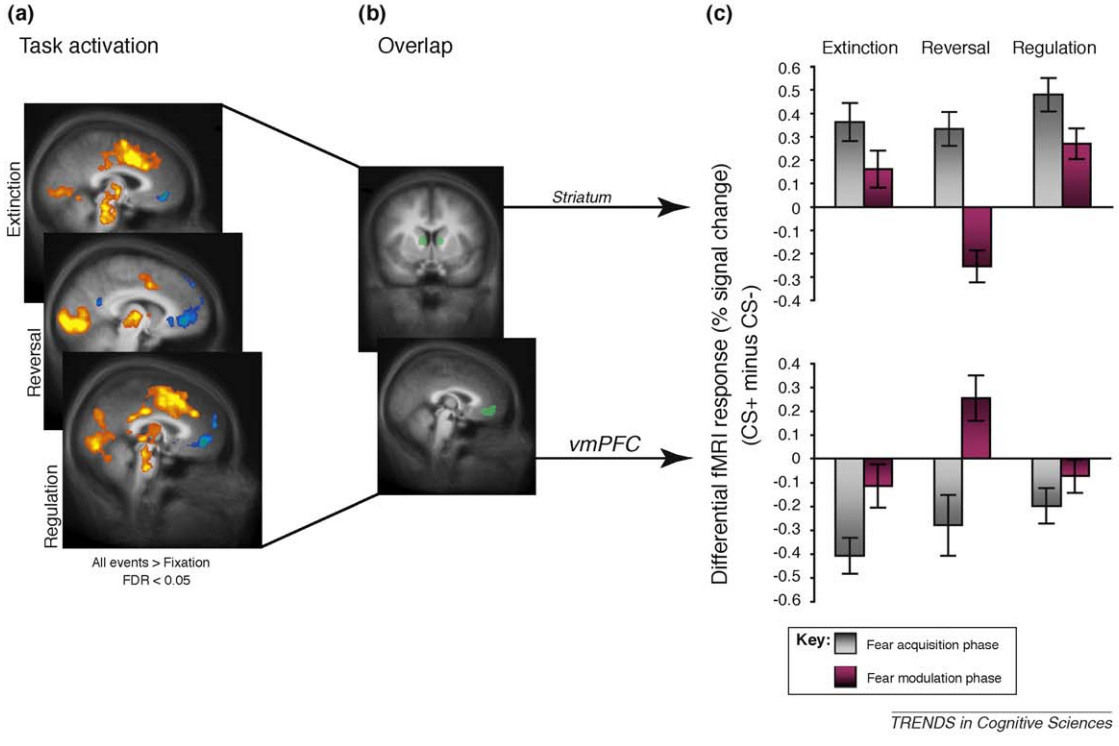

Figure I. Overlapping regions in the striatum and vmPFC show consistent activation patterns across three different fear modulation strategies

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 (see Box 2; Table S3). The amygdala is often observed during studies of aversive conditioning, whereas the striatum and the vmPFC are more typically associated with studies of positive reinforcement and affective learning [20–23] in which predictions about values of conditioned stimuli are acquired and updated dynamically (see also 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 became extinguished, reversed or regulated with cognitive strategies. This pattern was apparent in our 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 (Box 2; Table S3). 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 learned fear.

In addition to identifying overlap across tasks, examination of 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 paradigm [56] showed that vmPFC responses were stronger to a 'new CS−' during reversal (used to be a CS+) compared with a 'naïve' CS− during acquisition. These results indicate 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. Indeed, similar information is processed during extinction, but the reversal data further show that vmPFC responses 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 regulation [81] but not extinction or reversal (Tables S1 and S3). Emotion regulation involves cognitive re-evaluation [5] whereas extinction and reversal are based on the learning of a new competing association [1,54]. 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 conditioning [85]. This suggests general role for the striatum in affective learning irrespective of the emotional context (positive or negative). Recent rodent [89–91] and human

Box 3. Outstanding questions

- ***Representation of value in the striatum*** – The term valuation loosely refers to a process in which values are assigned to stimuli or actions that guide the computation of decisions [23]. Such values can be positive, as in the case of a reward, or negative, as in the context of fear. In a conditioning experiment, valuation might occur during the initial stages of acquisition, when a conditioned stimulus acquires a positive or negative valence, although changes in conditioned fear could 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 events [93], or even primarily to rewarding stimuli [22]. However, studies using secondary reinforcers such as money often report decreases in striatum activity during either anticipation [94] or receipt [95] of negative outcomes. Another possibility is that the striatum is involved in affective learning, irrespective of reinforcer valence, and is sensitive to the predictability of contingencies [96,97]. Future studies might 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 further 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 preferences [98]. 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 vmPFC [99].
- ***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 indicate 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 proposed that the expression of each state might depend on the balance between two adjacent regions in the medial PFC, the prelimbic and the infralimbic PFC [33], or that the transition between the states might be regulated by separate populations of neurons in the basal amygdala [100].
- ***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), indicating that these structures play a more general role in emotional flexibility [101,102].

[92] 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 could 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. 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 also Box 3). Nevertheless, 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 also Box 3). 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 further 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.

Acknowledgements

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 MIH R21 grant (MH072279) to Elizabeth Phelps.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tics.2010.04.002.

References

1 Pearce, J.M. and Hall, G. (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* 87, 532–552
2 Bouton, M.E. *et al.* (2006) Contextual and temporal modulation of extinction: behavioral and biological mechanisms. *Biol. Psychiatry* 60, 352–360
3 Kehagia, A.A. *et al.* (2010) Learning and cognitive flexibility: frontostriatal function and monoaminergic modulation. *Curr. Opin. Neurobiol.* DOI: 10.1016/j.conb.2010.1001.1007
4 LeDoux, J.E. (2000) Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184
5 Ochsner, K.N. and Gross, J.J. (2005) The cognitive control of emotion. *Trends Cogn. Sci.* 9, 242–249
6 Phelps, E.A. and LeDoux, J.E. (2005) Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron* 48, 175–187
7 Davis, M. and Shi, C. (1999) The extended amygdala: are the central nucleus of the amygdala and the bed nucleus of the stria terminalis differentially involved in fear versus anxiety? *Ann. N. Y. Acad. Sci.* 877, 281–291
8 Fanselow, M.S. and Poulos, A.M. (2005) The neuroscience of mammalian associative learning. *Annu. Rev. Psychol.* 56, 207–234
9 Maren, S. (2001) Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.* 24, 897–931
10 Milad, M.R. *et al.* (2006) Fear extinction in rats: implications for human brain imaging and anxiety disorders. *Biol. Psychol.* 73, 61–71
11 Phelps, E.A. and Whelan, P.J. (2009) *The Human Amygdala*, Gilford Press
12 Sehlmeyer, C. *et al.* (2009) Human fear conditioning and extinction in neuroimaging: a systematic review. *PLoS One* 4, e5865 DOI: 10.1371/journal.pone.0005865
13 Ehrlich, I. *et al.* (2009) Amygdala inhibitory circuits and the control of fear memory. *Neuron* 62, 757–771
14 LeDoux, J.E. and Schiller, D. (2009) The Human Amygdala: Insights from Other Animals. In *The Human Amygdala* (Whalen, P.J. and Phelps, E.A., eds), pp. 43–60, Gilford Press
15 Milad, M.R. and Quirk, G.J. (2002) Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* 420, 70–74
16 Myers, K.M. and Davis, M. (2007) Mechanisms of fear extinction. *Mol. Psychiatry* 12, 120–150
17 Pare, D. *et al.* (2004) New vistas on amygdala networks in conditioned fear. *J. Neurophysiol.* 92, 1–9
18 Quirk, G.J. and Mueller, D. (2008) Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology* 33, 56–72
19 Sotres-Bayon, F. *et al.* (2006) Brain mechanisms of fear extinction: historical perspectives on the contribution of prefrontal cortex. *Biol. Psychiatry* 60, 329–336
20 Balleine, B.W. and O'Doherty, J.P. (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology* 35, 48–69
21 Hare, T.A. *et al.* (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* 28, 5623–5630
22 Knutson, B. and Cooper, J.C. (2005) Functional magnetic resonance imaging of reward prediction. *Curr. Opin. Neurol.* 18, 411–417
23 Rangel, A. *et al.* (2008) A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.* 9, 545–556
24 Bouton, M.E. (2002) Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. *Biol. Psychiatry* 52, 976–986
25 Rescorla, R.A. (2001) Retraining of extinguished Pavlovian stimuli. *J. Exp. Psychol. Anim. Behav. Process* 27, 115–124
26 Rescorla, R.A. (2004) Spontaneous recovery. *Learn Mem.* 11, 501–509
27 Likhtik, E. *et al.* (2008) Amygdala intercalated neurons are required for expression of fear extinction. *Nature* 454, 642–645
28 Balleine, B.W. and Killcross, S. (2006) Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci.* 29, 272–279
29 Cardinal, R.N. *et al.* (2002) 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.* 116, 553–567
30 Sah, P. *et al.* (2003) The amygdaloid complex: anatomy and physiology. *Physiol. Rev.* 83, 803–834
31 Samson, R.D. *et al.* (2005) Synaptic plasticity in the central nucleus of the amygdala. *Rev. Neurosci.* 16, 287–302
32 Wilensky, A.E. *et al.* (2006) Rethinking the fear circuit: the central nucleus of the amygdala is required for the acquisition, consolidation, and expression of Pavlovian fear conditioning. *J. Neurosci.* 26, 12387–12396
33 Burgos-Robles, A. *et al.* (2009) Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *J. Neurosci.* 29, 8474–8482

34 Quirk, G.J. *et al.* (2003) Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. *J. Neurosci.* 23, 8800–8807
35 Amano, T. *et al.* (2010) Synaptic correlates of fear extinction in the amygdala. *Nat. Neurosci.* 13, 489–494
36 Pelletier, J.G. *et al.* (2005) Interaction between amygdala and neocortical inputs in the perirhinal cortex. *J. Neurophysiol.* 94, 1837–1848
37 Shi, C. and Davis, M. (2001) Visual pathways involved in fear conditioning measured with fear-potentiated startle: behavioral and anatomic studies. *J. Neurosci.* 21, 9844–9855
38 Milad, M.R. *et al.* (2007) Recall of fear extinction in humans activates the ventromedial prefrontal cortex and hippocampus in concert. *Biol. Psychiatry* 62, 446–454
39 Ongur, D. *et al.* (2003) Architectonic subdivision of the human orbital and medial prefrontal cortex. *J. Comp. Neurol.* 460, 425–449
40 Gottfried, J.A. and Dolan, R.J. (2004) Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nat. Neurosci.* 7, 1144–1152
41 Knight, D.C. *et al.* (2004) Amygdala and hippocampal activity during acquisition and extinction of human fear conditioning. *Cogn. Affect. Behav. Neurosci.* 4, 317–325
42 LaBar, K.S. and Disterhoft, J.F. (1998) Conditioning, awareness, and the hippocampus. *Hippocampus* 8, 620–626
43 Phelps, E.A. *et al.* (2004) Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* 43, 897–905
44 Kalisch, R. *et al.* (2006) Context-dependent human extinction memory is mediated by a ventromedial prefrontal and hippocampal network. *J. Neurosci.* 26, 9503–9511
45 Alvarez, R.P. *et al.* (2007) Contextual-specificity of short-delay extinction in humans: renewal of fear-potentiated startle in a virtual environment. *Learn Mem.* 14, 247–253
46 LaBar, K.S. and Phelps, E.A. (2005) Reinstatement of conditioned fear in humans is context dependent and impaired in amnesia. *Behav. Neurosci.* 119, 677–686
47 Milad, M.R. *et al.* (2005) Thickness of ventromedial prefrontal cortex in humans is correlated with extinction memory. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10706–10711
48 Schiller, D. *et al.* (2008) Evidence for recovery of fear following immediate extinction in rats and humans. *Learn Mem.* 15, 394–402
49 Bremner, J.D. *et al.* (2008) Structural and functional plasticity of the human brain in posttraumatic stress disorder. *Prog. Brain Res.* 167, 171–186
50 Liberzon, I. and Martis, B. (2006) Neuroimaging studies of emotional responses in PTSD. *Ann. N. Y. Acad. Sci.* 1071, 87–109
51 Milad, M.R. *et al.* (2009) Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biol. Psychiatry* 66, 1075–1082
52 Rauch, S.L. *et al.* (2006) Neurocircuitry models of posttraumatic stress disorder and extinction: human neuroimaging research–past, present, and future. *Biol. Psychiatry* 60, 376–382
53 Shin, L.M. *et al.* (2006) Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. *Ann. N. Y. Acad. Sci.* 1071, 67–79
54 Bouton, M.E. (1993) Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychol. Bull.* 114, 80–99
55 Brooks, D.C. and Bouton, M.E. (1993) A retrieval cue for extinction attenuates spontaneous recovery. *J. Exp. Psychol. Anim. Behav. Process* 19, 77–89
56 Schiller, D. *et al.* (2008) From fear to safety and back: reversal of fear in the human brain. *J. Neurosci.* 28, 11517–11525
57 Haber, S.N. and Knutson, B. (2010) The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35, 4–26
58 Peters, J. *et al.* (2009) Extinction circuits for fear and addiction overlap in prefrontal cortex. *Learn Mem.* 16, 279–288
59 Morris, J.S. and Dolan, R.J. (2004) Dissociable amygdala and orbitofrontal responses during reversal fear conditioning. *Neuroimage* 22, 372–380
60 Gottfried, J.A. *et al.* (2002) Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J. Neurosci.* 22, 10829–10837
61 Hampton, A.N. *et al.* (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.* 26, 8360–8367
62 O'Doherty, J. *et al.* (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* 4, 95–102
63 Kim, H. *et al.* (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol.* 4, e233 DOI: 10.1371/journal.pbio.0040233
64 Critchley, H.D. (2002) Electrodermal responses: what happens in the brain. *Neuroscientist* 8, 132–142
65 Cools, R. *et al.* (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J. Neurosci.* 22, 4563–4567
66 Evers, E.A. *et al.* (2005) Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. *Neuropsychopharmacology* 30, 1138–1147
67 Rolls, E.T. (2004) The functions of the orbitofrontal cortex. *Brain Cogn.* 55, 11–29
68 Schoenbaum, G. *et al.* (2003) Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* 39, 855–867
69 Schoenbaum, G. *et al.* (1998) Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat. Neurosci.* 1, 155–159
70 Glascher, J. and Buchel, C. (2005) Formal learning theory dissociates brain regions with different temporal integration. *Neuron* 47, 295–306
71 Gross, J.J. (2002) Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology* 39, 281–291
72 Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202
73 Goldin, P.R. *et al.* (2008) The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biol. Psychiatry* 63, 577–586
74 Harenski, C.L. and Hamann, S. (2006) Neural correlates of regulating negative emotions related to moral violations. *Neuroimage* 30, 313–324
75 Kalisch, R. *et al.* (2005) Anxiety reduction through detachment: subjective, physiological, and neural effects. *J. Cogn. Neurosci.* 17, 874–883
76 Kim, S.H. and Hamann, S. (2007) Neural correlates of positive and negative emotion regulation. *J. Cogn. Neurosci.* 19, 776–798
77 Levesque, J. *et al.* (2003) Neural circuitry underlying voluntary suppression of sadness. *Biol. Psychiatry* 53, 502–510
78 Ochsner, K.N. *et al.* (2002) Rethinking feelings: an FMRI study of the cognitive regulation of emotion. *J. Cogn. Neurosci.* 14, 1215–1229
79 Phan, K.L. *et al.* (2005) Neural substrates for voluntary suppression of negative affect: a functional magnetic resonance imaging study. *Biol. Psychiatry* 57, 210–219
80 Urry, H.L. *et al.* (2006) 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.* 26, 4415–4425
81 Delgado, M.R. *et al.* (2008) Neural circuitry underlying the regulation of conditioned fear and its relation to extinction. *Neuron* 59, 829–838
82 Boucsein, W. (1992) *Electrodermal Activity*, Plenum Press
83 Bechara, A. *et al.* (1999) Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *J. Neurosci.* 19, 5473–5481
84 Cheng, D.T. *et al.* (2006) Human amygdala activity during the expression of fear responses. *Behav. Neurosci.* 120, 1187–1195
85 Delgado, M.R. *et al.* (2008) The role of the striatum in aversive learning and aversive prediction errors. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363, 3787–3800
86 Kable, J.W. and Glimcher, P.W. (2009) The neurobiology of decision: consensus and controversy. *Neuron* 63, 733–745
87 O'Doherty, J.P. (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* 14, 769–776

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