A framework for studying the neurobiology of value-based decision making

Antonio Rangel, Colin Camerer* and P. Read Montague‡*

Abstract | Neuroeconomics is the study of the neurobiological and computational basis of value-based decision making. Its goal is to provide a biologically based account of human behaviour that can be applied in both the natural and the social sciences. This Review proposes a framework to investigate different aspects of the neurobiology of decision making. The framework allows us to bring together recent findings in the field, highlight some of the most important outstanding problems, define a common lexicon that bridges the different disciplines that inform neuroeconomics, and point the way to future applications.

Value-based decision making is pervasive in nature. It occurs whenever an animal makes a choice from several alternatives on the basis of a subjective value that it places on them. Examples include basic animal behaviours, such as bee foraging, and complicated human decisions, such as trading in the stock market. Neuroeconomics is a relatively new discipline that studies the computations that the brain carries out in order to make value-based decisions, as well as the neural implementation of those computations. It seeks to build a biologically sound theory of how humans make decisions that can be applied in both the natural and the social sciences.

The field brings together models, tools and techniques from several disciplines. Economics provides a rich class of choice paradigms, formal models of the subjective variables that the brain needs to compute to make decisions, and some experimental protocols for how to measure these variables. Psychology provides a wealth of behavioural data that shows how animals learn and choose under different conditions, as well as theories about the nature of those processes. Neuroscience provides the knowledge of the brain and the tools to study the neural events that attend decision making. Finally, computer science provides computational models of machine learning and decision making. Ultimately, it is the computations that are central to uniting these disparate levels of description, as computational models identify the kinds of signals and signal dynamics that are required by different value-dependent learning and decision problems. However, a full understanding of choice will require a description at all these levels.

In this Review we propose a framework for thinking about decision making. It has three components: first, it divides decision-making computations into five types; second, it shows that there are multiple types of valuation systems; and third, it incorporates modulating variables that affect the different valuation processes. This framework will allow us to bring together recent findings in the field, highlight some of the most important outstanding problems, define a common lexicon that bridges the different disciplines that inform neuroeconomics, and point the way to future applications. The development of a common lexicon is important because a lot of confusion has been introduced into the literature on the neurobiology of decision making by the use of the unqualified terms 'reward' and 'value'; as shown in the Review, these terms could apply to very different computations.

Computations involved in decision making

The first part of the framework divides the computations that are required for value-based decision making into five basic processes (FIG. 1). The categorization that we propose is based on existing theoretical models of decision making in economics, psychology and computer science[1–3]. Most models in these disciplines assume, sometimes implicitly, that all of these processes are carried out every time an animal makes a value-based decision.

The first process in decision making involves the computation of a representation of the decision problem. This entails identifying internal states (for example, hunger level), external states (for example, threat level)

**Division of the Humanities and Social Sciences (HSS) and Computational and Neural Systems Program, California Institute of Technology, Pasadena, California 91125, USA. ‡Department of Neuroscience, Computational Psychiatry Unit, Baylor College of Medicine, Houston, Texas 77030, USA. Correspondence to A.R. e-mail: rangel@hss.caltech.edu*
doi:10.1038/nrn2357
Published online 11 June 2008

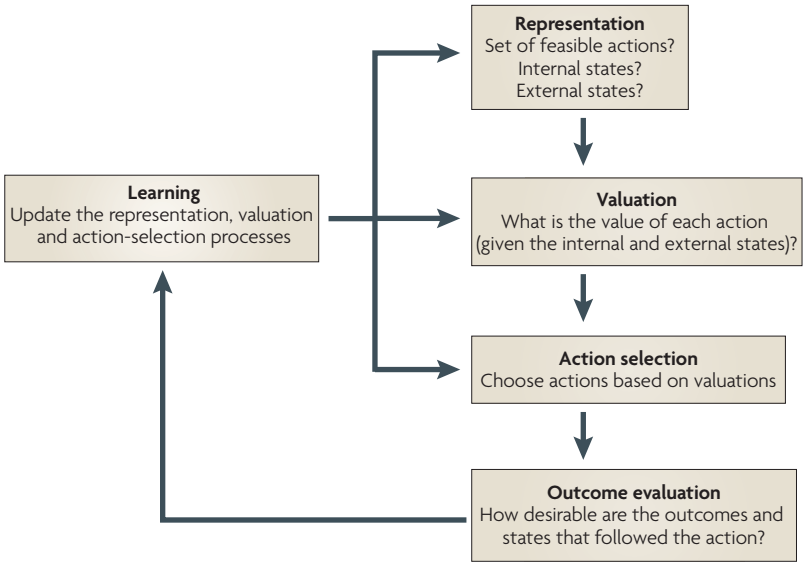


Figure 1 | **Basic computations involved in making a choice.** Value-based decision making can be broken down into five basic processes: first, the construction of a representation of the decision problem, which entails identifying internal and external states as well as potential courses of action; second, the valuation of the different actions under consideration; third, the selection of one of the actions on the basis of their valuations; fourth, after implementing the decision the brain needs to measure the desirability of the outcomes that follow; and finally, the outcome evaluation is used to update the other processes to improve the quality of future decisions.

and potential courses of action (for example, pursue prey). In the second process, the different actions that are under consideration need to be assigned a value (valuation). In order to make appropriate decisions, these values have to be reliable predictors of the benefits that are likely to result from each action. Third, the different values need to be compared for the animal to be able to make a choice (action selection). Fourth, after implementing the decision, the brain needs to measure the desirability of the outcomes. Finally, these feedback measures are used to update the other processes to improve the quality of future decisions (learning).

The five categories are not rigid, and many questions remain about how well they match the computations that are made by the brain. For example, it is not known whether valuation (step 2 in our model) must occur before action selection (step 3), or whether both computations are performed in parallel. Nevertheless, the taxonomy is conceptually useful because it breaks down the decision-making process into testable constituent processes, it organizes the neuroeconomics literature in terms of the computations that are being studied, and it makes predictions about the neurobiology of decision making, such as the hypothesis that the brain must encode distinct value signals at the decision and outcome stages, and the hypothesis that the brain computes a value signal for every course of action under consideration.

Representation

The representation process plays an essential part in decision making by identifying the potential courses of action that need to be evaluated, as well as the internal and external states that inform those valuations. For example, the valuation that a predator assigns to the action 'chasing prey' is likely to depend on its level of hunger (an internal state) as well as the conditions of the terrain (an external variable). Unfortunately, little is known about the computational and neurobiological basis of this step. Basic open questions include: how does the brain determine which actions to assign values to, and thus consider in the decision-making process, and which actions to ignore? Is there a limit to the number of actions that animals can consider at a time? How are internal and external states computed? How are the states passed to the valuation mechanisms described below?

Valuation at the time of choice

On the basis of a sizable body of animal and human behavioural evidence, several groups have proposed the existence of three different types of valuation systems: Pavlovian, habitual and goal-directed systems[4–6] (BOX. 1). These systems are sometimes in agreement but often in conflict (see section on action selection). It is important to emphasize that the precise neural basis of these three distinct valuation systems is yet to fully be established. Although the evidence described below points to neural dissociations between some of the components of the three hypothetical systems, it is possible that they do not map directly onto completely separate neural systems[6–9]. In fact, it is likely that they share common elements. Moreover, even the exact nature and number of valuation systems is still being debated. Nevertheless, conceptually the three systems provide a useful operational division of the valuation problem according to the style of the computations that are performed by each.

Pavlovian systems. Pavlovian systems assign values to a small set of behaviours that are evolutionarily appropriate responses to particular environmental stimuli. Typical examples include preparatory behaviours (such as approaching cues that predict the delivery of food) and consummatory responses to a reward (such as pecking at a food magazine). Analogously, cues that predict a punishment or the presence of an aversive stimulus can lead to avoidance behaviours. We refer to these types of behaviours as Pavlovian behaviours, and to the systems that assign value to them as the Pavlovian valuation systems.

Many Pavlovian behaviours are innate, or 'hard-wired', responses to specific predetermined stimuli. However, with sufficient training animals can also learn to deploy them in response to other stimuli. For example, rats and pigeons learn to approach lights that predict the delivery of food. An important difference between Pavlovian systems and the other two systems is that Pavlovian systems assign value to only a small set of 'prepared' behaviours and thus have a limited behavioural repertoire. Nonetheless, a wide range of human behaviours that have important economic consequences might be controlled by Pavlovian systems, such as overeating in the presence of food, behaviours displayed in people with obsessive–compulsive disorders (OCDs) and, perhaps, harvesting immediate smaller rewards at the expense of delayed larger rewards[5,9].

Box 1 | **Examples of behaviours driven by different valuation systems**

Valuation system	Valence	
	Appetitive (rewards)	**Avoidance (punishments)**
Pavlovian	Eat all food on plate	Cross street upon seeing dangerous person
	Reward obtained: food	Punishment avoided: potential harm
Habitual	Morning cup of coffee	Drive usual route to work
	Reward obtained: stimulant	Punishment avoided:traffic
Goal-directed	Movie selection	Go for a run
	Reward obtained: entertainment	Punishment avoided: obesity

Behaviour can be driven by different valuation systems. These systems can operate in the domain of rewards (that is, appetitive outcomes) and punishments (that is, aversive outcomes). Although the exact number of valuation systems and their scope remain to be determined, it is known that behaviour can be influenced by Pavlovian, habitual and goal-directed evaluators. The table contains examples of behaviours that are characteristic of each system. Consummatory actions, such as eating food that is within reach, are assigned a high value by the Pavlovian system regardless of the state of hunger. Routine actions, such as having a cup of coffee in the morning, are assigned a high value by the habitual system regardless of that morning's particular needs. Choices that are made infrequently, such as which movie to see, are assigned values by the goal-directed system.

At first glance, Pavlovian behaviours look like automatic, stimulus-triggered responses, and not like instances of value-based choice. However, as Pavlovian responses can be interrupted by other brain systems, they must be assigned something akin to a 'value' so that they can compete with the actions that are favoured by the other valuation systems.

Characterizing the computational and neural basis of Pavlovian systems has so far proven difficult. This is due in part to the fact that there might be multiple Pavlovian controllers, some of which might be responsible for triggering outcome-specific responses (for example, pecking at food or licking at water) whereas others might be responsible for triggering more general valence-dependent responses (for example, approaching for positive outcomes or withdrawing from negative ones).

The neural bases of active and passive Pavlovian responses to negative stimuli seem to have specific and spatial organizations along an axis of the dorsal periaqueductal grey[10]. With respect to valence-dependent responses, studies that used various species and methods suggested that a network that includes the basolateral amygdala, the ventral striatum and the orbitofrontal cortex (OFC) underlies the learning processes through which neutral stimuli become predictive of the value of outcomes[11,12]. In particular, the amygdala has been shown to play a crucial part in influencing some Pavlovian responses[8,13–15]. Specifically, the central nucleus of the amygdala, through its connections to the brainstem nuclei and the core of the nucleus accumbens, seems to be involved in nonspecific preparatory responses, whereas the basolateral complex of the amygdala seems to be involved in more specific responses through its connections to the hypothalamus and the periaqueductal grey.

Valence
The appetitive or aversive nature of a stimulus.

It is not currently known how many Pavlovian systems exist or how they interact with each other. Other important questions are whether there is a common carrier of Pavlovian value and, if so, how it is encoded; whether learning is possible within these systems; and how Pavlovian systems interact with the other valuation systems — for example, in phenomena such as Pavlovian-instrumental transfer[4].

Habit systems. In contrast to Pavlovian systems, which value only a small set of responses, habit systems can learn, through repeated training, to assign values to a large number of actions. Habit-valuation systems have a number of key characteristics. First, they learn to assign values to stimulus–response associations (which indicate the action that should be taken in a particular state of the world), on the basis of previous experience, through a process of trial-and-error (see the learning section below). Second, subject to some technical qualifications, habit systems learn to assign a value to actions that is commensurate with the expected reward that these actions generate, as long as sufficient practice is provided and the environment is sufficiently stable[3,6,16]. Third, because values are learned by trial-and-error, habit systems are believed to learn relatively slowly. As a consequence, they might forecast the value of actions incorrectly immediately after a change in the action–reward contingencies. Finally, these systems rely on 'generalization' when assigning action values in novel situations. For example, a rat that has learned to lever-press for liquids in response to a sound cue might respond with a similar behaviour when first exposed to a light cue. We refer to the actions that are controlled by these systems as 'habits' and the values that they compute as 'habit values'. Examples of habits include a smoker's desire to have a cigarette at particular times of day (for example, after a meal) and a rat's tendency to forage in a cue-dependent location after sufficient training.

Studies using several species and methods suggest that the dorsolateral striatum might play a crucial part in the control of habits[17,18]. As discussed below, the projections of dopamine neurons into this area are believed to be important for learning the value of actions. Furthermore, it has been suggested that stimulus–response representations might be encoded in cortico-thalamic loops[18]. Lesion studies in rats have shown that the infralimbic cortex is necessary for the establishment and deployment of habits[19,20].

There are many open questions regarding habit systems. Are there multiple habit systems? How do habitual systems value delayed rewards? What are the limits on the complexity of the environments in which the habit system can learn to compute adequate action values? How does the system incorporate risk and uncertainty? How much generalization is there from one state to another in this system (for example, from hunger to thirst)?

Goal-directed systems. In contrast to the habit system, the goal-directed system assigns values to actions by computing action–outcome associations and then evaluating the rewards that are associated with the different outcomes.

Under ideal conditions, the value that is assigned to an action equals the average reward to which it might lead. We refer to values computed by this system as 'goal values' and to the actions that it controls as 'goal-directed behaviours'. An example of a goal-directed behaviour is the decision of what to eat at a new restaurant.

Note that an important difference between habitual and goal-directed systems has to do with how they respond to changes in the environment. Consider, for example, the valuations made by a rat that has learned to press a lever to obtain food, after it is fed to satiation. The goal-directed system has learned to associate the action 'lever-press' with the outcome 'food' and thus assigns a value to the lever-press that is equal to the current value of food — which in this example is low because the animal has been fed to satiation. By contrast, the habit system assigns a high value to the lever-press because this is the value that it learned during the pre-satiation training. Thus, the goal-directed system updates the value of an action as soon as the value of its outcome changes, whereas the habit system does not.

To carry out the necessary computations, the goal-directed system needs to store action–outcome and outcome–value associations. Unfortunately, relatively little is known about the neural basis of these processes. Several rat lesion studies suggest that the dorsomedial striatum has a role in the learning and expression of action–outcome associations[21], whereas the OFC might be responsible for encoding outcome–value associations. Consistent with this, monkey electrophysiology studies have found appetitive goal-value signals in the OFC and in the dorsolateral prefrontal cortex (DLPFC)[22–25]. Electrophysiology experiments in rats point to the same conclusion[26]. In a further convergence of findings across methods and species, human functional MRI (fMRI) studies have shown that blood-oxygen-level-dependent (BOLD) activity in the medial OFC[27–31] and the DLPFC[28] correlates with behavioural measures of appetitive goal values, and individuals with damage to the medial OFC have problems making consistent appetitive choices[32]. Several lines of evidence from these various methods also point to an involvement of the basolateral amygdala and the mediodorsal thalamus (which, in combination with the DLPFC, form a network that Balleine has called the "associative cortico-basal-ganglia loop" (REF. 17)).

Several questions regarding this system remain unanswered. Are there specialized goal-directed systems for reward and punishment, and for different types of goals? How are action–outcome associations learned? How does the goal-directed system assign value to familiar and unfamiliar outcomes? How are action–outcome associations activated at the time that a choice has to be made?

For complex economic choices (such as choosing among detailed health-care plans), we speculate that, in humans, propositional logic systems have a role in constructing associations that are subsequently evaluated by the goal-directed system. For example, individuals might use a propositional system to try to forecast the consequences of a particular action, which are then evaluated by the goal-directed system. This highlights a limitation of the goal-directed system: the quality of its valuations is limited by the quality of the action–outcome associations that it uses.

Propositional logic system
A cognitive system that makes predictions about the world on the basis of known pieces of information.

Outstanding issues. Some general, important questions regarding the different valuation systems remain unanswered. First, are there multiple Pavlovian, habitual and goal-directed valuation systems, with each system specializing in particular classes of actions (in the case of the Pavlovian and habit systems) or outcomes (in the case of the goal-directed system)? For example, consider a dieter who is offered a tasty dessert at a party. If this is a novel situation, it is likely to be evaluated by the goal-directed system. The dieter is likely to experience conflict between going for the taste of the dessert and sticking to his health goals. This might entail a conflict between two goal-directed systems, one that is focused on the evaluation of immediate taste rewards and one that is focused on the evaluation of long-term outcomes. Second, are there more than three valuation systems? Lengyel and Dayan[5,33] have proposed the existence of an additional, 'episodic' system. At this point it is unclear how such a system differs both conceptually and neurally from the goal-directed system. Third, how does the brain implement the valuation computations of the different systems? Finally, how do long-term goals, cultural norms and moral considerations get incorporated into the valuation process? One possibility is that the habit and goal-directed systems treat violations of these goals and cultural and moral rules as aversive outcomes, and that compliance with them is treated as a rewarding outcome[34]. However, this can be the case only if the brain has developed the capacity to incorporate social and moral considerations into its standard valuation circuitry. Another possibility is that there are separate valuation systems for these types of considerations that are yet to be discovered.

Modulators of the valuation systems

Several factors can affect the values that the Pavlovian, habitual and goal-directed systems assign to actions. For example, the value that is assigned to an action might depend on the riskiness of its associated payoffs, the delay which with those payoffs occur and the social context. We refer to these types of variables as value modulators. Importantly, modulators might have different effects in each of the valuation systems. In this section we focus on the impact of risk and delay on the goal-directed valuation system, as most of the existing evidence pertains to this system. For reviews on social modulators, see REFS 35,36.

Risk and uncertainty. All decisions involve some degree of risk, in the sense that action–outcome associations are probabilistic (BOX 2). We refer to an action that has uncertain rewards as a 'prospect'. In order to make good decisions, the goal-directed system needs to take into account the likelihood of the different outcomes. Two hotly debated questions are: first, what are the computations that the goal-directed system uses to incorporate

Statistical moments
Properties of a distribution, such as mean and variance.

risks into its valuations; and second, how does the brain implement such computations[37]?

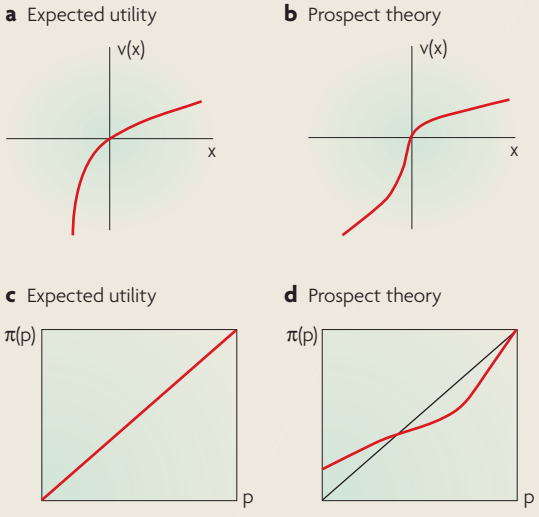
Early human neuroimaging studies in this topic identified some of the areas that are involved in making risky decisions, but were not able to characterize the nature of the computations made by these systems[38–41]. Currently, two main competing views regarding the nature of such computations are being tested. The first view, which is widely used in financial economics and behavioural ecology, asserts that the brain assigns value to prospects by first computing its statistical moments (such as its expected magnitude, its variance or coefficient of variation, and its skewness) and then aggregating them into a value signal[42,43]. The second view, which is widely used in other areas of economics and in psychology, asserts that the value is computed using either expected-utility theory (EU) or prospect theory (PT) (BOX 2). In this case the brain needs to compute a utility value for each potential outcome, which is then weighted by a function of the probabilities.

Decisions that result from an EU or PT valuation function can be approximated by a weighted sum of the prospects' statistical moments (and vice versa). This makes it difficult to distinguish the two views on the basis of behavioural data alone. Neuroimaging studies can provide important insights, although the debate between the two views has not yet been settled. In agreement with the first view, a number of recent human fMRI studies have found activity that is consistent with the presence of expected value signals in the striatum[44,45] and the medial OFC[46], and activity that is consistent with risk signals (as measured by the mathematical variance of the prospects) in the striatum[44,47], the insula[46,48] and the lateral OFC[45]. Similar risk and expected-value signals have been found in the midbrain dopamine system in electrophysiology studies in non-human primates[49]. Expected-value signals (BOX 2) have also been found in the lateral intraparietal cortex in non-human primate electrophysiology experiments[50]. Consistent with the second view, a recent human fMRI study found evidence for a PT-like value signal in a network that includes the ventral and dorsal striatum, the ventromedial and ventrolateral prefrontal cortex, the anterior cingulate cortex (ACC) and some midbrain dopaminergic regions[27]. The existence of evidence that is consistent with both views presents an apparent puzzle. A potential resolution that should be explored in future studies is that the striatal-prefrontal network might integrate the statistical moments that are encoded elsewhere into a value signal that exhibits EU- or PT-like properties.

In many circumstances, decision makers have incomplete knowledge of the risk parameters — a situation known as ambiguity that is different from the pure risk case in which the probabilities are known. Human behavioural studies have shown that people generally have an aversion to choices that are ambiguous[51], which suggests that a parameter that measures the amount of ambiguity might be encoded in the brain and might be used to modulate the value signal. Some preliminary human fMRI evidence points to the amygdala, the OFC[52] and the anterior insula[53] as areas where such a parameter might be encoded.

Some issues regarding risk and valuation are still unclear. First, little is known about how risk affects the computation of value in Pavlovian and habitual systems. For example, most reinforcement learning models (see below) assume that the habit learning system encodes a value signal that incorporates expected values but not risks. This assumption, however, has not been thoroughly tested. Second, little is known about how the brain learns the risk parameters. For example, some behavioural

Box 2 | **Risk modulators of value in the goal-directed system**



Many decisions involve the valuation of rewards and costs that occur probabilistically, often called 'prospects'. There are two dominant theories in economics about how valuation systems incorporate probability in the assignment of value. In expected-utility theory (EU), the value of a prospect equals the sum of the value of the individual outcomes, $v(x)$, weighted by their objective probability, $p(x)$, which is given by $\sum_x p(x)v(x)$. Under some special assumptions on the function $v(.)$, which are popular in the study of financial markets, the EU formula boils down to a weighted sum of the expected value and the variance of the prospect[42]. The appeal of EU comes from the fact that it is consistent with plausible normative axioms for decision making, from its mathematical tractability and from its success in explaining some aspects of market behaviour. An alternative approach, called prospect theory (PT), states that the value of a prospect equals $\sum_x \pi(p(x))v(x - r)$, where the values of the outcomes now depend on a reference point, r, and are weighted by a nonlinear function, $\pi(.)$, of the objective probabilities[122,123]. Reference-dependence can create framing effects (analogous to figure–ground switches in vision), in which different values are assigned to the same prospect depending on which reference point is cognitively prominent. The figure illustrates the usual assumptions that are imposed in the value and probability functions by the two theories. As shown in parts **a** and **c**, in EU the value function, $v(.)$, is a concave function of outcomes, and the probability function is the identity function. Note that a special case that is often used in the experimental neuroeconomics literature is $v(x) = x$, which makes the EU function reduce to the expected value of the prospect. The properties of PT are illustrated in parts **b** and **d**. The value function is usually revealed by choices to be concave for gains but convex for losses. This assumption is justified by the psychologically plausible assumption of diminished marginal sensitivity to both gains and losses starting from the reference point. PT also assumes that $v(x) < -v(-x)$ for $x > 0$, a property called 'loss-aversion', which leads to a kink in the value function. Part **d** illustrates the version of PT in which small probabilities are overweighted and large probabilities are underweighted. PT has been successful in explaining some behaviour that was inconsistent with EU theory in behavioural experiments with humans[123] and monkeys[124], as well as economic field evidence[125].

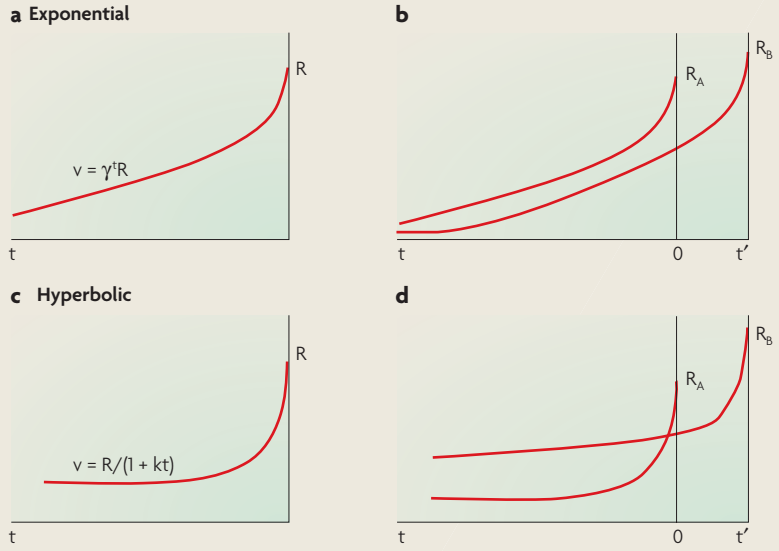
Neuroeconomists make a distinction between prospects that involve risk and those that involve ambiguity. Risk refers to a situation in which all of the probabilities are known. Ambiguity refers to a situation in which some of the probabilities are unknown. The EU and PT models described above apply to valuation under risk, but not under ambiguity. Several models of valuation under ambiguity have been proposed, but none of them has received strong empirical support[51,126,127].

Expected-utility theory
A theory that states that the value of a prospect (or of random rewards) equals the sum of the value of the potential outcomes weighted by their probability.

Prospect theory
An alternative to the expected utility theory that also pertains to how to evaluate prospects.

evidence suggests that habit and goal-directed systems learn about probabilities in different ways and that this leads to different probability weighting by the two systems[54]. Finally, more work is required to better characterize the nature of the computations that are made by the amygdala and the insula in decision making under uncertainty. Preliminary insights suggest that the amygdala might have an asymmetric role in the evaluation of gains and losses. For example, humans with amygdala damage made poor decisions if the decisions involved potential gains, but not if they involved losses[55], and a related study showed that the amygdala is differentially activated when subjects decide to take risks for large gains and when they decide to accept a sure loss[56].

Box 3 | **Temporal modulators of value in the goal-directed system**



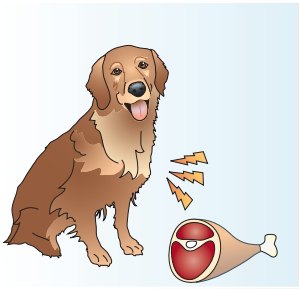
Many decisions involve the evaluation of rewards and costs that arrive with different delays. Thus, the valuation systems require a mechanism for incorporating the timing of rewards into their computations. Two prominent models of discounting have been proposed in psychology and economics. In the first model, known as hyperbolic discounting, rewards and costs that arrive t units of time in the future are discounted by a factor 1/(1+kt). Note that the discount factor is a hyperbolic function of time and that a smaller *k* is associated with less discounting (that is, more patience). In the second model, known as exponential discounting, the corresponding discount factor is γ^t. Note that a value of γ closer to one is associated with more patience. An important distinction between the two models is illustrated in parts **a** and **c** of the figure, which depict the value of a reward of size R *t* units of time before it arrives. Note that whereas every additional delay is discounted at the same rate (γ) in the exponential case, in hyperbolic discounting initial delays are discounted at a much higher rate and the discount curve flattens out for additional delays.

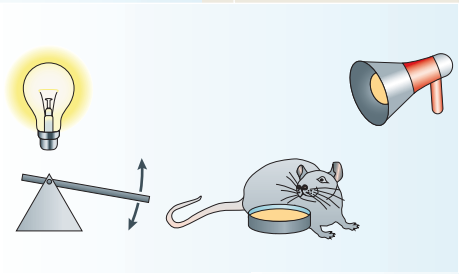
In most comparative behavioural studies of goal-directed behaviour with adequate statistical power, hyperbolic discount functions always fit the observed behaviour better than exponential functions[57]. Nevertheless, economists and computer scientists find the exponential function appealing because it is the only discount function that satisfies the normative principle of dynamic consistency, which greatly simplifies modelling. This property requires that if a reward, A, is assigned a higher value than another reward, B, at time *t*, then the same reward is also assigned a higher value when evaluated at any time *t–k*. Under hyperbolic discounting, by contrast, the relative valuation between the two actions depends on when the choice is made. This is known as dynamic inconsistency. Parts **b** and **d** of the figure illustrate this difference. They depict the comparative value of a reward, R_A, received at time *0* with a reward, R_B, received at time *t′* as a function of the time when the rewards are being evaluated. Note that in the exponential case the relative desirability of the two rewards is constant, whereas for the hyperbolic case it depends on the time of evaluation.

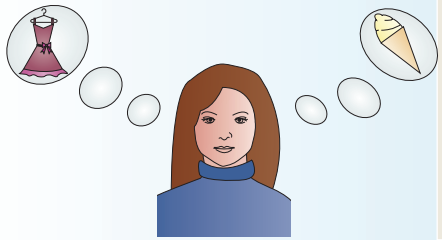
Time discounting. In all real-world situations there is a time lag between decisions and outcomes. From a range of behavioural experiments it is well-established that the goal-directed and habitual systems assign lower values to delayed rewards than to immediate ones; this phenomenon is known as time discounting[57]. The role of time discounting in the Pavlovian system is not as well-understood. As before, we focus on the impact of temporal discounting on the goal-directed system, as this is where most of the studies so far have focused.

The current understanding of time discounting parallels that for risk: two competing views have been proposed and are being tested using a combination of human-behavioural and neuroimaging experiments. One camp interprets the human fMRI evidence using the perspective of dual-process psychological models and has argued that discounting results from the interaction of at least two different neural valuation systems (BOX 3), one with a low discount rate and one with a high discount rate[58–60]. In this view, the patience that is exhibited by any given individual when making decisions depends on the relative activation of these two systems. In sharp contrast, the other camp has presented human fMRI evidence that suggests that there is a single valuation system that discounts future rewards either exponentially or hyperbolically[61] (BOX 3). As with the situation for risk valuation, this presents an apparent puzzle. A potential reconciliation is that the striatal-prefrontal network might integrate information that is encoded elsewhere in the brain into a single value signal, but that immediate and delayed outcomes might activate different types of information that are used to compute the value. For example, immediate rewards might activate 'immediacy markers' that increase the valuation signals in the striatal-prefrontal network. An understanding of these issues is also important from the perspective of brain development. When do value signals get computed in their 'adult' form and how do they contribute to choices made by children and adolescents? These and other related questions show that the economic framing of decision making will continue to provide new ways to probe the development and function of choice mechanisms in humans.

Time discounting remains a fruitful topic of investigation. First, the discounting properties of Pavlovian and habitual systems in humans have not been systematically explored. Second, the inputs to the valuation network are unknown, as is the reason why the aggregation of those inputs produces a hyperbolic-like signal in valuation areas such as the ventral striatum and the medial OFC. Third, the behavioural evidence suggests that discount factors are highly dependent on contextual variables. For example, subjects' willingness to delay gratification depends on whether the choice is phrased as a "delay" or as a "choice between two points in time" (REF. 62), on how they are instructed to think about the rewards[63] and on the subjects' arousal level[64]. The mechanisms through which such variables affect the valuation process







	Pavlovian	Habitual	Goal-directed
Pavlovian	Example: hungry animal presented with food and electric shock simultaneously Appetitive Pavlovian system: high value for food, low value for escape behaviours Avoidance Pavlovian system: high value for escape behaviours, low value for food	Example: animal rewarded for running away from food Appetitive Pavlovian system: high value for running towards food Avoidance habitual system: high value for running away from food	Example: individual considering taking an extra bite after feeling full Appetitive Pavlovian system: high value for food Health goal-directed system: low value for food
Habitual	Example: an animal trained to run towards a lever in response to a sound and away from a lever in response to a light being presented with both stimuli Approach habitual system: high value for lever approach Avoidance habitual system: high value for lever avoidance	Example: alcoholic considering having a drink at a bar Appetitive habitual system: high value for drink Avoidance goal-directed system: low value for drink	
Goal-directed	Example: dieter considering having ice-cream Appetitive goal-directed system: high value for ice-cream Avoidance goal-directed system: low value for ice-cream		

Figure 2 | **Conflict between the valuation systems.** The different valuation systems are often in agreement. For example, when an individual is hungry at meal time, the Pavlovian, habitual and goal-directed systems assign high value to the consumption of food. However, conflicts between the systems are also common and might lead to poor decision making. This figure provides examples of conflict among the different valuation systems and of conflict among different value signals of the same type.

are unknown. Fourth, several studies have shown that the anticipation of future rewards and punishments can affect subjects' behavioural discount rates[65,66]. The mechanisms through which anticipation affects valuation are also unknown. Finally, animals make very myopic choices that are consistent with large hyperbolic discount rates[67–70]. How do humans and animals differ in the way in which they incorporate temporal delays into the valuation process?

Action selection

Even for choices that involve only one of the valuation systems discussed above, options with different values need to be compared in order to make a decision. Little is known about how the brain does this. The only available theoretical models come from the literature on perceptual decision making, which has modelled binary perceptual choices as a race-to-barrier diffusion process[71–76]. However, it is unclear whether this class of model also applies to value-based decision making and, if so, how the models might be extended to cases of multi-action choice.

Another issue is the competition that arises among the different valuation systems when an animal has to make a choice between several potential actions that are assigned conflicting values (FIG. 2). Some preliminary theoretical proposals have been made, but the experimental evidence is scarce. Daw *et al.*[77] have suggested that the brain arbitrates between the habit and goal-directed valuation systems by assigning control to the system that at any given time has the less uncertain estimate of the true value of the actions. As the quality of the estimates that are made by the habit system increases with experience, this means in practice that the habit system should gradually take over from the goal-directed system[34]. Frank has proposed a neural-network model for choice between appetitive and aversive habitual valuations[78,79].

Understanding how the 'control assignment' problem is resolved is important for several reasons. First, as illustrated in FIG. 2 and as emphasized by Dayan *et al.*[9], many apparently puzzling behaviours are likely to arise as a result of the conflict between the different valuation systems. Second, in most circumstances the quality of decision making depends on the brain's ability to assign control to the valuation system that makes the best value forecasts. For example, it is probably optimal to assign control to the habit system in familiar circumstances, but not in rapidly changing environments. Third, some decision-making pathologies (for example, OCD and overeating) might be due to an inability to assign control to the appropriate system.

Dual-process models
A class of psychological models in which two processes with different properties compete to determine the outcome of a computation.

Race-to-barrier diffusion process
A stochastic process that terminates when the variable of interest reaches a certain threshold value.

There are many important open questions in the domain of action selection. First, in the case of goal-directed decisions, does the brain make decisions by comparing the value of outcomes, of the actions that are necessary to achieve those outcomes, or both? Second, what is the neural basis of the action-selection processes in the Pavlovian, habitual and goal-directed systems? Third, what are the neural mechanisms that are used to arbitrate between the different controllers, and is there a hierarchy of controllers so that some (for example, Pavlovian systems) tend to take precedence over others (for example, goal-directed systems)? Fourth, are there any neural markers that can be reliably used to identify goal-directed or habitual behavioural control?

Outcome evaluation

In order to learn how to make good decisions the brain needs to compute a separate value signal that measures the desirability of the outcomes that were generated by its previous decisions. For example, it is useful for an animal to know whether the last food that it consumed led to illness so that it can know whether it ought to avoid that food in the future.

The computations that are made by the outcome-evaluation system, as well as the neural basis of these computations, are slowly beginning to be understood. The existing evidence comes from several different methods and species. Human fMRI studies have shown that activity in the medial OFC at the time that a reward is being enjoyed correlates with subjective reports about the quality of the experience — this has been shown for olfactory[80–83], gustatory[84–86] and even musical rewards[87]. Moreover, the activity in the medial OFC parallels the reduction in outcome value that one would expect after a subject is fed to satiation[88,89]. This suggests that the medial OFC might be an area where positive outcome valuations are computed. Interestingly, other human fMRI studies have found positive responses in the medial OFC to the receipt of secondary reinforcers, such as monetary payoffs[90–92]. Analogous results have been found for negative experiences: in humans, subjective reports of pain intensity correlated with activity in the insula and the ACC[93,94].

Animal studies have also provided insight into the neural basis of the outcome-value signal. A recent electrophysiology experiment in monkeys found outcome-value signals in the dorsal ACC[95]. In addition, a series of provocative rat studies showed that it is possible to increase outward manifestations of 'liking' in rats (for example, tongue protrusions) by activating the nucleus accumbens and subsets of the ventral pallidum using opioid agonists[85,96–98]. Interestingly, and consistent with the hypothesis that outcome-evaluation signals play a part in learning, rats that received opioid agonists subsequently consumed more of the reward that was paired with the agonist.

Some recent human fMRI experiments have also provided novel insights into the computational properties of the outcome-value signal. For example, one study showed that activity in the medial OFC in response to an odour depended on whether subjects believed that they were smelling cheddar cheese or a sweaty sock[83]. In another study[99], activity in the medial OFC in response to the consumption of wine depended on beliefs about its price, and a third study[84] showed that the outcome-valuation signal after consumption of soda depended on beliefs about its brand. Together, these findings suggest that the outcome-valuation system is modulated by higher cognitive processes that determine expectancies and beliefs.

Much remains to be understood about the outcome-valuation system. What network is responsible for computing positive and negative outcome values in different types of domains? How are positive and negative outcome-valuation signals integrated? How are these signals passed to the learning processes described in the next section? Can they be modulated by variables such as long-term goals, social norms and moral considerations?

Learning

Although some Pavlovian behaviours are innate responses to environmental stimuli, most forms of behaviour involve some form of learning. In fact, in order to make good choices animals need to learn how to deploy the appropriate computations during the different stages of decision making. First, the brain must learn to activate representations of the most advantageous behaviours in every state. This is a non-trivial learning problem given that animals and humans have limited computational power, and yet they can deploy a large number of behavioural responses. Second, the valuation systems must learn to assign to actions values that match their anticipated rewards. Finally, the action-selection processes need to learn how to best allocate control among the different valuation systems.

Of all of these processes, the one that is best-understood is the learning of action values by the habit system. In this area there has been a productive interplay between theoretical models from computer science (BOX 4) and experiments using electrophysiology in rats and monkeys and fMRI in humans. In particular, various reinforcement learning models have been proposed to describe the computations that are made by the habit system[3,100]. The basic idea behind these models is that a prediction-error signal is computed after observing the outcome generated by every choice. The signal is called a prediction error because it measures the quality of the forecast that was implicit in the previous valuation (BOX 4). Every time a learning event occurs, the value of the actions is changed by an amount that is proportional to the prediction error. Over time, and under the appropriate technical conditions, the animal learns to assign the correct value to actions.

The existence of prediction-error-like signals in the brain is one of the best-documented facts in neuroeconomics. Schultz and colleagues initially observed such signals in electrophysiology studies performed in midbrain dopamine neurons of monkeys[101–106]. The connection between these signals and reinforcement-learning models was made in a series of papers by Montague and colleagues that were published in the 1990s[103,107]. Since then, several fMRI studies have

Box 4 | **Reinforcement learning models action-value learning in the habitual system**

Several models from computer science have proved to be useful in modelling how the habitual system learns to assign values to actions. All of these models have the following structure, which is known as a Markovian decision problem: first, the animal can be in a finite set of states and can take a finite set of actions; second, there is a transition function, *T(s,a,s')*, that specifies the probability that state *s* and action *a* at one time-step will result in the state *s'* at the next time-step; and third, at every time-step the animal obtains an action and a state-contingent reward, *r(a,s)*. A behavioural rule in this world (called a policy and denoted by $\pi(s)$) specifies the action that the animal takes in every state. In this world the habitual system needs to solve two problems. First, given a policy, it needs to compute the value of taking every action *a* in every state *s*. This is given by

$$Q^{\pi}(s, a) = E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \gamma^3 r_{t+3} + \ldots \mid s_t = s, a_t = a, a_{t+1} = \pi(s_{t+1}),\ldots]; \quad (1)$$

where r_{t+k} denotes the reward that is received at time $t + k$ and where $\gamma > 0$ is the discount rate. Second, it needs to identify the policy that generates the largest sum of exponentially discounted rewards (see BOX 3) in every state.

How could the habitual system learn $Q^{\pi}(s,a)$? Let $Q^{\pi}(s, a)$ denote the estimate that the system has at any point in time. Equation 1 can be rewritten in recursive form as

$$Q^{\pi}(s, a) = R(s) + \gamma \sum_{s' \in S} T(s, a, s') Q^{\pi}(s', \pi(s')) \quad (2)$$

Consider an estimator, $\hat{Q}(s, a)$, of $Q^{\pi}(s, a)$. Note that if $\hat{Q}(s, a)$ does not satisfy this expression, then it is not a good estimate of the value function. Define a prediction error

$$\delta_t = r_t + \gamma \, max_{a'} [\hat{Q}(s_{t+1}, a')] - \hat{Q}(s_t, a_t) \quad (3)$$

that is a sample measure of how close the estimate is to satisfying equation 2. If $\delta_t > 0$ the value of the action is overestimated; if $\delta_t < 0$ the value is underestimated. One can then use the prediction error to update the estimates of the action values as follows:

$$\hat{Q}(s_t, a_t) \leftarrow Q(\hat{s_t}, a_t) + \eta\delta_t \quad (4)$$

where η is a number between 0 and 1 that determines the speed of learning. This model is known as Q-learning and it satisfies one important property: subject to some technical conditions, the estimated action values converge to those that are generated by the optimal policy. It then follows that the animal can learn the optimal policy simply by following this algorithm and, at every step of the learning process, selecting the actions with the largest values. Two other variants of this model have been proposed as descriptions of how the habitual system learns. They are known as SARSA and the actor–critic model. They differ from Q-learning on the exact specification of the prediction error and the update rule, but they are based on essentially the same idea. Note that neither SARSA nor the actor–critic model is guaranteed to converge to the optimal policy.

It is worth emphasizing several properties of these learning models. First, they are model-free in the sense that the animal is not assumed to know anything about the transition function or the reward function. Second, they explain a wide range of conditioning behaviours that are associated with the habitual system, such as blocking, overshadowing and inhibitory conditioning. Finally, they are computationally simple in the sense that they do not require the animal to keep track of long sequences of rewards to learn the value of actions.

The reinforcement-learning models described here are often used to describe the process of action-value learning in the habitual system. The algorithms that the Pavlovian and goal-directed systems use to update their values based on feedback from the environment are currently unknown.

shown that, in humans, the BOLD signal in the ventral striatum (an important target of midbrain dopamine neurons) correlates with prediction errors in a wide range of tasks[29,90,108–113].

Although the existing evidence suggests that there is a remarkable match between the computational models and the activity of the dopamine system, recent experiments have demonstrated that much remains to be understood. First, a monkey electrophysiology study[114] suggested that the phasic firing rates of midbrain dopamine neurons might only encode the positive component of the prediction error (henceforth the 'positive prediction error'). This raises the question of which brain areas and neurotransmitter systems encode the negative component (henceforth the 'negative prediction error'), which is also essential for learning. Several possibilities have been proposed. A secondary analysis[115] of the monkey electrophysiology experiment[114] suggested that the magnitude of the negative prediction errors might be encoded in the timing of the fire-and-pause patterns of the dopamine cells[115]. Some human fMRI studies have found a BOLD signal in the amygdala that resembles a negative prediction error[108], but others have failed to replicate this finding and have instead found evidence for both types of prediction error in different parts of the striatum[116]. In turn, Daw and Dayan[117] proposed that the two prediction-error signals are encoded by the phasic responses of two neurotransmitter systems: dopamine for positive prediction errors and serotonin for negative prediction errors. Second, it was shown that midbrain dopamine neurons adjust their firing rates to changes in the magnitude of reward in a way that is inconsistent with the standard interpretation of prediction errors[49]. The exact nature of these adjustments remains an open

Box 5 | **From neuroeconomics to computational psychiatry**

Sometimes the brain's decision-making processes function so differently from societal norms that we label the ensuing behaviours and perceptions a psychiatric disease. The medical community recognizes and categorizes them according to well-accepted diagnostic criteria that, so far, have relied mostly on collections of behavioural features. Neuroscientists have accumulated a substantial amount of neurobiological data that impinges directly on these illnesses[128]. For example, there are now animal models for nicotine addiction, anxiety, depression and schizophrenia that have produced a veritable flood of data on neurotransmitter systems, receptors and gene expression[129,130]. Thus, there is a substantial body of biological data and detailed descriptions of the behavioural outcomes, but little is known about what connects them. This situation presents an opportunity for neuroeconomics and other computationally oriented sciences to connect the growing body of biological knowledge to the behavioural end points.

Computational models of reinforcement learning provide a new language for understanding mental illness and a starting point for connecting detailed neural substrates to behavioural outcomes. For example, reinforcement-learning models predict the existence of valuation malfunctions, in which a drug, a disease or a developmental event perturbs the brain's capacity to assign appropriate value to behavioural acts or mental states[34,131–133].

Disorders of decision making can also arise at the action-selection stage, especially when there are conflicts among the valuation systems. This presents the possibility of generating a new quantifiable taxonomy of mental-disease states. Interestingly, this set of issues is closely related to the problem of how to think about the 'will' and has applications to addiction, obsessive–compulsive disorder and obesity. These issues relate directly to the idea of executive control and the way that it is affected by mental disease. It is our opinion that future progress in this area will require more computational approaches, because only through such models can competing ideas of executive control be clearly differentiated. Such efforts are already well underway, and various modelling efforts have been applied to executive control and decision making in humans[79,134,135].

Another neuroeconomics concept that is ripe for applications to psychiatry is motivation, which is a measure of how hard an animal works in order to retrieve a reward. Disorders of motivation might play an especially important part in mood disorders, such as depression, and in Parkinson's disease[78,136].

question[43]. Finally, a study showed that the habit system can also learn from observing the outcomes of actions that it did not take, as opposed to only being able to learn from direct experience[118]. This form of 'fictive learning' is not captured by traditional reinforcement-learning models but is common in human strategic learning and suggests that the theory needs to be extended in new directions (to include, among others, imitative learning from observing the actions of others)[119].

Other important questions in the domain of value learning include the following: how does the goal-directed system learn the action-outcome and outcome-value representations that it needs to compute action values? What are the limitations of the habit system in situations in which there is a complex credit-assignment problem (because actions and outcomes are not perfectly alternated) and delayed rewards? How does the habit system learn to incorporate internal and external states in its valuations and generalize across them? How do the different learning systems incorporate expected uncertainty about the feedback signals[43]? To what extent can the different value systems learn by observation as opposed to through direct experience[120]?

The next 5 years and beyond

Although neuroeconomics is a new field and many central questions remain to be answered, rapid progress is being made. As illustrated by the framework provided in this Review, the field now has a coherent lexicon and research aims. The key challenge for neuroeconomics over the next few years is to provide a systematic characterization of the computational and neurobiological basis of the representation, valuation, action-comparision, outcome-valuation and value-learning processes described above. This will prove to be challenging because, as we have seen, at least three valuation systems seem to be at work, fighting over the control of the decision-making process.

Nevertheless, several welcome developments suggest that the next 5 years will produce significant progress in answering many of the questions outlined here. First, there is the close connection between theory and experiments, and the widespread use of theory-driven experimentation (including behavioural parameters inferred from choices that can be linked across subjects or trials to brain activity). Second, there is the rapid adoption of new technologies, such as fast cyclic voltammetry in freely moving animals[121], which permits quasi-real-time monitoring of neurotransmitter levels for long periods. Third, there is the investigation of decision-making phenomena using different species and experimental methods, which permits more rapid progress.

This is good news, because the range of potential applications is significant. The most important area in which knowledge from neuroeconomics can be applied is psychiatry. Many psychiatric disorders involve a failure of one or more of the decision-making processes described here (BOX 5). A better understanding of these processes should lead to improved diagnosis and treatment. Another area of application is the judicial system. A central question in many legal procedures is how to define and measure whether individuals are in full command of their decision-making faculties. Neuroeconomics has the potential to provide better answers to this question. Similarly, an improved understanding of why people experience failures of self-control should lead to better public-policy interventions in areas ranging from addiction and obesity to savings. The field also has the potential to improve our understanding of how marketing affects decisions and when it should be

Credit-assignment problem
The problem of crediting rewards to particular actions in complex environments.

regulated. Artificial intelligence is another fertile area of application: a question of particular interest is which features of the brain's decision-making mechanisms are optimal and should be imitated by artificial systems, and which mechanisms can be improved upon. Finally, neuroeconomics might advance our understanding of how to train individuals to become better decision-makers, especially in conditions of extreme time-pressure and large stakes, such as those that arise in policing, in war and in fast-paced financial markets.

1. Busemeyer, J. R. & Johnson, J. G. in *Handbook of Judgment and Decision Making* (eds Koehler, D. & Narvey, N.) 133–154 (Blackwell Publishing Co., New York, 2004).
2. Mas-Colell, A., Whinston, M. & Green, J. *Microeconomic Theory* (Cambridge Univ. Press, Cambridge, 1995).
3. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, Massachusetts, 1998).
4. Dickison, A. & Balleine, B. W. in *Steven's Handbook of Experimental Psychology Vol. 3 Learning, Motivation & Emotion* (ed. Gallistel, C.) 497–533 (Wiley & Sons, New York, 2002).
5. Dayan, P. in *Better Than Conscious? Implications for Performance and Institutional Analysis* (eds Engel, C. & Singer, W.) 51–70 (MIT Press, Cambridge, Massachusetts, 2008).
6. Balleine, B. W., Daw, N. & O'Doherty, J. in *Neuroeconomics: Decision-Making and the Brain* (eds Glimcher, P. W., Fehr, E., Camerer, C. & Poldrack, R. A.) 365–385 (Elsevier, New York, 2008).
7. Bouton, M. E. *Learning and Behavior: A Contemporary Synthesis* (Sinauer Associates, Inc., Sunderland, Massachusetts, 2007).
This book reviews a large amount of evidence pointing to multiple valuation systems being active in value-based decision making.
8. Dayan, P. & Seymour, B. in *Neuroeconomics: Decision Making and the Brain* (eds Glimcher, P. W., Camerer, C. F., Fehr, E. & Poldrack, R. A.) 175–191 (Elsevier, New York, 2008).
9. Dayan, P., Niv, Y., Seymour, B. & Daw, N. D. The misbehavior of value and the discipline of the will. *Neural Netw.* **19**, 1153–1160 (2006).
This paper provided several models of how "pathological behaviours" can arise from the competition process between Pavlovian, habitual and goal-directed valuation systems.
10. Keay, K. A. & Bandler, R. Parallel circuits mediating distinct emotional coping reactions to different types of stress. *Neurosci. Biobehav. Rev.* **25**, 669–678 (2001).
11. Cardinal, R. N., Parkinson, J. A., Hall, J. & Everitt, B. J. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* **26**, 321–352 (2002).
12. Holland, P. C. & Gallagher, M. Amygdala-frontal interactions and reward expectancy. *Curr. Opin. Neurobiol.* **14**, 148–155 (2004).
13. Fendt, M. & Fanselow, M. S. The neuroanatomical and neurochemical basis of conditioned fear. *Neurosci. Biobehav. Rev.* **23**, 743–760 (1999).
14. Adams, D. B. Brain mechanisms of aggressive behavior: an updated review. *Neurosci. Biobehav. Rev.* **30**, 304–318 (2006).
15. Niv, Y. in *Neuroscience* (Hebrew University, Jerusalem, 2007).
16. Dayan, P. & Abbott, L. R. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, Massachusetts, 1999).
17. Balleine, B. W. Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. *Physiol. Behav.* **86**, 717–730 (2005).
This important paper reviews a large amount of evidence pointing to multiple valuation systems being active in value-based decision making.
18. Yin, H. H. & Knowlton, B. J. The role of the basal ganglia in habit formation. *Nature Rev. Neurosci.* **7**, 464–476 (2006).
19. Killcross, S. & Coutureau, E. Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb. Cortex* **13**, 400–408 (2003).
20. Coutureau, E. & Killcross, S. Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behav. Brain Res.* **146**, 167–174 (2003).
21. Yin, H. H., Knowlton, B. J. & Balleine, B. W. Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur. J. Neurosci.* **22**, 505–512 (2005).
22. Wallis, J. D. & Miller, E. K. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* **18**, 2069–2081 (2003).
23. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).
This paper showed that neurons in the monkey OFC encode the goal value of individual rewarding objects (for example, different liquids) irrespective of the action that needs to be taken to obtain them.
24. Wallis, J. D. Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.* **30**, 31–56 (2007).
25. Barraclough, D. J., Conroy, M. L. & Lee, D. Prefrontal cortex and decision making in a mixed-strategy game. *Nature Neurosci.* **7**, 404–410 (2004).
26. Schoenbaum, G. & Roesch, M. Orbitofrontal cortex, associative learning, and expectancies. *Neuron* **47**, 633–636 (2005).
27. Tom, S. M., Fox, C. R., Trepel, C. & Poldrack, R. A. The neural basis of loss aversion in decision-making under risk. *Science* **315**, 515–518 (2007).
This fMRI study showed that the striatal-OFC network encodes a value signal at the time of the goal-directed choice that is consistent with the properties of PT. Furthermore, the study presented evidence that suggests that both the appetitive and the aversive aspects of goal-directed decisions might be encoded in a common valuation network.
28. Plassmann, H., O'Doherty, J. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27**, 9984–9988 (2007).
29. Hare, T., O'Doherty, J., Camerer, C. F., Schultz, W. & Rangel, A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* (in the press).
30. Paulus, M. P. & Frank, L. R. Ventromedial prefrontal cortex activation is critical for preference judgments. *Neuroreport* **14**, 1311–1315 (2003).
31. Erk, S., Spitzer, M., Wunderlich, A. P., Galley, L. & Walter, H. Cultural objects modulate reward circuitry. *Neuroreport* **13**, 2499–2503 (2002).
32. Fellows, L. K. & Farah, M. J. The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cereb. Cortex* **17**, 2669–2674 (2007).
33. Lengyel, M. & Dayan, P. Hippocampal contributions to control: the third way. *NIPS* [online] http://books.nips.cc/papers/files/nips20/NIPS2007_0927.pdf (2007).
34. Montague, P. R. *Why Choose This Book?* (Dutton, 2006).
35. Fehr, E. & Camerer, C. F. Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* **11**, 419–427 (2007).
36. Lee, D. Game theory and neural basis of social decision making. *Nature Neurosci.* **11**, 404–409 (2008).
37. Platt, M. L. & Huettel, S. A. Risky business: the neuroeconomics of decision making under uncertainty. *Nature Neurosci.* **11**, 398–403 (2008).
38. Paulus, M. P., Rogalsky, C., Simmons, A., Feinstein, J. S. & Stein, M. B. Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. *Neuroimage* **19**, 1439–1448 (2003).
39. Leland, D. S. & Paulus, M. P. Increased risk-taking decision-making but not altered response to punishment in stimulant-using young adults. *Drug Alcohol Depend.* **78**, 83–90 (2005).
40. Paulus, M. P. *et al.* Prefrontal, parietal, and temporal cortex networks underlie decision-making in the presence of uncertainty. *Neuroimage* **13**, 91–100 (2001).
41. Huettel, S. A., Song, A. W. & McCarthy, G. Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *J. Neurosci.* **25**, 3304–3311 (2005).
42. Bossaerts, P. & Hsu, M. in *Neuroeconomics: Decision Making and the Brain* (eds Glimcher, P. W., Camerer, C. F., Fehr, E. & Poldrack, R. A.) 351–364 (Elsevier, New York, 2008).
43. Preuschoff, K. & Bossaerts, P. Adding prediction risk to the theory of reward learning. *Ann. NY Acad. Sci.* **1104**, 135–146 (2007).
44. Preuschoff, K., Bossaerts, P. & Quartz, S. R. Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* **51**, 381–390 (2006).
45. Tobler, P. N., O'Doherty, J. P., Dolan, R. J. & Schultz, W. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.* **97**, 1621–1632 (2007).
46. Rolls, E. T., McCabe, C. & Redoute, J. Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. *Cereb. Cortex* **18**, 652–663 (2007).
47. Dreher, J. C., Kohn, P. & Berman, K. F. Neural coding of distinct statistical properties of reward information in humans. *Cereb. Cortex* **16**, 561–573 (2006).
48. Preuschoff, K., Quartz, S. R. & Bossaerts, P. Human insula activation reflects prediction errors as well as risk. *J. Neurosci.* **28**, 2745–2752 (2008).
This fMRI study shows that the human insula encodes risk-prediction errors that could be used to learn the riskiness of different options and that are complementary to reward-prediction errors.
49. Tobler, P. N., Fiorillo, C. D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642–1645 (2005).
50. Platt, M. L. & Glimcher, P. W. Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238 (1999).
51. Camerer, C. F. & Weber, M. Recent developments in modelling preferences: uncertainty and ambiguity. *J. Risk Uncertain.* **5**, 325–370 (1992).
52. Hsu, M., Bhatt, M., Adolphs, R., Tranel, D. & Camerer, C. F. Neural systems responding to degrees of uncertainty in human decision-making. *Science* **310**, 1680–1683 (2005).
53. Huettel, S. A., Stowe, C. J., Gordon, E. M., Warner, B. T. & Platt, M. L. Neural signatures of economic preferences for risk and ambiguity. *Neuron* **49**, 765–775 (2006).
54. Hertwig, R., Barron, G., Weber, E. U. & Erev, I. Decisions from experience and the effect of rare events in risky choice. *Psychol. Sci.* **15**, 534–539 (2004).
55. Weller, J. A., Levin, I. P., Shiv, B. & Bechara, A. Neural correlates of adaptive decision making for risky gains and losses. *Psychol. Sci.* **18**, 958–964 (2007).
56. De Martino, B., Kumaran, D., Seymour, B. & Dolan, R. J. Frames, biases, and rational decision-making in the human brain. *Science* **313**, 684–687 (2006).
57. Frederick, S., Loewenstein, G. & O'Donoghue, T. Time discounting and time preference: a critical review. *J. Econ. Lit.* **40**, 351–401 (2002).
58. McClure, S. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. Separate neural systems value immediate and delayed monetary rewards. *Science* **306**, 503–507 (2004).
This fMRI study argued that competing goal-directed valuation systems play a part in decisions that involve choosing between immediate small monetary payoffs and larger but delayed payoffs.
59. McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. Time discounting for primary rewards. *J. Neurosci.* **27**, 5796–5804 (2007).
60. Berns, G. S., Laibson, D. & Loewenstein, G. Intertemporal choice - toward an integrative framework. *Trends Cogn. Sci.* **11**, 482–488 (2007).
61. Kable, J. W. & Glimcher, P. W. The neural correlates of subjective value during intertemporal choice. *Nature Neurosci.* **10**, 1625–1633 (2007).
This fMRI study argued that a single goal-directed valuation system plays a part in decisions that involve choosing between immediate small monetary payoffs and larger but delayed payoffs.

62. Read, D., Frederick, S., Orsel, B. & Rahman, J. Four score and seven years ago from now: the "date/delay" effect in temporal discounting. *Manage. Sci.* **51**, 1326–1335 (1997).
63. Mischel, W. & Underwood, B. Instrumental ideation in delay of gratification. *Child Dev.* **45**, 1083–1088 (1974).
64. Wilson, M. & Daly, M. Do pretty women inspire men to discount the future? *Proc. Biol. Sci.* **271** (Suppl 4), S177–S179 (2004).
65. Berns, G. S. *et al.* Neurobiological substrates of dread. *Science* **312**, 754–758 (2006).
66. Loewenstein, G. Anticipation and the valuation of delayed consumption. *Econ. J.* **97**, 666–684 (1987).
67. Stevens, J. R., Hallinan, E. V. & Hauser, M. D. The ecology and evolution of patience in two New World monkeys. *Biol. Lett.* **1**, 223–226 (2005).
68. Herrnstein, R. J. Relative and absolute strength of response as a function of frequency of reinforcement. *J. Exp. Anal. Behav.* **4**, 267–272 (1961).
69. Mazur, J. E. Estimation of indifference points with an adjusting-delay procedure. *J. Exp. Anal. Behav.* **49**, 37–47 (1988).
70. Corrado, G. S., Sugrue, L. P., Seung, H. S. & Newsome, W. T. Linear-nonlinear-poisson models of primate choice dynamics. *J. Exp. Anal. Behav.* **84**, 581–617 (2005).
71. Newsome, W. T., Britten, K. H. & Movshon, J. A. Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54 (1989).
72. Kim, J. N. & Shadlen, M. N. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nature Neurosci.* **2**, 176–185 (1999).
73. Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
74. Gold, J. I. & Shadlen, M. N. Banburisms and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* **36**, 299–308 (2002).
75. Gold, J. I. & Shadlen, M. N. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* **5**, 10–16 (2001).
76. Heekeren, H. R., Marrett, S. & Ungerleider, L. G. The neural systems that mediate human perceptual decision making. *Nature Rev. Neurosci.* **9**, 467–479 (2008).
77. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neurosci.* **8**, 1704–1711 (2005).
This paper proposed a theoretical model of how the brain might assign control to the different goal and habitual systems.
78. Frank, M. J., Seeberger, L. C. & O'Reilly, R. C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306**, 1940–1943 (2004).
79. Frank, M. J. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw.* **19**, 1120–1136 (2006).
80. de Araujo, I. E., Rolls, E. T., Kringelbach, M. L., McGlone, F. & Phillips, N. Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. *Eur. J. Neurosci.* **18**, 2059–2068 (2003).
81. de Araujo, I. E., Kringelbach, M. L., Rolls, E. T. & McGlone, F. Human cortical responses to water in the mouth, and the effects of thirst. *J. Neurophysiol.* **90**, 1865–1876 (2003).
82. Anderson, A. K. *et al.* Dissociated neural representations of intensity and valence in human olfaction. *Nature Neurosci.* **6**, 196–202 (2003).
83. de Araujo, I. E., Rolls, E. T., Velazco, M. I., Margot, C. & Cayeux, I. Cognitive modulation of olfactory processing. *Neuron* **46**, 671–679 (2005).
84. McClure, S. M. *et al.* Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* **44**, 379–387 (2004).
85. Kringelbach, M. L., O'Doherty, J., Rolls, E. T. & Andrews, C. Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cereb. Cortex* **13**, 1064–1071 (2003).
86. Small, D. M. *et al.* Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* **39**, 701–711 (2003).
87. Blood, A. J. & Zatorre, R. J. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl Acad. Sci. USA* **98**, 11818–11823 (2001).
88. O'Doherty, J. *et al.* Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. *Neuroreport* **11**, 399–403 (2000).
89. Small, D. M., Zatorre, R. J., Dagher, A., Evans, A. C. & Jones-Gotman, M. Changes in brain activity related to eating chocolate: from pleasure to aversion. *Brain* **124**, 1720–1733 (2001).
90. Breiter, H. C., Aharon, I., Kahneman, D., Dale, A. & Shizgal, P. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* **30**, 619–639 (2001).
91. Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L. & Hommer, D. Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* **12**, 3683–3687 (2001).
92. Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C. & Berns, G. S. Human striatal responses to monetary reward depend on saliency. *Neuron* **42**, 509–517 (2004).
93. Peyron, R. *et al.* Haemodynamic brain responses to acute pain in humans: sensory and attentional networks. *Brain* **122**, 1765–1780 (1999).
94. Davis, K. D., Taylor, S. J., Crawley, A. P., Wood, M. L. & Mikulis, D. J. Functional MRI of pain- and attention-related activations in the human cingulate cortex. *J. Neurophysiol.* **77**, 3370–3380 (1997).
95. Seo, H. & Lee, D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* **27**, 8366–8377 (2007).
96. Pecina, S., Smith, K. S. & Berridge, K. C. Hedonic hot spots in the brain. *Neuroscientist* **12**, 500–511 (2006).
97. Berridge, K. C. & Robinson, T. E. Parsing reward. *Trends Neurosci.* **26**, 507–513 (2003).
98. Berridge, K. C. Pleasures of the brain. *Brain Cogn.* **52**, 106–128 (2003).
99. Plassmann, H., O'Doherty, J., Shiv, B. & Rangel, A. Marketing actions can modulate neural representations of experienced pleasantness. *Proc. Natl Acad. Sci. USA* **105**, 1050–1054 (2008).
This paper showed that the level of "experienced pleasantness" encoded in the medial OFC at the time of consuming a wine is modulated by subjects' beliefs about the price of the wine that they are drinking.
100. Montague, P. R., King-Casas, B. & Cohen, J. D. Imaging valuation models in human choice. *Annu. Rev. Neurosci.* **29**, 417–448 (2006).
101. Tremblay, L., Hollerman, J. R. & Schultz, W. Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J. Neurophysiol.* **80**, 964–977 (1998).
102. Hollerman, J. R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neurosci.* **1**, 304–309 (1998).
103. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
This seminal paper proposed the connection between the prediction-error component of reinforcement-learning models and the behaviour of dopamine cells.
104. Mirenowicz, J. & Schultz, W. Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* **72**, 1024–1027 (1994).
105. Schultz, W. Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.* **30**, 259–288 (2007).
106. Schultz, W. Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Curr. Opin. Neurobiol.* **14**, 139–147 (2004).
107. Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
108. Yacubian, J. *et al.* Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *J. Neurosci.* **26**, 9530–9537 (2006).
109. Tanaka, S. C. *et al.* Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neurosci.* **7**, 887–893 (2004).
110. Pagnoni, G., Zink, C. F., Montague, P. R. & Berns, G. S. Activity in human ventral striatum locked to errors of reward prediction. *Nature Neurosci.* **5**, 97–98 (2002).
111. O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H. & Dolan, R. J. Temporal difference models and reward-related learning in the human brain. *Neuron* **38**, 329–337 (2003).
112. Knutson, B., Westdorp, A., Kaiser, E. & Hommer, D. fMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage* **12**, 20–27 (2000).
113. Delgado, M. R., Nystrom, L. E., Fissell, C., Noll, D. C. & Fiez, J. A. Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.* **84**, 3072–3077 (2000).
114. Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
115. Bayer, H. M., Lau, B. & Glimcher, P. W. Statistics of midbrain dopamine neuron spike trains in the awake primate. *J. Neurophysiol.* **98**, 1428–1439 (2007).
116. Seymour, B., Daw, N., Dayan, P., Singer, T. & Dolan, R. Differential encoding of losses and gains in the human striatum. *J. Neurosci.* **27**, 4826–4831 (2007).
117. Daw, N. D., Kakade, S. & Dayan, P. Opponent interactions between serotonin and dopamine. *Neural Netw.* **15**, 603–616 (2002).
118. Lohrenz, T., McCabe, K., Camerer, C. F. & Montague, P. R. Neural signature of fictive learning signals in a sequential investment task. *Proc. Natl Acad. Sci. USA* **104**, 9493–9498 (2007).
119. Camerer, C. F. & Chong, J. K. Self-tuning experience weighted attraction learning in games. *J. Econ. Theory* **133**, 177–198 (2007).
120. Olsson, A. & Phelps, E. A. Social learning of fear. *Nature Neurosci.* **10**, 1095–1102 (2007).
121. Montague, P. R. *et al.* Dynamic gain control of dopamine delivery in freely moving animals. *J. Neurosci.* **24**, 1754–1759 (2004).
122. Tversky, A. & Kahneman, D. Advances in prospect theory cumulative representation of uncertainty. *J. Risk Uncertain.* **5**, 297–323 (1992).
123. Kahneman, D. & Tversky, A. Prospect Theory: an analysis of decision under risk. *Econometrica* **4**, 263–291 (1979).
This seminal paper proposed the PT model for goal-directed valuation in the presence of risk and provided some supporting evidence. It is one of the most cited papers in economics.
124. Chen, K., Lakshminarayanan, V. & Santos, L. How basic are behavioral biases? Evidence from capuchin-monkey trading behavior. *J. Polit. Econ.* **114**, 517–537 (2006).
125. Camerer, C. F. in *Choice, Values, and Frames* (eds Kahneman, D. & Tversky, A.) (Cambridge Univ. Press, Cambridge, 2000).
126. Gilboa, I. & Schmeidler, D. Maxmin expected utility with non-unique prior. *J. Math. Econ.* **28**, 141–153 (1989).
127. Ghirardato, P., Maccheroni, F. & Marinacci, M. Differentiating ambiguity and ambiguity attitude. *J. Econ. Theory* **118**, 133–173 (2004).
128. Nestler, E. J. & Charney, D. S. *The Neurobiology of Mental Illness* (Oxford Univ. Press, Oxford, 2004).
129. Kauer, J. A. & Malenka, R. C. Synaptic plasticity and addiction. *Nature Rev. Neurosci.* **8**, 844–858 (2007).
130. Hyman, S. E., Malenka, R. C. & Nestler, E. J. Neural mechanisms of addiction: the role of reward-related learning and memory. *Annu. Rev. Neurosci.* **29**, 565–598 (2006).
131. Redish, A. D. & Johnson, A. A computational model of craving and obsession. *Ann. NY Acad. Sci.* **1104**, 324–339 (2007).
132. Redish, A. D. Addiction as a computational process gone awry. *Science* **306**, 1944–1947 (2004).
This paper showed how addiction can be conceptualized as a disease of the habit valuation system, using a simple modification of the reinforcement-learning model.
133. Paulus, M. P. Decision-making dysfunctions in psychiatry—altered homeostatic processing? *Science* **318**, 602–606 (2007).
134. Miller, E. K. & Cohen, J. D. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001).
135. Hazy, T. E., Frank, M. J. & O'Reilly, R. C. Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 1601–1613 (2007).
136. Niv, Y., Joel, D. & Dayan, P. A normative perspective on motivation. *Trends Cogn. Sci.* **10**, 375–381 (2006).

Acknowledgements

Financial support from the National Science Foundation (SES-0134618, A.R.) and the Human Frontier Science Program (C.F.C.) is gratefully acknowledged. This work was also supported by a grant from the Gordon and Betty Moore Foundation to the Caltech Brain Imaging Center (A.R., C.F.C.). R.M. acknowledges support from the National Institute on Drug Abuse, the National Institute of Neurological Disorders and Stroke, The Kane Family Foundation, The Angel Williamson Imaging Center and The Dana Foundation.

FURTHER INFORMATION

Author's homepage: www.rnl.caltech.edu

ALL LINKS ARE ACTIVE IN THE ONLINE PDF