It has been suggested that many aspects of reward-guided behaviour can be understood within the framework of a computational account of decision making. The account emphasises representation of expectations about decision outcomes and the revision of future expectations in the light of the prediction error—the discrepancy between the actual outcome and prior expectation. Frontal cortex and striatum are implicated in such processes in humans, monkeys, and rats suggesting they are ubiquitous and found in many species. Disagreement remains over the exact contribution made by each brain region. A growing body of work even suggests analogous processes may account for behaviour outside the domain of reward-guided decision making, for example, when people and animals learn about visual and social environments.

**Introduction**

The value of an action is determined by our expectation that it will lead to a beneficial outcome, such as the receipt of food or money. Over recent years, researchers have become increasingly interested in how the value of actions may be represented in the brain of humans or other animals. According to one influential account that has its roots in psychology [1] and machine learning [2], our knowledge of the value of actions is updated when our expectations about the ensuing reward are violated. For example, if the action was expected to have only a low probability of reward but, in the end, reward was actually delivered, then the action may be expected to have a higher value in the future. Conversely, if an action associated with high reward expectations is followed by no reward, then future expectations about its value may be revised downwards. The extent to which future expectations are revised in the light of any single discrepancy that is encountered varies. If the learning rate is high then even a single outcome may lead to a large revision in expectations about the action’s value but if the learning rate is low then future expectations about an action’s value will be based on the full history of previous outcomes, rather than just the most recent one.

Such reinforcement learning (RL) approaches have guided several recent investigations of the neural mechanisms of reward processing with varying degrees of consistency. At the same time other researchers have begun to question whether analogous processes to those envisaged by RL theory are at work not just when we learn about reward but also when we learn about the visual and social environments.

**The representation of value and reward expectations in frontal cortex**

There is not a single, unitary representation of value confined to a single brain area. For example, within the frontal lobe (Figure 1a, b), the orbitofrontal cortex (OFC) and anterior cingulate cortex (ACC) are concerned with representing the values of stimuli and of actions respectively. Rudebeck et al. [3] tested the association between both stimuli and rewards and actions and rewards using a variety of reward reversal and reward matching decision making tasks. While lesions centred on the lateral part of the OFC disrupted stimulus–reward association they had little impact on action–reward associations. ACC lesions had the opposite effect (Figure 1c, d). In rats [4], monkeys [5], and humans [5] neural systems located in more lateral and more medial frontal cortex respectively represent stimulus–reward and action–reward associations.

Within OFC there is evidence of functional differentiation between lateral and medial regions. The more medial region is often referred to as the ventromedial prefrontal cortex (VMPFC). It can be difficult to integrate OFC findings from humans and other animals. Animal studies often focused on more lateral orbital sectors while human brain imaging studies often concentrated on VMPFC. However, there are dramatic differences in how lateral OFC and VMPFC interconnect with the rest of the brain in primates [6–8] including humans [9]. Another factor that complicates the interpretation of data gathered from various species is that, despite important similarities, the anatomy of these regions differs between rodents and primates [10].

Perhaps the most influential hypothesis is that VMPFC and lateral OFC are concerned with the registration of positive and negative reinforcement respectively [11].
We propose that an alternative framework for reconciling sometimes discrepant findings is that whereas the OFC learns and encodes the specific type of reward that is associated with specific stimuli, the VMPFC plays a role in value guided decision-making about which of several options to pursue. For example, one recent study revealed that OFC lesions provoke a failure to devalue a specific outcome (such as a grape-flavoured or banana-flavoured pellet) rather than the general expectation of positive affect [12]. Moreover, representations of specific aspects of the utility of an outcome, rather than utility per se – for example its absolute magnitude, and the delay-to-receipt – are encoded in discrete populations of neurons in the OFC [13]. Indeed, the OFC’s connections with temporal, opercular, and perihinal cortex means it is ideally placed to learn specific associations between particular environmental cues and particular reinforcement outcomes. According to this alternative hypothesis the prominence of activations to negative reinforcement in lateral OFC may simply reflect the often greater importance of negative outcomes for learning in many experimental paradigms. By contrast, the VMPFC represents the integrated utility of an outcome, for example the subjective utility of a time-devalued reward [14] and expectations of reward based on diverse sources of information. Adjacent ventromedial regions have also been implicated in the encoding of the economic exchange value, or utility, of food items [15]. Neurons in central orbital area 13 have been found that represent the relative value of different types of outcomes [16]. These results are reminiscent of RL accounts proposing that all rewards might ultimately be scaled in terms of a single common neural currency [17**]. Instead of linking events with reward, the VMPFC may thus be important for adjudicating and selecting between different possible options, on the basis of their global scalar utility.

Although its activity increases as expectations of reward increase [18**,19**,20] and when positive reinforcement is delivered [11] the VMPFC’s role may not be restricted to simply maintaining reward expectations after decisions have been made but it may have a role in actually making decisions. Patients with damage to this region make suboptimal choices when distinct aspects of an outcome must be integrated [21,22**] including reward probability and magnitude during gambling tasks [23]. Although VMPFC activity correlates with reward expectation one overlooked explanation for this pattern of activity is that it reflects the actual making of a decision by a network of mutually inhibiting neurons [24]. Over the course of a decision such a network changes from initially representing the value of both potential options to finally representing the value of just the option chosen. Evidence for such interactions, although not yet available for VMPFC, has been reported elsewhere in prefrontal cortex [25] (Figure 2).

**Reward prediction errors and learning rates**

Another key question concerns the neural representation of prediction error, that is, the discrepancy between expected and observed reward. Despite new evidence for the involvement of cortical structures such as the ACC (126**,27**), see also [28**]) (Figure 1e), most human fMRI research has continued to emphasise a role for the striatum in the encoding of prediction errors [19**,29**,30**]. Recent brain imaging studies have advanced our understanding of the nature of prediction error signals in this region, reporting for example that signal size varies in inverse proportion to a subject’s financial assets, suggesting that it codes for prediction errors in marginal utility [29**]. Additionally, striatal responses may incorporate a ‘bonus’ assigned to novel

---

**Figure 1 Legend** The frontal brain regions implicated in decision making include, on the medial surface (a), the anterior cingulate cortex (ACC) and ventromedial prefrontal cortex (VMPFC), and (b) the orbitofrontal cortex (OFC), ventrolateral prefrontal cortex (VLPFC) and dorsolateral prefrontal cortex (DLPFC). Several of these regions may maintain and update representations of value that are called upon when decisions are made. Different areas may represent distinct aspects of value, for example, the ACC and OFC are needed when decisions are made about the values of actions or of stimuli respectively. Rudebeck et al. [3] taught macaque monkeys to decide between two actions on the basis of their recent reward history in a reward matching task (c) or to choose one of two visual stimuli on the basis of their recent reward history in a reward matching task (d). In each case one of the action choices (c) or one of the stimulus choices (d) had a 0.75 probability of reward while the other had a 0.25 probability of reward. The optimal proportion of choices that should be of the more rewarding option (the 0.75 probability option) is indicated by the horizontal black dashed line. When animals allocated their responses at this ratio they received reward at the maximum rate. The red dashed lines indicate when reward was received at 97% of the maximum rate. The number of trials that were taken before control (CON) animals and animals with either OFC lesions or ACC lesions began to allocate their choices optimally is shown in each figure. The number of trials taken to reach criterion was significantly greater in the action and stimulus decision making tasks after ACC lesions (c) and after OFC lesions (d) respectively. In addition to representing the value of actions ACC neurons are also active when action values are updated [26**] (e). Macaque monkeys were taught to identify which of four possible directions of response was the correct one. The correct direction remained the same for a few trials but then changed. In this example ACC neuron there is an increase in activity after feedback (indicated by arrow) informs the monkey that the action just made is now correct and reward is delivered but this occurred only on the first occasion that the action was established as the correct one (CO1, second row) and not on subsequent trials (bottom row). This neuron also indicated the direction or sign of the change in action value; it was active when the action was rewarded for the first time but not when an exploratory choice established that the action was not rewarded (INC, top row). Other neurons exhibited the opposite pattern of activity. Once the action is established as rewarding the ACC activity occurs in expectation of reward, at the time that the lever is touched (red vertical line), rather than at the time of feedback (CO2, CO3, bottom row). In other words, ACC activity contains information both about action value expectations and errors in predictions when actions were better or worse than expected. Each row of the Figure shows both raster diagrams of the action potentials recorded on individual trials at the top and the mean frequency of action potentials at the bottom. (f) The ACC does not just encode information about action values and prediction errors but it also contains information about the current learning rate. Behrens et al. [38] showed that ACC activity that was time-locked to the delivery of a decision outcome increased in proportion to the volatility of the environment, which in turn determined subjects’ learning rates.
stimuli [30–32], as well as representing ‘fictive error’—the difference between an alternative possible outcome and the actual outcome received [31–33].

However, it is salutary to note that fMRI evidence that striatal activity correlates with prediction error [19–30] is not backed up by single-neuron recordings, which instead classically report such signals in upstream dopaminergic structures, such as the ventral tegmental area (VTA), or substantia nigra (SN) both in the monkey [33] and now also the rat [34]. Indeed, given that the blood oxygen level dependent (BOLD) signal seems to correlate better with the afferent input into a brain region and not just its spiking output activity [35], fMRI prediction error signals in the striatum may actually reflect the input signals from other dopaminergic structures, such as the VTA [36–37]. Despite its small size and proximity to pulsatile blood vessels it is now clear that similar signals can be recorded from the human dopaminergic midbrain [36–37]. A second caveat is that often predictions and prediction errors are difficult to disentangle under classical RL approaches; if an unheralded cue is presented that is associated with reward then the cue itself constitutes a prediction error and indicates that the future state of the world is likely to be better than anticipated. In other words, the representation of reward prediction errors can also entail the representation of reward expectations [34] (see also Figure 1e). One implication of such considerations is that the dopaminergic system and striatum may not just encode reward prediction errors but they may also contribute to encoding reward expectations.

Moreover, while many brain imaging studies have tested for correlations between the BOLD response and the prediction error signal generated by standard RL models, they do not always test whether other experimental parameters, such as the size of reward regardless of whether or not it is predicted, could also account for any activity changes. That both VTA and ventral striatum have been shown to have measurable responses even to predicted rewards suggests that these structures may not encode prediction errors in isolation [36,37]. Quite what other information may also be present in these structures and what implications it has for behaviour is unknown.

The prediction error does not, in isolation, determine the degree to which future reward expectations will be revised. The impact that any one prediction error will have on revising a value expectation depends on the learning rate, which in turn depends on the rate of environmental change or environmental volatility [38]. When volatility is high the situation is changing quickly; although old historical reward estimates may no longer be important each new reinforcement outcome should have a large impact on future expectations. The ACC does not
just encode prediction errors but, in addition, its activity reflects volatility (Figure 1f).

**Other types of prediction error**

Recent theoretical and empirical work has raised the possibility that the dissociable representation of prediction and prediction errors is a ubiquitous property of cerebral organisation, even outside of the reward system [39,40]. According to one view, during perception, predictions generated at higher levels of the processing hierarchy are fed back and ‘subtracted’ from incoming sensory signals, such that the neural information passed forward from stage to stage consists only of prediction error [39]. Computational implementations of this ‘predictive coding’ framework propose that separate classes of prediction and prediction error units exist at multiple stages of sensory systems [41,42], mirroring the dissociation seen in the reward system. A recent brain imaging study [43] offered support for this view, describing distinct voxels that respond to prior information about an expected visual target, and the ‘mismatch’ between expected and observed perception (Figure 4). In another fMRI study, an RL model was used to estimate sensory prediction error responses during implicit audio-visual learning, and visual cortex responded robustly both to the unexpected presence and the unexpected absence of visual stimulation, precisely as predicted by the model [44] (Figure 3). The reciprocal interconnectivity between sensory prediction and prediction error units also constitutes a plausible architecture for implementing the competitive lateral interactions by which stimuli compete for attention [42]. Indeed, visual surprise may be the key determinant of where we attend during natural viewing [45], and the unique information conveyed by an event is an underlying factor driving ubiquitous electrophysiological signatures of attention such as the P300 [46] (Figure 4).

Other, more abstract quantities, such as our knowledge about risks associated with a rewarding stimulus, may also be dissociably represented as predictions and prediction errors. Activity in the human anterior insula may encode the expected variance in reward outcome (i.e. the risk associated with a reward), but also respond when these risk expectations prove erroneous and need updating [47,48]. D’Acremont and Bossaerts [49] argue that comparing expected rewards and risks is a powerful and computationally efficient decision strategy that has similarities with the methods used by financial analysts. In accord with the notion that both reward and risk expectations influence our judgements, decisions not to purchase items in a financial decision making game were associated with increased insular activity but decreased striatal and medial frontal activity [50].

Recent studies have suggested that even higher order quantities such as our expectations and confidence in other social agents may be represented in a similar way [37,51**]. Behrens et al. [37] scanned human subjects while they interacted with what they took to be another player in semi-collaborative computer game. As well as looking for activity related to predictions about whether a
decision would lead to reward they also looked for activity related to predictions about whether the collaborator would lie about the best decision to take. Reward prediction error-related activity was found in ventral striatum but social prediction error-related activity was found in a dorsomedial paracingulate cortical region. As in a previous study activity in the sulcus of the ACC was related to the volatility, and hence the learning rate, for the reward prediction error.
environment but activity in the gyrus of the ACC was related to the volatility, and hence the learning rate, for the social environment.

Conclusions
There is increasing convergence across different neuroscience methods and across species about the nature of the representations and the processes that guide value-based decision making. Nevertheless many questions remain. The next few years will reveal whether similar theoretical frameworks apply to other cognitive domains such as visual and social decision making. It is clear already that there are some important differences between reward-guided decision making and visual decision making because prediction errors in sensory systems may not be ‘signed’ (see also Figure 1e); neural activity may simply represent surprise if it is similar both when expected stimuli fail to appear and when unexpected stimuli do appear. In the case of social learning, however, medial frontal activity is signed and related to the expectation of a lie rather than to the expectation of truth telling [37]. The applicability of expectation and prediction error based accounts of decision making is likely to come under further investigation in the coming years.

Acknowledgements
The work is funded by the MRC (MFSR). RBM was supported by a Marie Curie Intra-European Fellowship within the 6th European Community Framework Programme.

References and recommended readings
Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


An animal’s appreciation of the contingency between its behaviour and a rewarding outcome can be examined by using variable ratio reward contingencies in which reward is delivered as a function of the number of responses made or the temporal interval since the last reward but such approaches have only recently been used in human imaging studies. The BOLD signal in a number of human frontal areas, including VMPC, increases with both the objective contingency between behaviour and reward and the subject’s estimate of the contingency. Disentangling contingency-related signals from reward expectation signals may, however, be difficult.


It is important to try and dissociate the encoding of reward expectations associated with particular choice options, the net value of options available in a decision, and prediction errors. Some initial data suggest that it might be possible to link BOLD activity in different frontal and striatal regions to these different parameters. See also [20] for a discussion of this approach.
BOLD signal reflects the discrepancy, sometimes referred to as the fictive error, between this alternative possible outcome and the actual outcome received.


Macaque monkeys explored four possible spatially defined choices to find the one that was associated with reward. Once the correct option was identified it could be exploited for several more trials before it changed again. ACC neuron activity increased at the time of feedback on the exploration trials with some neurons encoding positive outcomes, some encoding negative outcomes, and some active on both types of trials. On exploitation trials, however, the activity moved from the outcome period to the response period. Although the change in ACC activity on moving from exploration to exploitation trials was more categorical it was reminiscent of prediction error coding in the dopamine system.


ACC neurons encode prediction errors when macaque monkeys learned which was the correct action to make in a given context. As in some other studies [25] some neurons encoded positive feedback, some negative feedback, and some encoded both types of feedback as long as there was still uncertainty about what choice to make. While positive and negative feedback specific activity is expected if ACC encodes prediction errors, non-specific feedback-related activity is expected if parameters such as the learning rate are also encoded by ACC neurons [35].


Reward outcome related activity is also seen in the posterior cingulate region, CGp, as well as in ACC. Some CGp neurons recorded while monkeys performed a probabilistic ally rewarded task fired particularly strongly when the value of the reward outcome was not as high as the maximum possible for a particular risky choice. After such outcomes, when CGp activity was high, monkeys were more likely to switch to a less risky choice that was always rewarded at the same level on every trial. CGp microstimulation induced a similar change in behaviour.


Human subjects learned a task motivated by small monetary rewards. The size of the BOLD prediction error signal in striatum varied between subjects and it did so in inverse proportion to their financial assets suggesting that it codes for prediction errors in marginal utility.


The size of the prediction error signal in the ventral striatum also reflects a ‘bonus’ assigned to especially novel stimuli when they are the subject of learning.


When people receive feedback about the outcome that would have ensued had they taken another course of action then the dorsal striatal

Human subjects played competitive games with one another while being scanned with fMRI. Different models of behaviour were compared. In some simple RL-based models actions were expected to be chosen again if they had recently been successful but more sophisticated models incorporated information about a player’s expectations of the influence their own behaviour would have on the behaviour of the other player. Such expectations about inter-agent interactions were associated with changes in activity in dorsomedial paracingulate regions.