The human amygdala and orbital prefrontal cortex in behavioural regulation

R. J. Dolan*

Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, UK

Survival in complex environments depends on an ability to optimize future behaviour based on past experience. Learning from experience enables an organism to generate predictive expectancies regarding probable future states of the world, enabling deployment of flexible behavioural strategies. However, behavioural flexibility cannot rely on predictive expectancies alone and options for action need to be deployed in a manner that is responsive to a changing environment. Important moderators on learning-based predictions include those provided by context and inputs regarding an organism's current state, including its physiological state. In this paper, I consider human experimental approaches using functional magnetic resonance imaging that have addressed the role of the amygdala and prefrontal cortex (PFC), in particular the orbital PFC, in acquiring predictive information regarding the probable value of future events, updating this information, and shaping behaviour and decision processes on the basis of these value representations.

Keywords: emotion; decision making; reward; fMRI; amygdala: orbital prefrontal cortex

1. AMYGDALA ENCODING OF VALUE

A general consensus that the human amygdala plays an important role in emotional processing begs a broader question as to the features of the sensory world to which it is responsive. There is a widely held view that emotion is reducible to dimensions of arousal and valence (Russell 1980; Lang 1995). Within this framework, an enhanced amygdala response to emotional stimuli has been proposed to reflect a specialization for processing emotional intensity (a surrogate for arousal), as opposed to processing valence. Consequently, amygdala activation by stimulus intensity, but not stimulus valence (Anderson & Sobel 2003; Small et al. 2003), is interpreted as supporting a view that external sensory events activate this structure by virtue of their arousal-inducing capabilities (McGaugh et al. 1996; Anderson & Sobel 2003; Hamann 2003).

If arousal is a critical variable in mediating the emotional value of sensory stimuli, then a prediction is that blockade of an arousal response should impair key functional characteristics of emotional stimuli, such as their ability to enhance episodic memory encoding. One means to experimentally influence an arousal response is by a pharmacological manipulation. Substantial evidence indicates that enhanced memory for emotional events engages a β-adrenergic central arousal system (Cahill & McGaugh 1998). β-Adrenergic blockade with the $\beta_1\beta_2$-receptor antagonist propranolol selectively impairs long-term human episodic memory for emotionally arousing material without affecting the memory for a neutral material (Cahill et al. 1994). This modulation of emotional memory by propranolol is centrally mediated because peripheral β-adrenergic

blockade has no effect on emotional memory function (van Stegeren et al. 1998). The fact that human amygdala lesions also impair emotional, but not non-emotional, memory (Cahill et al. 1995; Phelps et al. 1998) points to this structure as a critical locus for emotional effects on memory.

To test the impact of blockading arousal in response to emotional stimuli on episodic memory encoding, as well as to determine the locus of this effect, we presented human subjects with 38 lists of 14 nouns under either placebo or propranolol (Strange et al. 2003). Each list comprised 12 emotionally neutral nouns of the same semantic category, a perceptual oddball in a novel font, and an aversive emotional oddball of the same semantic category and perceptually equivalent to neutral nouns. As predicted from previous studies, we observed enhanced subsequent free recall for emotional nouns, relative to both control and perceptual oddball nouns. Furthermore, propranolol eliminated this enhancement such that memory for emotional nouns equated that for neutral nouns. Propranolol had no influence on memory for perceptual oddball items indicating that its effect was not related to an influence on oddball processing. Consequently, this finding supports the idea that arousal provides a key element in enhanced mnemonic processing, in this case encoding of emotional stimuli.

To determine the locus of this effect, we conducted an event-related functional magnetic resonance imaging (fMRI) experiment using identical stimulus sets. Twenty-four subjects received 40 mg of either propranolol or placebo in a double-blind experimental design (Strange & Dolan 2004). There were two distinct scanning sessions corresponding to encoding and retrieval, respectively. Drug/placebo was administered in the morning with the encoding scanning session coinciding with propranolol's peak plasma concentration. The retrieval session that took place 10 h later
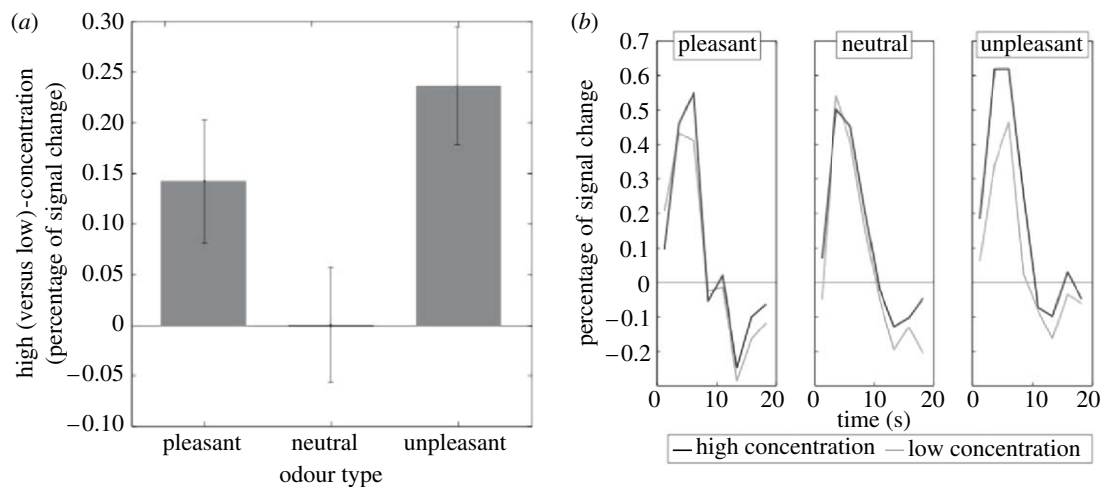
*r.dolan@fil.ion.ucl.ac.uk

Figure 1. Interaction between odour valence and intensity in the amygdala. (*a*) Plots represent the group-averaged peak fMRI signal in amygdala for high (versus low)-concentration odours at each valence level. An effect of intensity is evident for the pleasant and unpleasant, but not neutral, odours. This is reflected in a significant concentration–type interaction. (*b*) Time courses of amygdala activation for each level of odour concentration and odour type. Data highlight the effects of intensity on amygdala activity that are expressed only at the extremes of odour valence.

was not contaminated by the presence of drug. In the placebo group, successful encoding of emotional odd-balls, as assessed by successful retrieval 10 h later, engaged left amygdala relative to forgotten items. Under propranolol, amygdala activation no longer predicted subsequent memory for emotional nouns. Even more convincingly, the amygdala exhibited a significant three-way interaction for remembered versus forgotten emotional nouns, versus the same comparison for either control nouns or perceptual oddballs, under conditions of placebo compared with propranolol. Thus, adrener-gic-dependent amygdala responses do not simply reflect oddball encoding, but unambiguously show that successful encoding-evoked amygdala activation is β-adrenergic dependent. These findings fit with animal data demonstrating that inhibitory avoidance training increases noradrenaline/noradrenergic (NA) levels in the amygdala, where actual NA levels in individual animals correlate highly with later retention per-formance (McGaugh & Roozendaal 2002).

The fact that blockade of central arousal by propranolol impairs both emotional encoding and amygdala activation might seem to support an idea that an amygdala response indexes arousal. However, this conclusion is limited by the fact that most investigations, including the aforementioned, are pre-dicated on the idea that valence effects in respect of amygdala activation are linear. Few investigations have taken account of an alternative possibility, namely a nonlinearity of response such that effects of arousal are expressed only at the extremes of valence. One difficulty in addressing this question relates to the absence of a range of standard stimuli that have high arousal/inten-sity but are of neutral or low valence. Conveniently, it turns out that odour stimuli provide a means of unravelling these competing views as they can be independently classified in terms of hedonics (valence; Schiffman 1974) or intensity (an index of arousal; Bensafi *et al.* 2002). Valence as used in the present context is assumed to operate along a linear continuum of pleasantness, with stimuli of low (i.e. more negative)

valence representing a less pleasant sensory experience than those of higher (i.e. more positive) valence. Given that chemosensory strength or intensity takes on greater importance when a stimulus is pleasant or unpleasant, then a nonlinearity in response would predict that amygdala activation to intensity should be expressed maximally at valence extremes.

We tested two competing models of amygdala function. On a valence-independent hypothesis, amyg-dala response to intensity is similar at all levels of valence. We contrasted this with a valence-dependent model in which the amygdala is sensitive to intensity only at the outer bounds of valence. Using event-related functional magnetic resonance imaging, we then measured amygdala responses to high- and low-concentration variants of pleasant, neutral and unplea-sant odours. Our key finding was that amygdala exhibits an intensity–valence interaction in olfactory processing (figure 1). Put simply, the effect of intensity on amygdala activity is not the same across all levels of valence and amygdala responds differentially to high (versus low)-intensity odour for pleasant and unplea-sant smells, but not for neutral smells (Winston *et al.* 2005). This finding indicates that the amygdala codes neither intensity nor valence *per se*, but an interaction between intensity and valence, a combination we suggest reflects the overall emotional value of a stimulus. This suggestion is in line with more general theories of amygdala function which suggest that this structure contributes to encoding of salient events that are likely to invoke action (Whalen *et al.* 2004).

2. FLEXIBLE LEARNING OF STIMULUS–REWARD ASSOCIATIONS

Associative learning provides a phylogenetically highly conserved means to predict future events of value, such as the likelihood of food or danger, on the basis of predictive sensory cues. A key contribution of the amygdala to emotional processing relates to its role in acquiring associative or predictive information. In Pavlovian conditioning, a previously neutral item

(the conditioned stimulus or CS+) acquires predictive significance by pairing with a biologically salient reinforcer (the unconditioned stimulus or UCS). An example of this type of associative learning is a study where we scanned 13 healthy, hungry subjects using fMRI, while they learnt an association between arbitrary visual cues and two pleasant food-based olfactory rewards (vanilla and peanut butter), both before and after selective satiation (Cahill *et al.* 1996). Arbitrary visual images comprised the two conditioned stimuli (target and non-target CS+) that were paired with their corresponding UCS on 50% of all trials, resulting in paired (CS+p) and unpaired (CS+u) event types, enabling us to distinguish learning-related responses from sensory effects of the UCS (Cahill *et al.* 1994). A third visual image type was never paired with odour (the non-conditioned stimulus or CS−). One odour was destined for reinforcer devaluation (target UCS), while the other odour underwent no motivational manipulation (non-target UCS) (Gottfried *et al.* 2003).

Olfactory associative learning engaged amygdala, rostromedial orbitofrontal cortex (OFC), ventral midbrain, primary olfactory (piriform) cortex, insula and hypothalamus, highlighting the involvement of these regions in acquiring picture–odour contingencies. While these contingencies enable the generation of expectancies in response to sensory cues, it is clear that these predictions have limitations in optimizing future behaviour. The value of the states associated with predictive cues can change in the absence of pairing with these cues, for example when the physiological state of the organism changes. Consequently, it is important for an organism to have a capability of maintaining accessible and flexible representations of the current value of sensory–predictive cues.

Reinforcer devaluation offers an experimental methodology for dissociating among stored representations of value accessed by a CS+. For example, in animals, food value can be decreased by pairing a meal with a toxin. In humans, a more acceptable way of achieving the same end is through sensory-specific satiety, where the reward value of a food eaten to satiety is reduced (devalued) more than to foods not eaten. Animal studies of appetitive (reward-based) learning show that damage to amygdala and OFC interferes with the behavioural expression of reinforcer devaluation (Hatfield *et al.* 1996; Malkova *et al.* 1997; Gallagher *et al.* 1999; Baxter *et al.* 2000). We reasoned that if amygdala and OFC maintain representations of predictive reward value, then CS+-evoked neural responses within these regions should be sensitive to their current reward value and, by implication, to experimental manipulations that devalue a predicted reward. On the other hand, insensitivity to devaluation would indicate that the role of these areas relates more to associative learning that is independent of, or precedes linkage to, central representations of their reward value.

When we compared responses elicited by CS+ stimuli associated with a post-training devalued and non-devalued reward outcome, we found significant response decrements in left amygdala, and both rostral and caudal areas of OFC in relation to the predictive cue that signalled a devalued outcome (Gottfried *et al.* 2003).

Thus, activity in the amygdala and OFC showed a satiety-related decline for target CS+u activity but remained unchanged for the non-target CS+u activity, paralleling the behavioural effects of satiation. Conversely, satiety-sensitive neural responses in ventral striatum, insular cortex and anterior cingulate exhibited a different pattern of activity, reflecting decreases to the target CS+u, which contrasted with increases to the non-target CS+u (figure 2). Thus, amygdala and OFC activities evoked by the target CS+u decreased from pre- to post-satiety in a manner that paralleled the concurrent reward value of the target UCS. This response pattern within amygdala and OFC suggests that these regions are involved in representing reward value of predictive stimuli in a flexible manner, observations that accord with animal data demonstrating that amygdala and OFC lesions impair the effects of reinforcer devaluation (Hatfield *et al.* 1996; Malkova *et al.* 1997; Gallagher *et al.* 1999; Baxter *et al.* 2000).

A question raised by these findings is whether satiety-related devaluation effects reflect the ability of a CS+ to access UCS representations of reward value. Evidence in support of this is our finding that devaluation effects are expressed in the same regions that encode representations of the odour UCS (figure 2). Similarly, brain regions that encode predictive reward value participated in the initial acquisition of stimulus–reward contingencies (Gottfried *et al.* 2003) as evidenced by common foci of CS+-evoked responses at initial learning, and during reinforcer devaluation, in amygdala and OFC. The inference from these data is that brain regions maintaining representations of predictive reward are a subset of those that actually participate in associative learning.

Computational models of reward learning postulate motivational 'gates' that facilitate information flow between internal representations of CS+ and UCS stimuli (Dayan & Balleine 2002; Ikeda *et al.* 2002). These are the targets of motivational signals and determine the likelihood that stimulus–reward associations activate appetitive systems. The observation that neural responses evoked by a CS+ in amygdala and OFC are directly modulated by hunger states indicates that these regions underpin Pavlovian incentive behaviour in a manner that accords with specifications of a motivation gate. These findings also inform an understanding of the impact of pathologies within mediotemporal and basal orbitofrontal lobes. Damage to these regions causes a wide variety of maladaptive behaviours. Defective encoding of (or impaired access to) updated reward value in amygdala and OFC could explain the inability of such patients to modify their responses when expected outcomes change (Rolls *et al.* 1981). Thus, these findings can potentially explain the feeding abnormalities observed in both the Kluver–Bucy syndrome (Bechara *et al.* 1994) and frontotemporal dementias (Terzian & Ore 1955). Patients with these conditions may show increased appetite, indiscriminate eating, food cramming, change in food preference, hyperorality and even attempts to eat non-food items. Our data suggest that, in these pathologies, food cues no longer evoke updated representations of their reward value. A disabled anatomical network involving OFC and amygdala
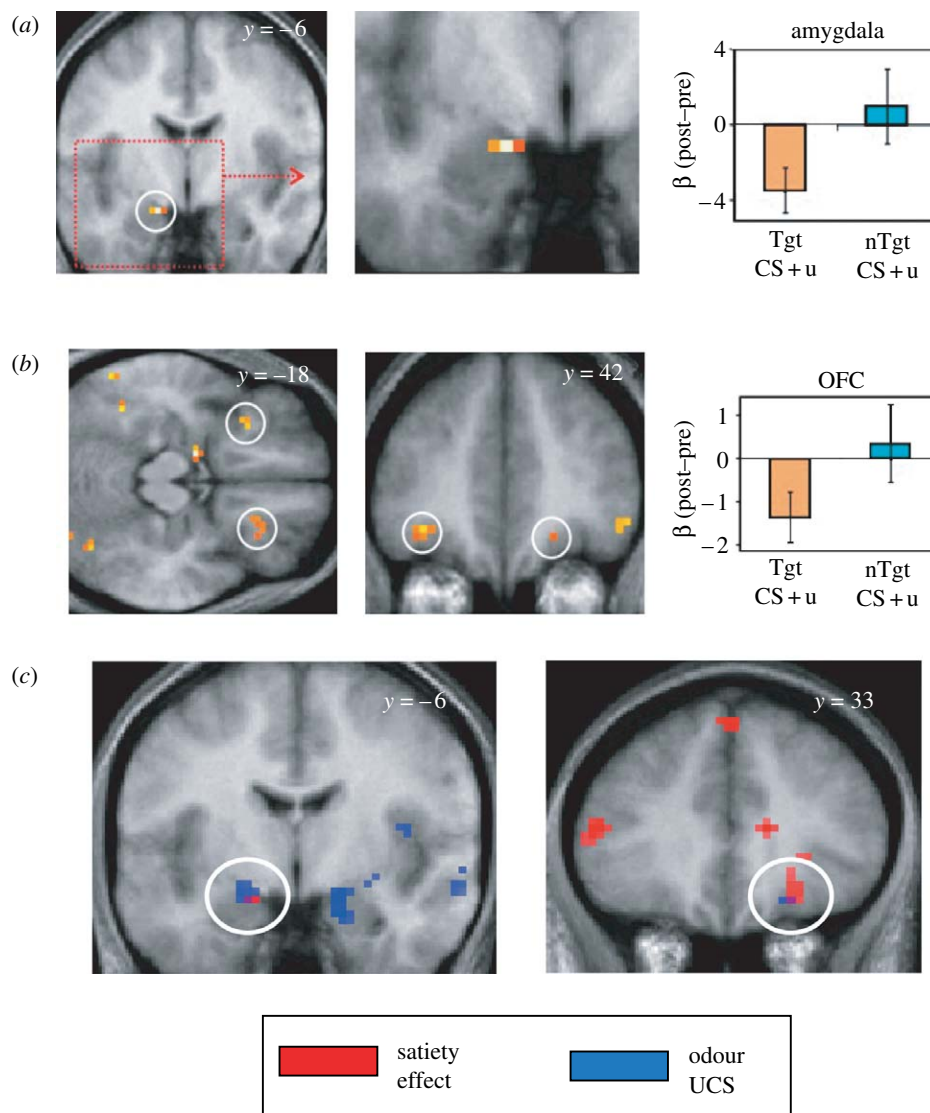
Figure 2. (*a*) Dorsal amygdala region showing altered response to a predictive stimulus that was devalued (CS + Tgt) compared with a control non-devalued stimulus (CS + nTgt). (*b*) Region of orbital prefrontal cortex showing altered response to a predictive stimulus that was devalued (CS + Tgt) compared with a control non-devalued stimulus (CS + nTgt). (*c*) Areas of overlap in amygdala and OFC for responses to UCS (blue) and devaluation (satiety) effects (red). Tgt, target; nTgt, non-target.

would result in food cues being unable to recruit motivationally appropriate representations of food-based reward value.

3. EMOTIONAL LEARNING EVOKES A TEMPORAL DIFFERENCE TEACHING SIGNAL

There are a number of theoretical accounts of emotional learning; among them the most notable have been based on the Rescorla–Wagner rule (Rescorla 1972). These models, and their real-time extensions, provide some of the best descriptions of computational processes underlying associative learning (Sutton & Barto 1981). The characteristic teaching signal within these models is the prediction error, which is used to direct acquisition and refine expectations relating to cues. In simple terms, a prediction error records change in an expected affective outcome and is expressed whenever predictions are generated, updated or violated. The usefulness of a computational approach in fostering an understanding of biological processes ultimately rests on an empirical test of the degree to which these processes are approximated in real biological systems (Sutton & Barto 1981).

To determine whether emotional learning is implemented using a temporal difference learning algorithm, we used fMRI to investigate the brain activity in healthy subjects as they learned to predict the occurrence of phasic relief or exacerbations of background tonic pain, using a first-order Pavlovian conditioning procedure with a probabilistic (50%) reinforcement schedule (Seymour *et al.* 2005). Tonic pain was induced using the capsaicin thermal hyperalgesia model, while visual cues (abstract coloured images) acted as Pavlovian-conditioned stimuli such that subjects learned that certain images tended to predict either imminent relief or exacerbation of pain. We implemented a computational learning model, the temporal difference (TD) model, which generates a prediction error signal, enabling us to identify brain responses that correlated with this signal. The temporal difference model learns the predictive values (expectations) of neutral cues by assessing their previous associations with appetitive or aversive outcomes.

We treated relief of pain as a reward and exacerbation of pain as a negative reward. Usage of the temporal difference algorithm to represent positive and negative deviations of pain intensity from a tonic background level approximates a class of reinforcement learning model termed average-reward models (Price 1999; O'Doherty *et al.* 2001; Tanaka *et al.* 2004). In these models, predictions are judged relative to the average level of pain, rather than some absolute measures; a comparative approach consistent with both neurobiological and economic accounts of homeostasis that rely crucially on change in affective state (Small *et al.* 2001; Craig 2003). The individual sequence of stimuli given to each subject provides a means to calculate both the value of expectations (in relation to the cues) and the prediction error as expressed at all points throughout the experiment. The use of a partial reinforcement strategy, in which the cues are only 50% predictive of their outcomes, ensures constant learning and updating of expectations, generating both positive and negative prediction errors throughout the course of the experiment.

Using a TD model, as described earlier, to index a representation of an appetitive (reward/relief) prediction error in the brain, we found that activity in left amygdala and left midbrain, a region encompassing the substantia nigra, correlated with this prediction error signal. Time-course analysis of the average pattern of response associated with the different trial types in this area showed a strong correspondence with the average pattern of activity predicted by the model. These data allow two general inferences. Firstly, the functions of the amygdala are not only confined to learning about aversive events, but also reflect learning about rewarding events. Secondly, predictive learning in the amygdala involves a neuronal signature that accords with the outputs of a TD computational model involving a prediction error teaching signal. This suggests that emotional learning in the amygdala involves implementation of a TD-like learning algorithm.

4. LEARNING TO AVOID DANGER

Predictive learning may be temporally specific and a cue that predicts danger at one time may not predict danger at a subsequent future time. It would clearly be maladaptive if an organism continued to be governed by these earlier predictions. How cues that no longer signal threat are disregarded can be studied using extinction paradigms. During extinction, successive presentations of a non-reinforced CS+ (following conditioning) diminish conditioned responses (CRs). Animal research indicates that extinction is not simply unlearning an original contingency, but evokes new learning that opposes, or inhibits, expression of conditioning (Rescorla 2001; Myers & Davis 2002). This account proposes that extinction leads to the formation of two distinct memory representations: a 'CS : UCS' excitatory memory and a 'CS : no UCS' inhibitory memory. Which competing memory is activated by a given CS+ is influenced by a range of contingencies including sensory, environmental and temporal contexts (Bouton 1993; Garcia 2002; Hobin *et al.* 2003).

In rodent models of extinction, both ventral prefrontal cortex (PFC) (White & Davey 1989; Repa *et al.* 2001; Milad & Quirk 2002) and amygdala (Herry & Garcia 2002; Hobin *et al.* 2003; Quirk & Gehlert 2003) are implicated in extinction-related processes. The most interesting mechanistic account of extinction is a proposal that excitatory projections from medial PFC to interneurons in lateral amygdaloid nucleus (Mai *et al.* 1997; Rosenkranz *et al.* 2003) or neighbouring intercalated cell masses (Pearce & Hall 1980; Quirk *et al.* 2003) gate excitatory impulses into the central nucleus of the amygdala in a manner that attenuates the expression of CRs.

To assess how extinction learning is expressed in the human brain, we measured region-specific brain activity in human subjects who had undergone olfactory aversive conditioning (Gottfried & Dolan 2004). This involved pairing two CS+ faces repetitively with two different UCS odours, while two other faces were never paired with odour and acted as non-conditioned control stimuli (CS-1 and CS-2). This manipulation was followed in the same session by extinction, permitting direct comparison between neural responses evoked during conditioning and extinction learning. In a further manipulation, we used a revaluation procedure to alter, post-conditioning and pre-extinction, the value of one target odour reinforcer using UCS inflation, resulting in target and non-target CS+ stimuli (Falls *et al.* 1992; Critchley *et al.* 2002). In effect, we presented subjects a more intense and aversive exemplar of this UCS, a manipulation that enabled us to tag this CS : UCS memory and index its persistence during the extinction procedure.

We again showed that neural substrates associated with learning involved the amygdala (Gottfried & Dolan 2004) with additional activations seen in ventral midbrain, insula, caudate and ventral striatum, comprising structures previously implicated in associative learning (Morgan *et al.* 1993; Gottfried *et al.* 2002; O'Doherty *et al.* 2002). The crucial finding in this study was our observation that significant activations in rostral and caudal OFC, ventromedial PFC (VMPFC) and lateral amygdala were evident during extinction. Strikingly, these findings implicate similar regions highlighted in animal studies of extinction learning, with the caveat that there are difficulties in identifying homologies between human and rodent models (White & Davey 1989; Repa *et al.* 2001; Herry & Garcia 2002; Milad & Quirk 2002; Hobin *et al.* 2003; Quirk & Gehlert 2003).

One obvious question that arises from these data is whether neural substrates of extinction learning overlap those involved in acquisition? Areas mutually activated across both conditioning and extinction contexts included medial amygdala, rostromedial OFC, insula, and dorsal and ventral striatum. In comparison, a direct contrast of extinction−conditioning allowed us to test for functional dissociations between these sessions. This analysis indicated that neural responses in lateral amygdala, rostromedial OFC and hypothalamus were preferentially enhanced during extinction learning, over and above any conditioning-evoked activity for both CS+ types. These peak activations occurred in the absence of significant interactions

between phase (conditioning versus extinction) and CS+ type (target versus non-target).

The above findings do not enable a distinction between CS+-evoked activation of UCS memory traces and those related more generally to extinction learning. The fact that we used UCS inflation to create an updated trace of UCS value meant it could be selectively indexed during extinction. The logic here is that CRs subsequently elicited by the corresponding CS+ should become accentuated, as the predictive cue accesses an updated and inflated representation of UCS value. The contrast of target CS+u versus non-target CS+u at extinction (each minus their respective CS−) demonstrated a significant activity in left lateral OFC, with an adjacent area of enhanced activity evident when we examined the interaction between phase (extinction versus conditioning) and CS+ type (target versus non-target). Furthermore, a significant positive correlation was evident between lateral OFC activity and ratings of target CS+ aversiveness, implying that relative magnitude of predictive (aversive) value is encoded, and updated, within this structure. Note that as UCS inflation enhanced target CS+ aversiveness, the non-target CS+ concurrently became less aversive, relative to the target CS+. In this respect, post-inflation value of the non-target UCS became relatively more rewarding. Consequently, when we compared non-target and target CS+ activities at extinction (minus CS− baselines) with index areas sensitive to predictive reward value (i.e., relatively 'less aversive' value), we found significant VMPFC activity driven by the non-target CS+u response at extinction. Regression analysis demonstrated that neural responses in ventromedial PFC were significantly and negatively correlated with differential CS+ aversiveness.

These findings indicate that discrete regions of OFC, including lateral/medial and rostral sectors, as well as lateral amygdala are preferentially activated during extinction learning. However, these findings cannot be attributed to general mechanisms of CS+ processing, as extinction-related activity was selectively enhanced in these areas over and above that evoked during conditioning. Thus, CS+-evoked recruitment of an OFC–amygdala network provides the basis for memory processes that regulate expression of conditioning. Our findings suggest that ventral PFC supports dual mnemonic representations of UCS value, which are accessible to a predictive cue. The presence of a dual representational system that responds as a function of the degree of preference (or non-preference) could provide a basis for fine-tuned regulation over conditioned behaviour and other learned responses. Indeed, an organism that needs to optimize its choices from among a set of different predictive cues would be well served by a system that integrates information about their relative values in such a parallel and differentiated manner. The general idea that orbital PFC synthesizes sensory, affective and motivational cues in the service of goal-directed behaviour also accords well with animal (Tremblay & Schultz 1999; Arana et al. 2003; Pickens et al. 2003; Schoenbaum et al. 2003) and human (Morgan et al. 1993; Gottfried et al. 2003) studies of associative learning and incentive states.

Medial–lateral dissociations in ventral PFC activity have been described in the context of a diverse set of rewards and punishments. Pleasant and unpleasant smells (Hatfield et al. 1996; O'Doherty et al. 2003) and tastes (Malkova et al. 1997), as well as more abstract valence representations (Gallagher et al. 1999; Gottfried et al. 2002), all exhibit functional segregation along this axis. On neuroanatomical grounds, these regions can be regarded as distinct functional units with unique sets of cortical and subcortical connections (Baxter et al. 2000). Notably, projections between OFC and amygdala are reciprocal (Baxter et al. 2000; Morris & Dolan 2004) and it is thus plausible that differences in input patterns from amygdala might contribute to the expression of positive and negative values in medial and lateral prefrontal subdivisions, respectively.

5. CONTEXTUAL CONTROL IN THE EXPRESSION OF EMOTIONAL MEMORY

The idea that extinction represents a form of new learning, while leaving intact associations originally established during conditioning, receives strong support from animal studies. As we have seen, the VMPFC is strongly implicated in the storage and recall of extinction memories (Morgan & LeDoux 1995; Milad & Quirk 2002; Phelps et al. 2004; Milad et al. 2005) and this region may exert control in conditioned memory expression via suppression of the amygdala (Quirk et al. 2003; Rosenkranz et al. 2003).

Conditioning has usually been discussed within a framework of associative models, but an alternative perspective invokes the concept of decision making (Wasserman & Miller 1997; Gallistel & Gibbon 2000). Consequently, extinction is proposed to reflect a decision mechanism designed to detect change in the rate of reinforcement attributed to a CS where recent experience is inconsistent with earlier experience. Background context is a critical regulatory variable in this decision process (Bouton 2004) such that, with extinction training, the subsequent recall of an extinction memory with CS presentation (i.e. the CR) shows a relative specificity to contexts that resemble those present during extinction training ('extinction context').

We used context-dependency in recall of extinction memory to study its neurobiological underpinnings in human subjects using a within-subject AB–AB design consisting of Pavlovian fear conditioning in context A and extinction in context B on day 1 with testing of CS-evoked responses in both a conditioning (A) and extinction (B) context on day 2 (delayed recall of extinction) (Kalisch et al. 2006). The two CSs (CS+, which was occasionally paired with the UCS, and a CS−, which was never paired with a UCS) consisted of one male and one female face, while contexts were distinguished by background screen colour and auditory input. Conditioned fear responses were extinguished by presenting the CSs in the same fashion as during conditioning but now omitting shock. We reasoned that areas supporting context-dependent recall of extinction memory would show a contextual modulation of CS+-evoked activation. We examined this by testing for a categorical CS–context interaction $(CS+ > CS-)_B > (CS+ > CS-)_A$ on day 2 (delayed recall of extinction).

We found a significant interaction in OFC and left anterior hippocampus driven by a relatively greater activation to the CS + than to the CS − in the extinction (figure 3). Thus, CS +-evoked activation of OFC and (left anterior) hippocampus was specifically expressed in an extinction context. The hippocampus is known to process contextual information supporting recall of memory (Delamater 2004) and our data suggest that it uses this information to confer (extinction) context-dependency on CS +-evoked VMPFC activity. Thus, contextually regulated recall of extinction memory in humans seems to be mediated by a network of brain areas including the OFC and the anterior hippocampus. Our data indicate that these regions form a neurobiological substrate for the context-dependence of extinction recall (Bouton 2004). The finding of extinction context-specific relative activations in our study, as opposed to the extinction-related deactivations observed previously by LaBar & Phelps (2005), supports the general idea that extinction (and its recall) is not simply a process of forgetting the CS–UCS association, but consists of active processes that encode and retrieve a new CS–no UCS memory trace (Myers & Davis 2002; Bouton 2004; Delamater 2004).

6. INSTRUMENTAL BEHAVIOUR AND VALUE REPRESENTATIONS

Optimal behaviour relies on using past experience to guide future decisions. In behavioural economics, expected utility (a function of probability, magnitude and delay to reward) provides a guiding perspective on decision making (Camerer 2003). Despite limitations, utility theory of Neumann–Morgenstern (Loomes 1988) continues to dominate models of decision making and converges with reinforcement learning on the idea that decision making involves integration of reward, reward magnitude and reward timing to provide a representation of action desirability (Glimcher & Rustichini 2004). On this basis, a key variable in optimal decision, particularly under conditions of uncertainty, reflects the use of information regarding the likely value of distinct courses of actions. As we have already seen, there is good evidence from studies of associate learning that OFC is involved in representing value and in updating value representations in a flexible manner.

A key question that arises is whether such representations guide more instrumental-type actions where reward values of options for action are unknown or can only be approximated. Classically, these situations pose a conflict between exploiting what is estimated to be the current best option versus sampling an uncertain, but potentially more rewarding, alternative (see also Cohen *et al.* 2007). This scenario is widely known as the explore–exploit dilemma. We studied this class of decision making while subjects performed an *n*-armed bandit task with four slots that paid money as reward (Daw *et al.* 2006). Simultaneous data on brain responses were acquired using fMRI. Pay-offs for each slot varied from trial to trial around a mean value corrupted by Gaussian noise. Thus, information regarding the value of an individual slot can only be obtained by active sampling. As the values of each of the actions cannot be determined from single outcomes, subjects need to optimize their choices by exploratory sampling of each of the slots in relation to a current estimate of what is the optimal or greedy slot.

We characterized subjects' exploratory behaviour by examining a range of reinforcement learning models of exploration, where the best approximation was what is known as a softmax rule. A softmax solution implies that actions are chosen as a ranked function of their estimated value. This rule ensures that the action with the highest value is still selected in an exploitative manner, with other actions chosen with a frequency that reflects a ranked estimate of their value. Using the softmax model, we could calculate value predictions, prediction errors and choice probabilities for each subject on each trial. These regressors were then used to identify brain regions where activity was significantly correlated with the model's internal signals.

Implementing this computational approach to our fMRI data analysis demonstrated that activity in medial OFC correlated with the magnitude of the obtained pay-off, a finding consistent with our previous evidence that this region codes the relative value of different reward stimuli, including abstract rewards (O'Doherty *et al.* 2001; O'Doherty 2004). Furthermore, activity in medial and lateral OFC, extending into VMPFC, correlated with the probability assigned by the model to the action actually chosen trial-to-trial. This probability provides a relative measure of the expected reward value of the chosen action, and the associated profile of activity is consistent with a role for orbital and adjacent medial PFC in encoding predictions of future reward as indicated in our studies of devaluation (Gottfried *et al.* 2003; Tanaka *et al.* 2004). Thus, these data accord with a general framework wherein action choice is optimized by accessing the likely future reward value of chosen actions.

A crucial aspect of our model is that it affords a characterization of neural activity as exploratory as opposed to exploitative. Consequently, we classified subjects' behaviour according to whether the actual choice was one predicted by the model to be determined by the dominant slot machine with the highest expected value (exploitative) or a dominated machine with a lower expected value (exploratory). Comparing the pattern of brain activity associated with these exploratory and exploitative trials showed right anterior frontopolar cortex (BA 10) as more active during decisions classified as exploratory (figure 4). This anterior frontopolar cortex activity indicates that this region provides a control mechanism facilitating switching between exploratory and exploitative strategies. Indeed, this fits with what is known about the role of this most rostral prefrontal region in high-level control (Ramnani & Owen 2004), mediating between different goals, subgoals (Braver & Bongiolatti 2002) or cognitive processes (Ramnani & Owen 2004; see also Burgess *et al.* 2007)

The conclusions from these data are that choices under uncertainty are strongly correlated with activity in the OFC, emphasizing the key role played by this structure in behavioural control. Within the context of an instrumental task, activity in the OFC encodes the value, and indeed relative value, of individual

(a) (i)

LVMPFC

(ii)

left anterior
hippocampus

(b) (i)

LVMPFC

(ii)

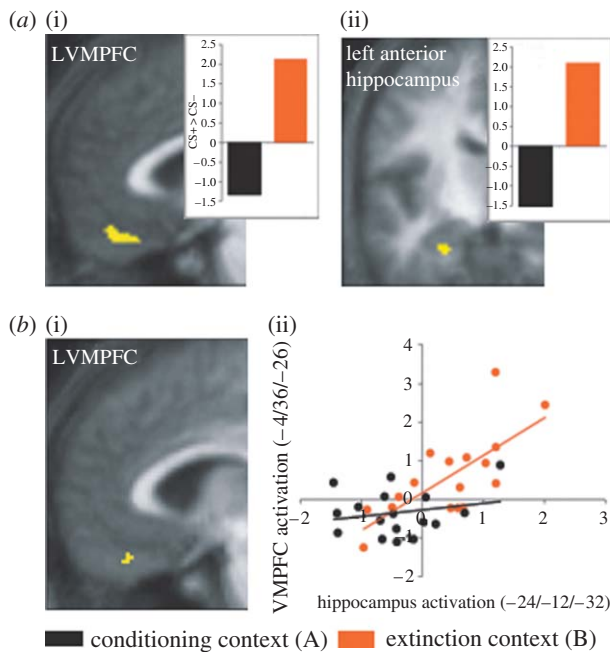conditioning context (A)   extinction context (B)

Figure 3. (a) Recall of extinction memory: figure (i) displays a region of orbital prefrontal cortex (left ventromedial prefrontal cortex: LVMPFC) showing enhanced activity during contextual recall of extinction memory (in other words signalling that a sensory stimulus that had earlier predicted an aversive event no longer predicts its occurrence). Figure (ii) shows a region of left hippocampus where there is also enhanced activity during extinction memory recall. (b) Context-specific correlation: (i and ii) region of OFC where there is a significant correlation in activity with hippocampal regions (as highlighted in (a)) that is expressed solely in an extinction context.

actions. More rostral PFC (BA 10), activated during exploratory choices, would seem to represent a control region that can override decisions based solely on value to allow a less deterministic sampling of the environment and possibly reveal a richer seam of rewards.

7. EMOTIONAL BIASES ON DECISION MAKING

Decision-making theory emphasises the role of analytic processes based upon utility maximisation, which incorporates the concept of reward, in guiding choice behaviour. There is now good evidence that more intuitive, or emotional responses, play a key role in human decision-making (Damasio *et al.* 1994; Loewenstein *et al.* 2001; Greene & Haidt 2002). In particular, decisions under conditions when available information is incomplete or overly complex can invoke simplifying heuristics, or efficient rules of thumb, rather than extensive algorithmic processing (Gilovich *et al.* 2002). Deviations from predictions of utility theory can, in some instances, be explained by emotion, as proposed by disappointment (Bell 1985; Loomes & Sugden 1986) and regret theory (Bell 1982; Loomes & Sugden 1983). Disappointment is an emotion that occurs when an outcome is worse than an outcome one would have obtained under a different state of reward. Regret is the emotion that occurs when an outcome is worse than what one would have experienced, had one made a different choice. It has been shown that an ability to anticipate emotions, such as disappointment or regret, has consequences for

future options for action and profoundly influences our decisions (Mellers *et al.* 1999). In other words, when faced with mutually exclusive options, the choice we make is influenced as much by what we hope to gain (expected value or utility) as by how much we anticipate we will feel after the choice (Payne *et al.* 1992).

Counterfactual thinking is a comparison between obtained and unattained outcomes that determines the quality and intensity of the ensuing emotional response. Neuropsychological studies have shown that this effect is abolished by lesions to OFC (Camille *et al.* 2004). A cumulative regret history can also exert a biasing influence on the decision process, such that subjects are biased to choose options likely to minimize future regret, an effect mediated by the OFC and amygdala (Coricelli *et al.* 2005). Theoretically, these findings can be seen as approximating a forward model of choice that incorporates predictions regarding future emotional states as inputs into decision processes.

While there is now good evidence that emotion can bias decision making, there were other striking instances of the effects of emotion on rationality. A key assumption in rational decision making is logical consistency across decisions, regardless of how choices are presented. This assumption of description invariance (Tversky & Kahneman 1986) is now challenged by a wealth of empirical data (McNeil *et al.* 1982; Kahneman & Tversky 2000), most notably in the 'framing effect', a key component within Prospect Theory (Kahneman & Tversky 1979; Tversky & Kahneman 1981). One theoretical consideration is that the framing effect results from a systematic bias in choice behaviour arising from an affect heuristic underwritten by an emotional system (Slovic *et al.* 2002, Gabaix, 2003 #4968).

We investigated the neurobiological basis of the framing effect using fMRI and a novel financial decision-making task. Participants were shown a message indicating the amount of money that they would initially receive in that trial (e.g. 'You receive £50'). Subjects then had to choose between a 'sure' or a 'gamble' option presented in the context of two different 'frames'. The sure option was formulated as either the amount of money retained from the initial starting amount (e.g. keep £20 of a total of £50—'Gain' frame), or as the amount of money lost from the initial amount (e.g. lose £30 of a total of £50—'Loss' frame). The gamble option was identical in both frames and represented as a pie chart depicting the probability of winning or losing (De Martino *et al.* 2006).

Subjects' behaviour in this task showed a marked framing effect evident in being risk-averse in the Gain frame, tending to choose the sure option over the gamble option (gambling on 42.9% trials; significantly different from 50%, $p < 0.05$), and risk-seeking in the Loss frame, preferring the gamble option (gambling on 61.6% trials; significantly different from 50%). This effect was consistently observed across different probabilities and initial endowment amounts. During simultaneous acquisition of fMRI data on regional brain activity, we observed bilateral amygdala activity when subjects' choices were influenced by the frame. In other words, amygdala activation was significantly
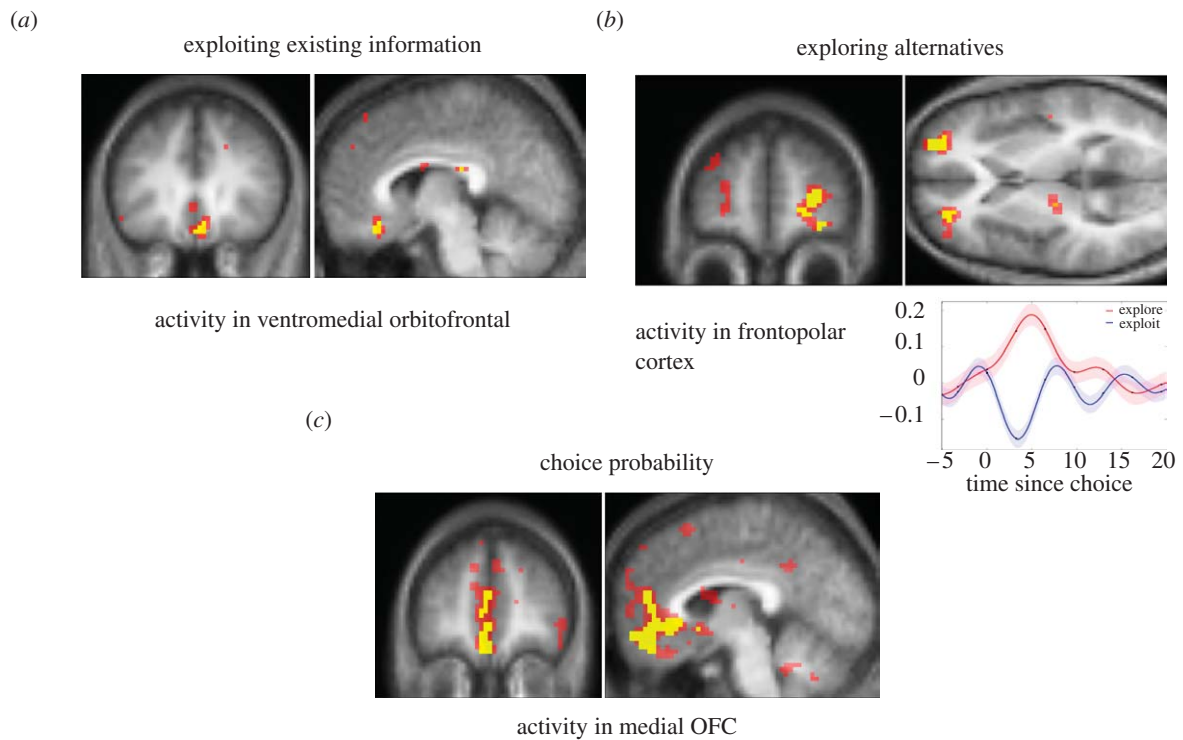
Figure 4. (*a*) Regions of ventromedial prefrontal cortex where activity increases as a function of actual reward received. (*b*) Regions of left and right frontopolar cortex where activity increases during trials that are exploratory. (*c*) Regions of medial and lateral OFC and adjacent medial prefrontal cortex where activity correlates with the probability assigned by the computational model to subject's choice of slot.

greater when subjects chose the sure option in the Gain frame (G_sure–G_gamble), and the gamble option in the Loss frame (L_gamble–L_sure). When subjects made choices that ran counter to this general behavioural tendency, there was enhanced activity in anterior cingulate cortex (ACC), suggesting an opponency between two neural systems. Activation of ACC in these situations is consistent with the detection of conflict between more 'analytic' response tendencies and an obligatory effect associated with a more 'emotional' amygdala-based system (Botvinick *et al.* 2001; Balleine & Killcross 2006).

A striking feature of our behavioural data was a marked intersubject variability in susceptibility to the frame. This variability allowed us to index subject-specific differences in neural activity associated with a decision by frame interaction. Using a measure of overall susceptibility of each subject to the frame manipulation, we constructed a 'rationality index' and found a significant correlation between decreased susceptibility to the framing effect and enhanced activity in orbital and medial PFC (OMPFC) and VMPFC. In other words, subjects who acted more rationally exhibited greater activation in OMPFC and VMPFC associated with the frame effect.

We have already seen that the amygdala plays a key role in value-related prediction and learning, both for negative (aversive) and positive (appetitive) outcomes (LeDoux 1996; Baxter & Murray 2002; Seymour *et al.* 2005). Furthermore, in simple instrumental decision-making tasks in animals, the amygdala appears to mediate biases in decision that come from value-related predictions (Paton *et al.* 2006). In humans, the amygdala is also implicated in the detection of

emotionally relevant information present in contextual and social emotional cues. Increased activation in amygdala associated with subjects' tendency to be risk-averse in the Gain frame and risk-seeking in the Loss frame supports the hypothesis that the framing effect is driven by an affect heuristic underwritten by an emotional system.

The observation that the frame has such a pervasive impact on complex decision making supports the emerging central role for the amygdala in decision making (Kim *et al.* 2004; Hsu *et al.* 2005). These data extend the role of the amygdala to include processing contextual positive or negative emotional information communicated by a frame. Note that activation of amygdala was driven by the combination of a subject's decision in a given frame, rather than by the valence of the frame *per se*. It would seem that frame-related valence information is incorporated into the relative assessment of options to exert control over the apparent risk sensitivity of individual decisions.

An intriguing question is why is the frame so potent in driving emotional responses that engender deviations from rationality? Information about motivationally important outcomes may come from a variety of sources, not only from those based on analytic processes. In animals, the provision of cues that signal salient outcomes, for example, Pavlovian contingencies, can have a strong impact on ongoing instrumental actions. Intriguingly, interactions of the impact of Pavlovian cues on instrumental performance involve brain structures such as the amygdala. Similar processes may be involved in the 'framing effect' where an option for action is accompanied by non-contingent affective cues. Such affective cues ('frames')

invoke risk, typically either positive (you could win £x) or negative (you might lose £y), and may cause individuals to adjust how they value distinct options. From the perspective of economics, the resulting choice biases seem irrational, but in real-life decision-making situations, sensitivity to these cues may provide a valuable source of additional information.

Susceptibility to the frame showed a robust correlation with neural activity in OMPFC across subjects consistent with the idea that this region and the amygdala each contributes distinct functional roles in decision making. As already argued, the OFC, by incorporating inputs from the amygdala, represents the motivational value of stimuli (or choices), which allows it to integrate and evaluate the incentive value of predicted outcomes in order to guide future behaviour (Rolls *et al.* 1994; Schoenbaum *et al.* 2006). Lesions of the OFC cause impairments in decision making, often characterized as an inability to adapt behavioural strategies according to current contingencies (as in extinction) and consequences of decisions expressed in forms of impulsivity (Winstanley *et al.* 2004; Bechara *et al.* 1994). One interpretation of enhanced activation with increasing resistance to the frame is that more 'rational' individuals have a better and more refined representation of their own emotional biases. Such a representation would allow subjects to modify their behaviour appropriate to circumstances, as for example when such biases might lead to suboptimal decisions. In this model, OFC evaluates and integrates emotional and cognitive information that underpins more 'rational' behaviour, operationalized here as description-invariant.

8. CONCLUSIONS

The amygdala and orbital PFC have a pivotal role in emotional processing and guidance of human behaviour. The rich sensory connectivity of the amygdala provides a basis for its role in encoding predictive value for both punishments and rewards. Acquisition of value representations by the amygdala would appear to involve implementation of a TD-like reinforcement learning algorithm. On the other hand, the OFC appears to provide the basis for a more flexible representation of value that is sensitive to multiple environmental factors, including context and internal physiological state (e.g. state satiety). Flexible representation of value in OFC also provides a basis for optimization of behaviour on the basis of reward value accruing from individual choice behaviour. There are suggestions that OFC may integrate value over both short and long time frames (Cohen *et al.* 2007). An intriguing, though as yet unanswered, question is whether there are distinct regions within OFC that encode different aspects of value such as reward and punishment. Our findings indicate that a more rostral PFC region appears to exert control of action, enabling switching between actions that are exploitative and those that allow exploratory sampling of other options for action (see also Cohen *et al.* 2007). However, high-level behaviour is also susceptible to more low-level influences from the amygdala, perhaps mediated via low-level Pavlovian processing, and these influences can bias choice behaviour in a manner that is rationally suboptimal as seen in regret and in the framing effect.

REFERENCES

Anderson, A. K. & Sobel, N. 2003 Dissociating intensity from valence as sensory inputs to emotion. *Neuron* **39**, 581–583. (doi:10.1016/S0896-6273(03)00504-X)

Arana, F. S., Parkinson, J. A., Hinton, E., Holland, A. J., Owen, A. M. & Roberts, A. C. 2003 Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. *J. Neurosci.* **23**, 9632–9638.

Balleine, B. W. & Killcross, S. 2006 Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci.* **29**, 272–279. (doi:10.1016/j.tins.2006.03.002)

Baxter, M. G. & Murray, E. A. 2002 The amygdala and reward. *Nat. Rev. Neurosci.* **3**, 563–573. (doi:10.1038/nrn875)

Baxter, M. G., Parker, A., Lindner, C. C., Izquierdo, A. D. & Murray, E. A. 2000 Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. *J. Neurosci.* **20**, 4311–4319.

Bechara, A., Damasio, A. R., Damasio, H. & Anderson, S. W. 1994 Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* **50**, 7–15. (doi:10.1016/0010-0277(94)90018-3)

Bell, D. 1982 Regret and decision making under uncertainty. *Oper. Res.* **30**, 961–981.

Bell, D. E. 1985 Disappointment in decision making under uncertainty. *Oper. Res.* **33**, 1–27.

Bensafi, M., Rouby, C., Farget, V., Bertrand, B., Vigouroux, M. & Holley, A. 2002 Autonomic nervous system responses to odours: the role of pleasantness and arousal. *Chem. Senses* **27**, 703–709. (doi:10.1093/chemse/27.8.703)

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S. & Cohen, J. D. 2001 Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624–652. (doi:10.1037/0033-295X.108.3.624)

Bouton, M. E. 1993 Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychol. Bull.* **114**, 80–99. (doi:10.1037/0033-2909.114.1.80)

Bouton, M. E. 2004 Context and behavioral processes in extinction. *Learn. Mem.* **11**, 485–494. (doi:10.1101/lm.78804)

Braver, T. S. & Bongiolatti, S. R. 2002 The role of frontopolar cortex in subgoal processing during working memory. *Neuroimage* **15**, 523–536. (doi:10.1006/nimg.2001.1019)

Burgess, P. W., Gilbert, S. J. & Dumontheil, I. 2007 Function and localization within rostral prefrontal cortex (area 10). *Phil. Trans. R. Soc. B* **362**, 887–899. (doi:10.1098/rstb.2007.2095)

Cahill, L. & McGaugh, J. L. 1998 Mechanisms of emotional arousal and lasting declarative memory. *Trends Neurosci.* **21**, 294–299. (doi:10.1016/S0166-2236(97)01214-9)

Cahill, L., Prins, B., Weber, M. & McGaugh, J. L. 1994 Beta-adrenergic activation and memory for emotional events. *Nature* **371**, 702–704. (doi:10.1038/371702a0)

Cahill, L., Babinsky, R., Markowitsch, H. J. & McGaugh, J. L. 1995 The amygdala and emotional memory. *Nature* **377**, 295–296. (doi:10.1038/377295a0)

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., Wu, J. & McGaugh, J. L. 1996 Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proc. Natl Acad. Sci. USA* **93**, 8016–8021. (doi:10.1073/pnas.93.15.8016)

Camerer, C. F. 2003 *Behavioral gain theory.* New Jersey, NJ: Princeton University Press.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J. R. & Sirigu, A. 2004 The involvement of the orbitofrontal cortex in the experience of regret. *Science* **304**, 1167–1170. (doi:10.1126/science.1094550)

Cohen, J., McLure, S. M. & Yu, A. 2007 Should I stay or should I go? How the human brain manages the tradeoff between exploration and exploitation. *Phil. Trans. R. Soc. B* **362**, 933–942. (doi:10.1098/rstb.2007.2098)

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A. & Dolan, R. J. 2005 Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. Neurosci.* **8**, 1255–1262. (doi:10.1038/nn1514)

Craig, A. D. 2003 A new view of pain as a homeostatic emotion. *Trends Neurosci.* **26**, 303–307. (doi:10.1016/S0166-2236(03)00123-1)

Critchley, H. D., Mathias, C. J. & Dolan, R. J. 2002 Fear conditioning in humans: the influence of awareness and autonomic arousal on functional neuroanatomy. *Neuron* **33**, 653–663. (doi:10.1016/S0896-6273(02)00588-3)

Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M. & Damasio, A. R. 1994 The return of Phineas Gage: clues about the brain from the skull of a famous patient. *Science* **264**, 1102–1105. (doi:10.1126/science.8178168) [Erratum in *Science* 1994 **265**, 1159.]

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. 2006 Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879. (doi:10.1038/nature04766)

Dayan, P. & Balleine, B. W. 2002 Reward, motivation, and reinforcement learning. *Neuron* **36**, 285–298. (doi:10.1016/S0896-6273(02)00963-7)

De Martino, B., Kumaran, D., Seymour, B. & Dolan, R. J. 2006 Frames, biases, and rational decision-making in the human brain. *Science* **313**, 684–687. (doi:10.1126/science.1128356)

Delamater, A. R. 2004 Experimental extinction in Pavlovian conditioning: behavioural and neuroscience perspectives. *Q. J. Exp. Psychol. B* **57**, 97–132. (doi:10.1080/02724990344000097)

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

Gallagher, M., McMahan, R. W. & Schoenbaum, G. 1999 Orbitofrontal cortex and representation of incentive value in associative learning. *J. Neurosci.* **19**, 6610–6614.

Gallistel, C. R. & Gibbon, J. 2000 Time, rate, and conditioning. *Psychol. Rev.* **107**, 289–344. (doi:10.1037/0033-295X.107.2.289)

Garcia, R. 2002 Postextinction of conditioned fear: between two CS-related memories. *Learn. Mem.* **9**, 361–363. (doi:10.1101/lm.56402)

Gilovich, T., Griffin, D. W. & Kahneman, D. 2002 *Heuristics and biases: the psychology of intuitive judgment*. New York, NY: Cambridge University Press.

Glimcher, P. W. & Rustichini, A. 2004 Neuroeconomics: the consilience of brain and decision. *Science* **306**, 447–452. (doi:10.1126/science.1102566)

Gottfried, J. A. & Dolan, R. J. 2004 Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nat. Neurosci.* **7**, 1144–1152. (doi:10.1038/nn1314)

Gottfried, J. A., O'Doherty, J. & Dolan, R. J. 2002 Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J. Neurosci.* **22**, 10 829–10 837.

Gottfried, J. A., O'Doherty, J. & Dolan, R. J. 2003 Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* **301**, 1104–1107. (doi:10.1126/science.1087919)

Greene, J. & Haidt, J. 2002 How (and where) does moral judgment work? *Trends Cogn. Sci.* **6**, 517–523. (doi:10.1016/S1364-6613(02)02011-9)

Hamann, S. 2003 Nosing in on the emotional brain. *Nat. Neurosci.* **6**, 106–108. (doi:10.1038/nn0203-106)

Hatfield, T., Han, J. S., Conley, M., Gallagher, M. & Holland, P. 1996 Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects. *J. Neurosci.* **16**, 5256–5265.

Herry, C. & Garcia, R. 2002 Prefrontal cortex long-term potentiation, but not long-term depression, is associated with the maintenance of extinction of learned fear in mice. *J. Neurosci.* **22**, 577–583.

Hobin, J. A., Goosens, K. A. & Maren, S. 2003 Context-dependent neuronal activity in the lateral amygdala represents fear memories after extinction. *J. Neurosci.* **23**, 8410–8416.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D. & Camerer, C. F. 2005 Neural systems responding to degrees of uncertainty in human decision-making. *Science* **310**, 1680–1683. (doi:10.1126/science.1115327)

Ikeda, M., Brown, J., Holland, A. J., Fukuhara, R. & Hodges, J. R. 2002 Changes in appetite, food preference, and eating habits in frontotemporal dementia and Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* **73**, 371–376. (doi:10.1136/jnnp.73.4.371)

Kahneman, D. & Tversky, A. 1979 Prospect theory: an analysis of decision under risk. *Econometrica* **47**, 263–291. (doi:10.2307/1914185)

Kahneman, D. & Tversky, A. 2000 *Choices, values, and frames*. New York, NY: Cambridge University Press.

Kalisch, R., Korenfeld, E., Stephan, K. E., Weiskopf, N., Seymour, B. & Dolan, R. J. 2006 Context-dependent human extinction memory is mediated by a ventromedial prefrontal and hippocampal network. *J. Neurosci.* **26**, 9503–9511. (doi:10.1523/JNEUROSCI.2021-06.2006)

Kim, H., Somerville, L. H., Johnstone, T., Polis, S., Alexander, A. L., Shin, L. M. & Whalen, P. J. 2004 Contextual modulation of amygdala responsivity to surprised faces. *J. Cogn. Neurosci.* **16**, 1730–1745. (doi:10.1162/0898929042947865)

LaBar, K. S. & Phelps, E. A. 2005 Reinstatement of conditioned fear in humans is context dependent and impaired in amnesia. *Behav. Neurosci.* **119**, 677–686. (doi:10.1037/0735-7044.119.3.677)

Lang, P. J. 1995 The emotion probe. Studies of motivation and attention. *Am. J. Psychol.* **50**, 372–385. (doi:10.1037/0003-066X.50.5.372)

LeDoux, J. E. 1996 *The emotional brain*. New York, NY: Simon and Schuster.

Loewenstein, G. F., Weber, E. U., Hsee, C. K. & Welch, N. 2001 Risk as feelings. *Psychol. Bull.* **127**, 267–286. (doi:10.1037/0033-2909.127.2.267)

Loomes, G. 1988 Further evidence of the impact of regret and disappointment in choice under uncertainty. *Econometrica* **55**, 47–62.

Loomes, G. & Sugden, R. 1983 A rationale for preference reversal. *Am. Econ. Rev.* **73**, 428–432.

Loomes, G. & Sugden, R. 1986 Disappointment and dynamic consistency in choice under uncertainty. *Rev. Econ. Stud.* **53**, 272–282.

Mai, J. K., Assheuer, J. & Paxinos, G. 1997 *Atlas of the human brain*. San Dieto, CA: Academie Press.

Malkova, L., Gaffan, D. & Murray, E. A. 1997 Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. *J. Neurosci.* **17**, 6011–6020.

McGaugh, J. L. & Roozendaal, B. 2002 Role of adrenal stress hormones in forming lasting memories in the brain. *Curr. Opin. Neurobiol.* **12**, 205–210. (doi:10.1016/S0959-4388 (02)00306-9)

McGaugh, J. L., Cahill, L. & Roozendaal, B. 1996 Involvement of the amygdala in memory storage: interaction with other brain systems. *Proc. Natl Acad. Sci. USA* **93**, 13 508–13 514. (doi:10.1073/pnas.93.24.13508)

McNeil, B. J., Pauker, S. G., Sox, H. C. & Tversky, A. 1982 On the elicitation of preferences for alternative therapies. *N. Engl. J. Med.* **306**, 1259–1262.

Mellers, B., Ritov, I. & Schwartz, A. 1999 Emotion-based Choice. *J. Exp. Psychol. General* **128**, 332–345. (doi:10. 1037/0096-3445.128.3.332)

Milad, M. R. & Quirk, G. J. 2002 Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* **420**, 70–74. (doi:10.1038/nature01138)

Milad, M. R., Quinn, B. T., Pitman, R. K., Orr, S. P., Fischl, B. & Rauch, S. L. 2005 Thickness of ventromedial prefrontal cortex in humans is correlated with extinction memory. *Proc. Natl Acad. Sci. USA* **102**, 10 706–10 711. (doi:10.1073/pnas.0502441102)

Morgan, M. A. & LeDoux, J. E. 1995 Differential contribution of dorsal and ventral medial prefrontal cortex to the acquisition and extinction of conditioned fear in rats. *Behav. Neurosci.* **109**, 681–688. (doi:10.1037/0735-7044.109.4.681)

Morgan, M. A., Romanski, L. M. & LeDoux, J. E. 1993 Extinction of emotional learning: contribution of medial prefrontal cortex. *Neurosci. Lett.* **163**, 109–113. (doi:10. 1016/0304-3940(93)90241-C)

Morris, J. S. & Dolan, R. J. 2004 Dissociable amygdala and orbitofrontal responses during reversal fear conditioning. *Neuroimage* **22**, 372–380. (doi:10.1016/j.neuroimage. 2004.01.012)

Myers, K. M. & Davis, M. 2002 Behavioral and neural analysis of extinction. *Neuron* **36**, 567–584. (doi:10.1016/ S0896-6273(02)01064-4)

O'Doherty, J. P. 2004 Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* **14**, 769–776. (doi:10.1016/j.conb.2004.10.016)

O'Doherty, J., Kringelbach, M. L., Hornak, J., Andrews, C. & Rolls, E. T. 2001 Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat. Neurosci.* **4**, 95–102. (doi:10.1038/82959)

O'Doherty, J. P., Deichmann, R., Critchley, H. D. & Dolan, R. J. 2002 Neural responses during anticipation of a primary taste reward. *Neuron* **33**, 815–826. (doi:10.1016/ S0896-6273(02)00603-7)

O'Doherty, J., Critchley, H., Deichmann, R. & Dolan, R. J. 2003 Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J. Neurosci.* **23**, 7931–7939.

Paton, J. J., Belova, M. A., Morrison, S. E. & Salzman, C. D. 2006 The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439**, 865–870. (doi:10.1038/nature04490)

Payne, J. W., Bettman, J. R. & Johnson, E. J. 1992 Behavioral decision research: a constructive processing pespective. *Annu. Rev. Psychol.* **43**, 87–131. (doi:10.1146/annurev.ps. 43.020192.000511)

Pearce, J. M. & Hall, G. 1980 A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552. (doi:10. 1037/0033-295X.87.6.532)

Phelps, E. A., LaBar, K. S., Anderson, A. K., O'Connor, K. J., Fulbright, R. K. & Spencer, D. D. 1998 Specifying the contributions of the human amygdala to emotional memory: a case study. *Neurocase* **4**, 527–540. (doi:10. 1093/neucas/4.6.527)

Phelps, E. A., Delgado, M. R., Nearing, K. I. & LeDoux, J. E. 2004 Extinction learning in humans: role of the amygdala and vmPFC. *Neuron* **43**, 897–905. (doi:10.1016/j.neuron. 2004.08.042)

Pickens, C. L., Saddoris, M. P., Setlow, B., Gallagher, M., Holland, P. C. & Schoenbaum, G. 2003 Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *J. Neurosci.* **23**, 11 078–11 084.

Price, D. D. 1999 *Psychological mechanisms of pain and analgesia*. Seattle, WA: IASP.

Quirk, G. J. & Gehlert, D. R. 2003 Inhibition of the amygdala: key to pathological states? *Ann. NY Acad. Sci.* **985**, 263–272.

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

Ramnani, N. & Owen, A. M. 2004 Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat. Rev. Neurosci.* **5**, 184–194. (doi:10.1038/nrn1343)

Repa, J. C., Muller, J., Apergis, J., Desrochers, T. M., Zhou, Y. & LeDoux, J. E. 2001 Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nat. Neurosci.* **4**, 724–731. (doi:10. 1038/89512)

Rescorla, R. A. 1972 "Configural" conditioning in discrete-trial bar pressing. *Q. J. Exp. Psychol. B* **79**, 307–317.

Rescorla, R. A. 2001 Experimental extinction. In *Handbook of contemporary learning theories* (eds R. R. Mowrer & S. Klein), pp. 119–154. Mahwah, NJ: Lawrence Erlbaum.

Rolls, B. J., Rolls, E. T., Rowe, E. A. & Sweeney, K. 1981 Sensory specific satiety in man. *Physiol. Behav.* **27**, 137–142. (doi:10.1016/0031-9384(81)90310-3)

Rolls, E. T., Hornak, J., Wade, D. & McGrath, J. 1994 Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J. Neurol. Neurosurg. Psychiatry* **57**, 1518–1524.

Rosenkranz, J. A., Moore, H. & Grace, A. A. 2003 The prefrontal cortex regulates lateral amygdala neuronal plasticity and responses to previously conditioned stimuli. *J. Neurosci.* **23**, 11 054–11 064.

Russell, J. 1980 A circumplex model of affect. *J. Pers. Soc. Psychol.* **39**, 1161–1178. (doi:10.1037/h0077714)

Schiffman, S. S. 1974 Physicochemical correlates of olfactory quality. *Science* **185**, 112–117. (doi:10.1126/science.185. 4146.112)

Schoenbaum, G., Setlow, B., Saddoris, M. P. & Gallagher, M. 2003 Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* **39**, 855–867. (doi:10.1016/S0896-6273(03)00474-4)

Schoenbaum, G., Roesch, M. R. & Stalnaker, T. A. 2006 Orbitofrontal cortex, decision-making and drug addiction. *Trends Neurosci.* **29**, 116–124. (doi:10.1016/j.tins.2005.12. 006)

Seymour, B., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K. & Dolan, R. 2005 Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nat. Neurosci.* **8**, 1234–1240. (doi:10.1038/nn1527)

Slovic, P., Finucane, M., Perers, E. & MacGregor, D. 2002 The affect heuristic. In *Heuristics and biases: the psychology of intuitive judgment* (eds T. Gilovich, D. W. Griffin & D. Kahneman), pp. 397–421. New York, NY: Cambridge University Press.

Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C. & Jones-Gotman, M. 2001 Changes in brain activity related to eating chocolate: from pleasure to aversion. *Brain* **124**, 1720–1733. (doi:10.1093/brain/124.9.1720)

Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M. & Parrish, T. 2003 Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* **39**, 701–711. (doi:10.1016/S0896-6273(03)00467-7)

Strange, B. A. & Dolan, R. J. 2004 Beta-adrenergic modulation of emotional memory-evoked human amygdala and hippocampal responses. *Proc. Natl Acad. Sci. USA* **101**, 11 454–11 458. (doi:10.1073/pnas.0404282101)

Strange, B. A., Hurlemann, R. & Dolan, R. J. 2003 An emotion-induced retrograde amnesia in humans is amygdala- and β-adrenergic-dependent. *Proc. Natl Acad. Sci. USA* **100**, 13 626–13 631. (doi:10.1073/pnas.1635116100)

Sutton, R. S. & Barto, A. G. 1981 Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* **88**, 135–170. (doi:10.1037/0033-295X.88.2.135)

Tanaka, S. C., Doya, K., Okada, G., Ueda, K., Okamoto, Y. & Yamawaki, S. 2004 Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat. Neurosci.* **7**, 887–893. (doi:10.1038/nn1279)

Terzian, H. & Ore, G. D. 1955 Syndrome of Kluver and Bucy; reproduced in man by bilateral removal of the temporal lobes. *Neurology* **5**, 373–380.

Tremblay, L. & Schultz, W. 1999 Relative reward preference in primate orbitofrontal cortex. *Nature* **398**, 704–708. (doi:10.1038/19525)

Tversky, A. & Kahneman, D. 1981 The framing of decisions and the psychology of choice. *Science* **211**, 453–458. (doi:10.1126/science.7455683)

Tversky, A. & Kahneman, D. 1986 Rational choice and the framing of decisions. *Business* **59**, 251–278. (doi:10.1086/296365)

van Stegeren, A. H., Everaerd, W., Cahill, L., McGaugh, J. L. & Gooren, L. J. 1998 Memory for emotional events: differential effects of centrally versus peripherally acting beta-blocking agents. *Psychopharmacology (Berlin)* **138**, 305–310. (doi:10.1007/s002130050675)

Wasserman, E. A. & Miller, R. R. 1997 What's elementary about associative learning? *Annu. Rev. Psychol.* **48**, 573–607. (doi:10.1146/annurev.psych.48.1.573)

Whalen, P. J. *et al.* 2004 Human amygdala responsivity to masked fearful eye whites. *Science* **306**, 2061. (doi:10.1126/science.1103617)

White, K. & Davey, G. C. 1989 Sensory preconditioning and UCS inflation in human 'fear' conditioning. *Behav. Res. Ther.* **27**, 161–166. (doi:10.1016/0005-7967(89)90074-0)

Winstanley, C. A., Theobald, D. E., Cardinal, R. N. & Robbins, T. W. 2004 Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. *J. Neurosci.* **24**, 4718–4722. (doi:10.1523/JNEUROSCI.5606-03.2004)

Winston, J. S., Gottfried, J. A., Kilner, J. M. & Dolan, R. J. 2005 Integrated neural representations of odor intensity and affective valence in human amygdala. *J. Neurosci.* **25**, 8903–8907. (doi:10.1523/JNEUROSCI.1569-05.2005)