Neuroeconomics: The neurobiology of value-based decision-making

Antonio Rangel[1,2], **Colin Camerer**[1], and **P. Read Montague**[3]

[1]HSS and computational and Neural Systems, Caltech, Pasadena, CA 91125

[3]Department of Neuroscience, Computational Psychiatry Unit, Baylor College of Medicine, Houston, TX 77030

Abstract

Neuroeconomics studies the neurobiological and computational basis of value-based decision-making. Its goal is to provide a biologically-based account of human behavior that can be applied in both the natural and the social sciences. In this review we propose a framework for thinking about decision-making that 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 organism makes a choice from several alternatives based on the subjective value that it places on them. Examples include basic animal behaviors such as bee foraging, and complicated human decisions such as trading in the stock-market. Neuroeconomics is a relatively new discipline, which studies the computations that the brain makes 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 should need to make decisions, and some experimental protocols for how to measure them. Psychology provides a wealth of behavioral data showing 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 since computational models identify the kinds of signals and their dynamics required by different value-dependent learning and decision problems. But a full understanding of choice will require descriptions at all these levels.

In this review we propose a framework for thinking about decision-making that allows us to bring together recent findings in the field, highlight some of the most important outstanding

[2]Corresponding author: Antonio Rangel, Caltech, MC 228-77, 1200 E. California Blvd, Pasadena, CA 91125. rangel@hss.caltech.edu.

problems, define a common lexicon that bridges the different disciplines that inform neuroeconomics, and point the way to future applications.

Computations involved in decision-making

The first part of the framework divides the computations required for value-based decision-making into five basic types (FIG 1). The categorization that we propose is based on existing theoretical models of decision-making in economics, psychology, and computer science.[11181] Most models in these disciplines assume, sometimes implicitly, that all of these computations are carried out.

The five steps include the following. First, a representation of the decision problem needs to be computed. This entails identifying internal states (e.g., hunger level), external states (e.g., threat level), and potential courses of action (e.g., pursue a prey or not). Second, the different actions under consideration need to be assigned a value ("valuation"). In order to make sound decisions, these values have to be good predictors of the benefits that are likely to result from each action. Third, the different values need to be compared in order for the organism 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 in order to improve the quality of future decisions ("learning").

We emphasize that these are conceptually useful categories, rather than rigid ones, and that many open questions remain about how well they match the computations made by the brain. For example, it is not know if valuation (step 2) must occur before action selection (step 3), or if both computations are performed in parallel. Nevertheless, the taxonomy is useful because it provides a decomposition of the decision-making process into workable (and testable) constituent processes, it organizes the neuroeconomics literature in terms of the computations that are being studied (for example, by emphasizing that distinct reward-related computations can take place at the valuation, outcome, or learning stages), 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 role 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 or neurobiological basis of this step. Basic open questions include the following: 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: multiple systems

Another important piece of the framework is the existence of multiple valuation systems which are sometimes in agreement, but often in conflict. Based on a sizable body of animal and human behavioral evidence several groups have proposed the existence of three different types of valuation systems: a Pavlovian, a habitual, and a goal-directed system.[4, 29, 30] Here we discuss the computational properties that define them and summarize what is known about their neural basis. There are still many questions open about the physical separability of these learning systems in the brain; however, conceptually, they provide an excellent operational division of the problem according to the style of computations required by each.

It is important to begin the discussion by emphasizing several points. First, the existence of these three distinct valuation systems is based on a rising consensus about how to make sense of a large amount of animal and human behavioral evidence. As we suggest above, these distinction do not necessarily map directly onto separate neural systems[34437, 38] In fact, although the evidence described below points to neural dissociations between some of the components of the three systems, it is likely that that they share common elements. Second, theory has progressed well ahead of the neural data and the precise neural basis of these three distinct valuation is yet to fully be established. Finally, even the exact nature and number of valuation systems is still being debated.

Pavlovian systems

Pavlovian systems assign values to a small set of behaviors that are evolutionarily appropriate responses to particular environmental stimuli. Typical examples include preparatory behaviors 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 behaviors. We refer to these types of behaviors as 'Pavlovian behaviors' and to the systems that assign value to them as the Pavlovian valuation systems.

Many Pavlovian behaviors are innate, or 'hard-wired', responses to specific predetermind stimuli. However, with sufficient training organisms 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 the Pavlovian system and the other two systems is that the Pavlovian system only assigns value to a small set of "prepared" behaviors, and thus has a limited behavioral repertoire. Nonetheless, a wide range of human behaviors that have important economic consequences might be controlled by the Pavlovian valuation system, such as overeating in the presence of food, obsessive-compulsive disorders, and, perhaps, harvesting of immediate and present smaller rewards at the expense of delayed non-present larger rewards.[30, 38]

At first glance, Pavlovian behaviors look like automatic stimulus-triggered responses, and not like instances of value-based choice. However, since Pavlovian responses can be interrupted by other brain systems, Pavlovian behaviors must be assigned something akin to

a 'value' so that they can compete with the actions that are favored by the other valuation systems.

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

The neural basis for Pavlovian responses to specific negative outcomes appear to have a specific and topographic organization along an axis of the dorsal periaqueductal gray.[40] With respect to valence-dependent responses, studies using various species and methods suggest that a network that includes the basolateral amygdala, ventral striatum and the orbitofrontal cortex, is involved in the learning processes by which neutral stimuli become predictive of the value of outcomes.[41][42] In particular, the amygdala has been shown to play a crucial role in influencing some Pavlovian responses.[37][43, 44][45] Specifically, the central nucleus of the amygdala, through its connections to the brainstem nuclei and the core of nucleus accumbens, seems to be involved in non-specific preparatory responses, whereas the basolateral complex of the amygdala seems to be more involved in specific responses through its connections to the hypothalamus and the periaqueductal grey.

Some important questions regarding the Pavlovian valuation systems remained unanswered. How many Pavlovian systems are there and how do they interact with each other? Is there a common carrier of Pavlovian value and if so, how is it encoded? Is learning possible within these systems? How do Pavlovian systems interact with the other valuation systems, for example, in phenomena such as Pavlovian-instrumental-transfer[29]?

Habit systems

In contrast to the Pavlovian system, which values only a small set of responses, the habit systems can learn to assign values to a large number of actions with repeated training. Habit valuation systems exhibit the following key characteristics. First, they learn to assign values to stimulus-response associations (which indicate the action to be taken in a particular state of the world) based on previous experience through a process of trial-and-error (see BOX 3 and the learning section below). Second, subject to some technical qualifications, the system learns 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.[13][94] Third, according to the algorithms that are used to describe learning by this system, since values are learned by trial-and-error, it learns relatively slowly. As a consequence, it may forecast the value of actions incorrectly immediately after a change in the action-reward contingencies. Finally, this system relies on "generalization" to assign action values in novel situations. For example, a rat that has learned to lever-press for liquids in response to a sound cue might repose with a similar behavior when first exposed to a light cue. We refer to the actions controlled by these systems as 'habits' and to 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 (e.g., 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 in the control of habits.[46][47] As discussed below, the projections of dopamine neurons into this area are believed to be important to learning the value of actions. Furthermore, it has been suggested that stimulus-response representations may be encoded in cortico-thalamic loops.[47] Lesion studies in rats have shown that the infralimbic cortex is necessary for the establishment and deployment of habits.[48, 49]

There are many open questions regarding the habit system. 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 (e.g., 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 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 behaviors'. An example of a goal-directed behavior is the decision of what to eat at a new restaurant.

Note that an important difference between the 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 the action-outcome association "lever-press = food" and thus assigns a value to the lever-press equal to the current value of food, which is low since the animal has been fed to satiation. In contrast, the habit system assigns a high value to the action "lever-press" since this is the value that it learned during the pre-satiation training. Thus, while the goal-directed system updates the value of the actions immediately upon a change on the value of outcome, 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,[50] whereas the orbitofrontal cortex (OFC) might be responsible for the encoding of 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)[51–54] Electrophysiology experiments in rats point to the same conclusion.[55] In a further convergence of findings across methods and species, human fMRI studies have shown that BOLD activity in the medial OFC[56–60] and in the DLPFC[57] correlate with behavioral measures of appetitive goal values, and human lesion studies have shown that individuals with damage to the medial OFC have problems making consistent appetitive choices.[61]

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'.[46]

Several questions regarding this system remained unanswered. Are there specialized goal-directed systems for rewards 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 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 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.

Outstanding issues

In addition to the issues listed above, there are several other open questions regarding the different valuation systems. 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 other valuation systems? Lengyel and Dayan[30,62] 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.[17] But 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 there are separate evaluation systems for these types of considerations that are yet to be discovered.

Modulators of the valuation systems

Several variables have been shown to affect the value that the Pavlovian, habitual and goal-directed systems assign to actions. For example, the value assigned to an action might depend, among others, on the riskiness of its associated payoffs, the temporal delay which with they 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 temporal delays on the goal-directed valuation system, since most of the existing evidence pertains to this system. We do not review the literature on social modulators, but refer the reader to recent reviews by Camerer-Fehr[63] and Lee[64].

Risk and uncertainty

All decisions involve some degree of risk in the sense that action-reward associations are probabilistic (BOX 1). We refer to an action with uncertain rewards as a prospect. In order to make good choices the goal-directed system needs to take into account the likelihood of the different outcomes. Two hotly debated questions are: Which are the computations used by the goal-directed system to incorporate risks into its valuations? How does the brain implements such computations?[65]

Early human neuroimaging studies in this topic identified some of the areas involved in making decisions under risk, but were not able to characterize the nature of the computations made by these systems.[66, 676869] The emphasis of more recent work has been in identifying the nature of such computations. Currently, two main competing views are being tested. The first view, which is widely used in financial economics and behavioral ecology, asserts that the brain assigns value to prospects by first computing its statistical moments (such as expected value, variance or coefficient of variation, and skewness) and then aggregating them into a value signal.[370] 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 or prospect theory (BOX 1). 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.

Choices resulting from an expected utility or prospect-theoretic valuation function can be approximated by a weighted sum of the prospects' statistical moments (and vice versa). This makes distinguishing the two models from behavioral data alone complicated. Neural data can provide important insights, although the debate has not yet been settled. Consistent 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[7172] and the medial orbitofrontal cortex[73], and activity that is consistent with risk signals (as measured by the mathematical variance of the prospects) in the striatum[7174], the insula[7573], and the lateral orbitofrontal cortex.[72] Similar risk and expected signals have been found in the midbrain dopamine system in electrophysiology studies in non-human primates.[76] Expected value signals (see BOX 1) have also been found in LIP in non-human primate electrophysiology experiments.[77] Consistent with the second view, a recent human fMRI study has found evidence for a prospect-theory like value signal in a network that includes the ventral and dorsal striatum, ventromedial and ventrolateral PFC, anterior cingulate cortex, and some midbrain dopaminergic regions.[56] Other human fMRI studies have found ventral striatal activation that is correlated with the non-linearity in the transformation of probabilities.[78] The existence of evidence consistent with both views presents an apparent puzzle. A potential resolution that ought to be explored in future studies is that the striatal-prefrontal

network might integrate the statistical moments encoded elsewhere into a value signal that exhibits expected-utility or prospect-theory 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 where probabilities are known. Human behavioral studies have shown that subjects generally exhibit an aversion to choices that are ambiguous,[28] which suggests that a parameter that measures the amount of ambiguity might be encoded in the brain and used to modulate the value signal. Some preliminary human fMRI evidence points to the amygdala, orbitofrontal cortex[79] and anterior insula[80] as areas where such a parameter might be encoded.

Many questions remain to be answered on the topic of risk and valuation. First, little is known about how risk affects the computation of value in the Pavlovian and habitual systems. For example, most reinforcement learning models (see BOX 3) assume that the habit learning system encodes a value signal that cares about expected values, but not about risks. This assumption, however, has not been tested thoroughly. Second, little is known about how the brain learns the risk parameters. For example, some behavioral evidence suggests that the habit and goal-directed systems learn about probabilities differently and that this leads to different probability weighting by the two systems.[81] Finally, more work is required to characterize better the nature of the computations made by the amygdala and insula in decision-making under uncertainty. Preliminary insights suggest that the amygdala might play an asymmetric role in the evaluation of gains and losses. For example, a human lesion study showed that amygdala damage led to poor decision-making in the domain of gains, but not of losses,[82] and a related human fMRI study has shown that the amygdala is differentially activated when subjects are making choices to take risks for large gains and to accept a sure loss.[83]

Temporal discounting

In all real-world situations there is a time lag between decisions and outcomes. From a range of behavioral experiments it is well established that the goal-directed and habitual systems assign lower values to delayed rewards than to immediate ones, a phenomenon known as time discounting.[8] The role of time discounting in the Pavlovian system is not as well understood. As before, this part of the review focuses on the impact of temporal discounting on the goal-directed system, where most of the studies have focused so far.

The current understanding of time discounting parallels the one for risk. Two competing views have been proposed and are being tested using a combination of human behavioral and neuroimaging experiments. One camp has interpreted 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 2), one with a low discount rate and one with a high discount rate.[84–86] In this view, the patience exhibited by any given individual depends on the relative activation of these two systems. In sharp contrast, the other camp has presented human fMRI evidence suggestive of the existence of a single valuation system that discounts future rewards either exponentially or hyperbolically (BOX 2).[87] The existence of evidence consistent with both camps again presents an apparent puzzle. A potential reconciliation is that the striatal-prefrontal network

might integrate information encoded elsewhere in the brain into a single value signal, but that immediate and delayed outcomes might activate different types of information used to compute the value. For example, immediate rewards might activate 'immediacy markers' that increase the valuation signals seen in the striatal-prefrontal network. These same questions are also important to understand from the perspective of brain development. When do value signals get computed in their adult form and how do they contribute to important 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.

Many open questions remain in the domain of temporal discounting. First, the discounting properties of the habitual and Pavlovian systems in humans have not been systematically explored. Second, what are the inputs to the valuation network and why does the aggregation of those inputs produce a hyperbolic-like signal in valuation areas such as the ventral striatum and the medial OFC? Third, the behavioral 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[88], on how they are instructed to think about the rewards[89], and on their arousal level.[90] The mechanisms through which such variables affect the valuation process are unknown. Fourth, several studies have shown that the anticipation of future rewards and punishment can affect subjects' behavioral discount rates.[91][92] The mechanisms through which anticipation affects valuation are also unknown. Finally, several studies have shown that animals make very myopic choices consistent with large hyperbolic discount rates.[93–95][71][96] What is different about how humans 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 choice. Little is known about how the brain does this. The only available theoretical models come from the literature on perceptual decision-making, which has modeled binary perceptual choices as a race-to-barrier diffusion process.[97][98][99–101][102] However, it is unclear whether this class of models also applies to value-based decision-making, and if so, how they might be extended to cases of multi-action choice.

Another important open question has to do with the issue of competition among the different valuation systems that arises 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 evidence is scarce. Daw *et al.*[103] have suggested that the brain arbitrates between the habit and goal-directed valuation systems by assigning control at any given time to the system that has the less uncertain estimate of the true value of the actions. Since the quality of the estimates made by the habit system increase with experience, in practice this means that the habit system should gradually take over the goal-directed system with experience.[17] Frank has proposed a neural-network model for choice between appetitive and aversive habitual valuations.[21, 32]

Understanding how the 'control assignment' problem is resolved is important for several reasons. First, as illustrated in FIG 2, and emphasized by Dayan *et al.*,[38] many apparently puzzling behaviors 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 (e.g., OCD and over-eating) might be due to an inability to assign control to the appropriate system.

There are many important open questions in the domain of action selection. First, in the case of goal-directed choice, does the brain make choices over outcomes, the actions necessary to achieve those outcomes, or both? Second, what is the neural basis of the action selection process in the Pavlovian, habitual, and goal-directed systems? Third, what are the neural mechanisms used to arbitrate between the different controllers, and is there a hierarchy of controllers such that some controllers (e.g., Pavlovian) tend to take precedence over others (e.g., goal-directed)? Fourth, are there any neural markers that can be reliably used to identify goal-directed or habitual behavioral 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 generated by its previous decisions. For example, it is useful for an animal to know if the last food that it consumed led to illness so that it can avoid it in the future.

The computations made by this system, as well as their neural basis, 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 a reward is being consumed correlates with subjective reports about the quality of the experience for odors[104, 105106107], tastes[108109110] and even music.[111] Several related studies have also shown that the activity in the medial OFC parallels the reduction in outcome value that one would expect after a subject is fed to satiation.[112113] 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 reinforces such as monetary payoffs.[114115116] Analogous results have been found for negative experiences: subjective reports of pain intensity correlated in human fMRI studies with activity in the insula and the anterior cingulate cortex.[117, 118]

Evidence for the neural basis of the outcome value signal also come from animal studies. A monkey electrophysiology experiment by Lee and colleagues has recently found outcome value signals in the dorsal anterior cingulate cortex.[119] In a series of provocative rat studies, Berridge and colleagues have shown that it is possible to increase outward manifestations of 'liking' in rats (e.g., tongue protrusions) by activating subsets of the ventral pallidum and nucleus accumbens using opioid agonists.[109, 120–122] Interestingly, and consistent with the

hypothesis that outcome evaluation signals play a role in learning, the rats exposed to opioids agonists subsequently consume more of the reward that was paired with it.

Some recent human fMRI experiments have also provided novel insights about the computational properties of the outcome-value signal. De Araujo et al[107] showed that activity in the medial OFC in response to an odor depended on whether subjects believed that they were smelling cheddar cheese or a sweaty sock. In addition, Plassman et al.[123] showed that activity in the medial OFC in response to the consumption of a wine depended on beliefs about its price, and McClure et al.[108] showed that the outcome-valuation signal after consumption of a 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 is the precise network responsible for computing positive and negative outcome values in different types of domains? How do positive and negative outcome-valuation signals get integrated? How are they 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 behaviors are innate responses to environmental stimuli, most forms of behavior 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 behaviors in every state. This is a non-trivial learning problem given that animals and humans have limited computational power, yet they can deploy a large number of behavioral responses. Second, the valuation systems must learn to assign values to actions 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 questions, the one that is best understood is the learning of action-values by the habit system. In this area there has been a very productive interplay between theoretical models from computer science (BOX 3) 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 made by the habit system.[1124] The basic idea behind these models is that a prediction-error signal is computed after every choice. The signal is called a prediction error because it measures the quality of the forecast implicit in the previous valuation (BOX 3). Every time a learning event occurs, the value of 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.[125–128][129, 130] The connection between these signals and the reinforcement-learning models was made in a series of papers in the 1990s by Montague and

colleagues.[127, 131] In later years several fMRI studies have 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.[58, 114, 132–137]

Although the existing evidence suggests 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 by Bayer and Glimcher[138] suggests that the phasic firing rates of midbrain dopamine neurons might only encode for 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 part (henceforth, the negative prediction error), which is also essential for learning. Several possibilities have been proposed. A subsequent analsysis of the previous experiment suggested that the magnitude of the negative prediction errors might be encoded in the timing of the firing-and-pause patterns of the dopamine cells.[139] Some human fMRI studies have found a BOLD signal in the amygdala that resembles a negative-prediction error[132], but others have failed to replicate the finding and have found instead evidence for both types of prediction error in different parts of the striatum.[140] In turn, Daw and Dayan[141] have 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, Tobler *et al.*[76] have 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. The exact nature of these adjustments remains an open question.[70] Finally, Lohrenz *et al.*[142] have shown that the habit system can also learn from observing the outcomes of actions that it did not take, as opposed to learning using only direct experience. 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 (including imitative learning from actions of others).[143]

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 where 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 to generalize across them? How do the different learning systems incorporate expected uncertainty about the feedback signals?[70] To what extent can the different value systems learn by observation as opposed to direct experience?[144]

The next five 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 challenging since, as we have seen, there seems to be at least three valuation systems at work that fight over control of the decision-making process.

Nevertheless, several developments are welcome and give reason to be hopeful that significant progress will be made over the next five years on answering many of the questions outlined here. First, is the close connection between theory and experiments, and the widespread use of theory-driven experimentation (including behavioral parameters inferred from choices that can be linked across subjects or trials to brain activity). Second is the rapid adoption of new technologies, such as fast cyclic voltametry in freely-moving animals[145], which permits quasi-real-time monitoring of neurotransmitter levels for long periods of time. Third, is the investigation of decision-making phenomena using different species and experimental methods, which permits more rapid progress that would be made otherwise.

This is good news because the range of potential applications is significant. The most important area of application is psychiatry. Many psychiatric diseases involve a failure of one or more of the decision-making processes described here (BOX 4). A better understanding of these processes should lead to improved diagnoses 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, a better 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 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 policing, war, and fast-paced financial markets.

Acknowledgments

GLOSSARY

Propositional logic system.	A cognitive system that makes predictions about the world based on known pieces of information.
Race-to-barrier diffusion process.	A stochastic process that terminates when the variable of interest reaches a certain threshold value.

Credit assignment problem.	The problem of crediting rewards to particular actions in complex environments.
Expected utility theory.	A theory that states that the value of a prospect (or random rewards) equals the sum of the value of the potential outcomes weighted by their probability.
Prospect theory.	An alternative theory of how to evaluate prospects (see BOX 1 for details).

References

1. Sutton, RS.; Barto, AG. Reinforcement Learning: An Introduction. Cambridge: MIT Press; 1998.

2. Association AP. 2000

3. Bossaerts, P.; Hsu, M. Neuroeconomics: Decision Making and the Brain. Glimcher, PW.; Camerer, CF.; Fehr, E.; Poldrack, RA., editors. New York: Elsevier; 2008.

4. Balleine, BW.; Daw, N.; O'Doherty, J. Neuroeconomics: Decision-Making and the Brain. Glimcher, PW.; Fehr, E.; Camerer, C.; Poldrack, RA., editors. New York: Elsevier; 2008.

5. Nestler, EJ.; Charney, DS. The Neurobiology of Mental Illness. Oxford: Oxford University Press; 2004.

6. Kauer JA, Malenka RC. Synaptic plasticity and addiction. Nat Rev Neurosci. 2007; 8:844–858. [PubMed: 17948030]

7. Hyman SE, Malenka RC, Nestler EJ. Neural Mechanisms of Addiction: The Role of Reward-Related Learning and Memory. Annu Rev Neurosci. 2006

8. Frederick S, Loewenstein G, O'Donoghue T. Time discounting and time preference: a critical review. Journal of Economic Literature. 2002; 40:351–401.

9. Niv, Y.; Montague, PR. Multiple Forms of Value Learning and the Function of Dopamine. Glimcher, PW.; Fehr, E.; Camerer, C.; Poldrack, RA., editors. New York: Elsevier; 2008.

10. Puterman, ML. Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York: Wiley; 1994.

11. Busemeyer, JR.; Johnson, JG. Handbook of Judgment and Decision Making. Koehler, D.; Narvey, N., editors. New York: Blackwell Publishing, Co.; 2004. p. 133-154.

12. Tversky A, Kahneman D. Advances in prospect theory cumulative representation of uncertainty. J Risk and Uncertainty. 1992; 5:297–323.

13. Kahneman D, Tversky A. Prospect Theory: an analysis of decision under risk. Econometrica. 1979; 4:263–291.

14. Redish AD, Johnson A. A computational model of craving and obsession. Ann N Y Acad Sci. 2007; 1104:324–339. [PubMed: 17595292]

15. Redish AD. Addiction as a computational process gone awry. Science. 2004; 306:1944–1947. [PubMed: 15591205]

16. Paulus MP. Decision-making dysfunctions in psychiatry--altered homeostatic processing? Science. 2007; 318:602–606. [PubMed: 17962553]

17. Montague, PR. Why Choose This Book?. Dutton Press; 2006.

18. Mas-Colell, A.; Whinston, M.; Green, J. Microeconomic Theory. Cambridge: Cambridge University Press; 1995.

19. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 2001; 24:167–202. [PubMed: 11283309]

20. Hazy TE, Frank MJ, O'Reilly RC. Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. Philos Trans R Soc Lond B Biol Sci. 2007; 362:1601–1613. [PubMed: 17428778]

21. Frank MJ. Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. Neural Netw. 2006; 19:1120–1136. [PubMed: 16945502]

22. Watkins, CJCH. Cambridge, UK: Cambridge University; 1989.

23. Chen K, Lakshminarayanan V, Santos L. How Basic Are Behavioral Biases? Evidence from Capuchin-Monkey Trading Behavior. Journal of Political Economy. 2006; 114:517–537.

24. Watkins CJCH, Dayan P. Q-learning. Machine Learning. 1192; 8:279–292.

25. Niv Y, Joel D, Dayan P. A normative perspective on motivation. Trends Cogn Sci. 2006; 10:375–381. [PubMed: 16843041]

26. Camerer, CF. Choice, Values, and Frames. Kahneman, D.; Tversky, A., editors. Cambridge: Cambridge University Press; 2000.

27. Rummery, GA.; Niranjan, M. Technical Report CUED/F-INENG/TR. Cambridge University: Engineering Department; 1994.

28. Camerer CF, Weber M. Recent Developments in Modelling Preferences -- Uncertainty and Ambiguity. Journal of Risk and Uncertainty. 1992; 5:325–370.

29. Dickison, A.; Balleine, BW. Learning, Motivation & Emotion, Volume 3 of Steven's Handbook of Experimental Psychology. Gallistel, C., editor. New York: John Wiley & Sons; 2002. p. 497-533.

30. Dayan, P. Better Than Conscious? Implications for Performance and Institutional Analysis. Engel, C.; Singer, W., editors. Cambridge, MA: MIT Press; 2008.

31. Barto, AG. Models of information processing in the basal ganglia. Houk, JC.; Davis, JL.; Beiser, DG., editors. Cambridge: MIT Press; 1995.

32. Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science. 2004; 306:1940–1943. [PubMed: 15528409]

33. Gilboa I, Schmeidler D. Maxmin expected utility with non-unique prior. Journal of Mathematical Economics. 1989; 28:141–153.

34. Bouton, ME. Learning and Behavior: A Contemporary Synthesis. Sunderland, Massachusetts: Sinauer Associates, Inc. Publishers; 2007.

35. Dayan P, Balleine BW. Reward, motivation, and reinforcement learning. Neuron. 2002; 36:285–298. [PubMed: 12383782]

36. Ghirardato P, Maccheroni F, Marinacci M. Differentiating ambiguity and ambiguity attitude. Journal of Economic Theory. 2004; 118:133–173.

37. Dayan, P.; Seymour, B. Neuroeconomics: Decision Making and the Brain. Glimcher, PW.; Camerer, CF.; Fehr, E.; Poldrack, RA., editors. New York: Elsevier; 2008.

38. Dayan P, Niv Y, Seymour B, Daw ND. The misbehavior of value and the discipline of the will. Neural Netw. 2006; 19:1153–1160. [PubMed: 16938432]

39. Dayan, P.; Abbott, LR. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. Cambridge, MA: MIT Press; 1999.

40. Keay KA, Bandler R. Parallel circuits mediating distinct emotional coping reactions to different types of stress. Neurosci Biobehav Rev. 2001; 25:669–678. [PubMed: 11801292]

41. Cardinal RN, Parkinson JA, Hall J, Everitt BJ. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev. 2002; 26:321–352. [PubMed: 12034134]

42. Holland PC, Gallagher M. Amygdala-frontal interactions and reward expectancy. Curr Opin Neurobiol. 2004; 14:148–155. [PubMed: 15082318]

43. Fendt M, Fanselow MS. The neuroanatomical and neurochemical basis of conditioned fear. Neurosci Biobehav Rev. 1999; 23:743–760. [PubMed: 10392663]

44. Adams DB. Brain mechanisms of aggressive behavior: an updated review. Neurosci Biobehav Rev. 2006; 30:304–318. [PubMed: 16289283]

45. Niv, Y. Neuroscience. Jerusalem: Hebrew University; 2007.

46. Balleine BW. Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. Physiol Behav. 2005; 86:717–730. [PubMed: 16257019]

47. Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. Nat Rev Neurosci. 2006; 7:464–476. [PubMed: 16715055]

48. Killcross S, Coutureau E. Coordination of actions and habits in the medial prefrontal cortex of rats. Cereb Cortex. 2003; 13:400–408. [PubMed: 12631569]

49. Coutureau E, Killcross S. Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. Behav Brain Res. 2003; 146:167–174. [PubMed: 14643469]

50. Yin HH, Knowlton BJ, Balleine BW. Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. Eur J Neurosci. 2005; 22:505–512. [PubMed: 16045503]

51. Wallis JD, Miller EK. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. Eur J Neurosci. 2003; 18:2069–2081. [PubMed: 14622240]

52. Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. Nature. 2006; 441:223–226. [PubMed: 16633341]

53. Wallis JD. Orbitofrontal cortex and its contribution to decision-making. Annu Rev Neurosci. 2007; 30:31–56. [PubMed: 17417936]

54. Barraclough DJ, Conroy ML, Lee D. Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci. 2004; 7:404–410. [PubMed: 15004564]

55. Schoenbaum G, Roesch M. Orbitofrontal cortex, associative learning, and expectancies. Neuron. 2005; 47:633–636. [PubMed: 16129393]

56. Tom SM, Fox CR, Trepel C, Poldrack RA. The Neural Basis of Loss Aversion in Decision-Making Under Risk. Science. 2007; 315:515–518. [PubMed: 17255512]

57. Plassmann H, O'Doherty J, Rangel A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J Neurosci. 2007; 27:9984–9988. [PubMed: 17855612]

58. Hare T, O'Doherty J, Camerer CF, Schultz W, Rangel A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. J. Neuroscience. *in press*.

59. Paulus MP, Frank LR. Ventromedial prefrontal cortex activation is critical for preference judgments. Neuroreport. 2003; 14:1311–1315. [PubMed: 12876463]

60. Erk S, Spitzer M, Wunderlich AP, Galley L, Walter H. Cultural Objects Modulate Reward Circuitry. Neuroreport. 2002; 13:2499–2503. [PubMed: 12499856]

61. Fellows LK, Farah MJ. The Role of Ventromedial Prefrontal Cortex in Decision Making: Judgment under Uncertainty or Judgment Per Se? Cerebral Cortex. 2007 *****Advanced publication*****.

62. Lengyel M, Dayan P. Hippocampal contributions to control: The third way. NIPS. 2007

63. Fehr E, Camerer CF. Social neuroeconomics: the neural circuitry of social preferences. Trends Cogn Sci. 2007; 11:419–427. [PubMed: 17913566]

64. Lee D. Game theory and neural basis of social decision making. Nat Neurosci. 2008; 11:404–409. [PubMed: 18368047]

65. Platt ML, Huettel SA. Risky business: the neuroeconomics of decision making under uncertainty. Nat Neurosci. 2008; 11:398–403. [PubMed: 18368046]

66. Paulus MP, Rogalsky C, Simmons A, Feinstein JS, Stein MB. Increased activation in the right insula during risk-taking decision making is related to harm avoidance and neuroticism. Neuroimage. 2003; 19:1439–1448. [PubMed: 12948701]

67. Leland DS, Paulus MP. Increased risk-taking decision-making but not altered response to punishment in stimulant-using young adults. Drug Alcohol Depend. 2005; 78:83–90. [PubMed: 15769561]

68. Paulus MP, et al. Prefrontal, parietal, and temporal cortex networks underlie decision-making in the presence of uncertainty. Neuroimage. 2001; 13:91–100. [PubMed: 11133312]

69. Huettel SA, Song AW, McCarthy G. Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. J Neurosci. 2005; 25:3304–3311. [PubMed: 15800185]

70. Preuschoff K, Bossaerts P. Adding prediction risk to the theory of reward learning. Ann N Y Acad Sci. 2007; 1104:135–146. [PubMed: 17344526]

71. Preuschoff K, Bossaerts P, Quartz SR. Neural Differentiation of Expected Reward and Risk in Human Subcortical Structures. Neuron. 2006; 51:381–390. [PubMed: 16880132]

72. Tobler PN, O'Doherty JP, Dolan RJ, Schultz W. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. J Neurophysiol. 2007; 97:1621–1632. [PubMed: 17122317]

73. Rolls ET, McCabe C, Redoute J. Expected Value, Reward Outcome, and Temporal Difference Error Representations in a Probabilistic Decision Task. Cereb Cortex. 2007

74. Dreher JC, Kohn P, Berman KF. Neural coding of distinct statistical properties of reward information in humans. Cereb Cortex. 2006; 16:561–573. [PubMed: 16033924]

75. Preuschoff K, Quartz SR, Bossaerts P. Human Insula Activation Reflects Prediction Errors As Well As Risk. 2007 *under review*.

76. Tobler PN, Fiorillo CD, Schultz W. Adaptive coding of reward value by dopamine neurons. Science. 2005; 307:1642–1645. [PubMed: 15761155]

77. Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. Nature. 1999; 400:233–238. [PubMed: 10421364]

78. Hsu M, Krajbich I, Zhao M, Camerer CF. Neural evidence for nonlinear weighting of probabilities in risky choice. 2007 *under review*.

79. Hsu M, Bhatt M, Adolphs R, Tranel D, Camerer CF. Neural Systems Responding to Degrees of Uncertainty in Human Decision-Making. Science. 2005; 310:1680–1683. [PubMed: 16339445]

80. Huettel SA, Stowe CJ, Gordon EM, Warner BT, Platt ML. Neural Signatures of Economic Preferences for Risk and Ambiguity. Neuron. 2006; 49:765–775. [PubMed: 16504951]

81. Hertwig R, Barron G, Weber EU, Erev I. Decisions from experience and the effect of rare events in risky choice. Psychol Sci. 2004; 15:534–539. [PubMed: 15270998]

82. Weller JA, Levin IP, Shiv B, Bechara A. Neural correlates of adaptive decision making for risky gains and losses. Psychol Sci. 2007; 18:958–964. [PubMed: 17958709]

83. De Martino B, Kumaran D, Seymour B, Dolan RJ. Frames, biases, and rational decision-making in the human brain. Science. 2006; 313:684–687. [PubMed: 16888142]

84. McClure SM, Laibson DI, Loewenstein G, Cohen JD. Separate neural systems value immediate and delayed monetary rewards. Science. 2004; 306:503–507. [PubMed: 15486304]

85. McClure SM, Ericson KM, Laibson DI, Loewenstein G, Cohen JD. Time discounting for primary rewards. J Neurosci. 2007; 27:5796–5804. [PubMed: 17522323]

86. Berns GS, Laibson D, Loewenstein G. Intertemporal choice - toward an integrative framework. Trends Cogn Sci. 2007; 11:482–488. [PubMed: 17980645]

87. Kable JW, Glimcher PW. The neural correlates of subjective value during intertemporal choice. Nat Neurosci. 2007; 10:1625–1633. [PubMed: 17982449]

88. Read D, Frederick S, Orsel B, Rahman J. Four score and seven years ago from now: the "date/delay" effect in temporal discounting. Management Science. 1997; 51:1326–1335.

89. Mischel W, Underwood B. Instrumental ideation in delay of gratification. Child Dev. 1974; 45:1083–1088. [PubMed: 4143856]

90. Wilson M, Daly M. Do pretty women inspire men to discount the future? Proc Biol Sci. 2004; 271(Suppl 4):S177–S179. [PubMed: 15252976]

91. Berns GS, et al. Neurobiological substrates of dread. Science. 2006; 312:754–758. [PubMed: 16675703]

92. Loewenstein G. Anticipation and the valuation of delayed consumption. Economic Journal. 1987; 97

93. Stevens JR, Hallinan EV, Hauser MD. The ecology and evolution of patience in two New World monkeys. Biol Lett. 2005; 1:223–226. [PubMed: 17148172]

94. Herrnstein RJ. Relative and absolute strength of response as a function of frequency of reinforcement. J Exp Anal Behav. 1961; 4:267–272. [PubMed: 13713775]

95. Mazur JE. Estimation of indifference points with an adjusting-delay procedure. J Exp Anal Behav. 1988; 49:37–47. [PubMed: 3346621]

96. Corrado GS, Sugrue LP, Seung HS, Newsome WT. Linear-Nonlinear-Poisson models of primate choice dynamics. J Exp Anal Behav. 2005; 84:581–617. [PubMed: 16596981]

97. Newsome WT, Britten KH, Movshon JA. Neuronal correlates of a perceptual decision. Nature. 1989; 341:52–54. [PubMed: 2770878]

98. Kim JN, Shadlen MN. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat Neurosci. 1999; 2:176–185. [PubMed: 10195203]

99. Gold JI, Shadlen MN. The neural basis of decision making. Annu Rev Neurosci. 2007; 30:535–574. [PubMed: 17600525]

100. Gold JI, Shadlen MN. Banburisms and the Brain: Decoding the Relationship Between Sensory Stimuli, Decisions, and Reward. Neuron. 2002; 36:299–308. [PubMed: 12383783]

101. Gold JI, Shadlen MN. Neural computations that underlie decisions about sensory stimuli. TRENDS in Cognitive Sciences. 2001; 5:10–16. [PubMed: 11164731]

102. Krajbich I, Armel C, Rangel A. Visual attention drives the computation of value in goal-directed choice. 2007 *under review*.

103. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 2005; 8:1704–1711. [PubMed: 16286932]

104. de Araujo IE, Rolls ET, Kringelbach ML, McGlone F, Phillips N. Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain. Eur J Neurosci. 2003; 18:2059–2068. [PubMed: 14622239]

105. de Araujo IE, Kringelbach ML, Rolls ET, McGlone F. Human cortical responses to water in the mouth, and the effects of thirst. J Neurophysiol. 2003; 90:1865–1876. [PubMed: 12773496]

106. Anderson AK, et al. Dissociated neural representations of intensity and valence in human olfaction. Nat Neurosci. 2003; 6:196–202. [PubMed: 12536208]

107. de Araujo IE, Rolls ET, Velazco MI, Margot C, Cayeux I. Cognitive modulation of olfactory processing. Neuron. 2005; 46:671–679. [PubMed: 15944134]

108. McClure SM, et al. Neural correlates of behavioral preference for culturally familiar drinks. Neuron. 2004; 44:379–387. [PubMed: 15473974]

109. Kringelbach ML, O'Doherty J, Rolls ET, Andrews C. Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. Cereb Cortex. 2003; 13:1064–1071. [PubMed: 12967923]

110. Small DM, et al. Dissociation of neural representation of intensity and affective valuation in human gustation. Neuron. 2003; 39:701–711. [PubMed: 12925283]

111. Blood AJ, Zatorre RJ. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc Natl Acad Sci U S A. 2001; 98:11818–11823. [PubMed: 11573015]

112. O'Doherty J, et al. Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. Neuroreport. 2000; 11:399–403. [PubMed: 10674494]

113. Small DM, Zatorre RJ, Dagher A, Evans AC, Jones-Gotman M. Changes in brain activity related to eating chocolate: from pleasure to aversion. Brain. 2001; 124:1720–1733. [PubMed: 11522575]

114. Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron. 2001; 30:619–639. [PubMed: 11395019]

115. Knutson B, Fong GW, Adams CM, Varner JL, Hommer D. Dissociation of reward anticipation and outcome with event-related fMRI. Neuroreport. 2001; 12:3683–3687. [PubMed: 11726774]

116. Zink CF, Pagnoni G, Martin-Skurski ME, Chappelow JC, Berns GS. Human striatal responses to monetary reward depend on saliency. Neuron. 2004; 42:509–517. [PubMed: 15134646]

117. Peyron R, et al. Haemodynamic brain responses to acute pain in humans: sensory and attentional networks. Brain. 1999; 122(Pt 9):1765–1780. [PubMed: 10468515]

118. Davis KD, Taylor SJ, Crawley AP, Wood ML, Mikulis DJ. Functional MRI of pain- and attention-related activations in the human cingulate cortex. J Neurophysiol. 1997; 77:3370–3380. [PubMed: 9212281]

119. Seo H, Lee D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J Neurosci. 2007; 27:8366–8377. [PubMed: 17670983]

120. Pecina S, Smith KS, Berridge KC. Hedonic hot spots in the brain. Neuroscientist. 2006; 12:500–511. [PubMed: 17079516]

121. Berridge KC, Robinson TE. Parsing reward. Trends in Neurosciences. 2003; 26

122. Berridge KC. Pleasures of the brain. Brain Cogn. 2003; 52:106–128. [PubMed: 12812810]

123. Plassmann H, O'Doherty J, Shiv B, Rangel A. Marketing Actions Can Modulate Neural Representations of Experienced Pleasantness. 2007 *under review*.

124. Montague PR, King-Casas B, Cohen JD. Imaging Valuation Models in Human Choice. Annu Rev Neurosci. 2006

125. Tremblay L, Hollerman JR, Schultz W. Modifications of reward expectation-related neuronal activity during learning in primate striatum. J Neurophysiol. 1998; 80:964–977. [PubMed: 9705482]

126. Hollerman JR, Schultz W. Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci. 1998; 1:304–309. [PubMed: 10195164]

127. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997; 275:1593–1599. [PubMed: 9054347]

128. Mirenowicz J, Schultz W. Importance of unpredictability for reward responses in primate dopamine neurons. J Neurophysiol. 1994; 72:1024–1027. [PubMed: 7983508]

129. Schultz W. Multiple dopamine functions at different time courses. Annu Rev Neurosci. 2007; 30:259–288. [PubMed: 17600522]

130. Schultz W. Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. Curr Opin Neurobiol. 2004; 14:139–147. [PubMed: 15082317]

131. Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci. 1996; 16:1936–1947. [PubMed: 8774460]

132. Yacubian J, et al. Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. J Neurosci. 2006; 26:9530–9537. [PubMed: 16971537]

133. Tanaka SC, et al. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat Neurosci. 2004; 7:887–893. [PubMed: 15235607]

134. Pagnoni G, Zink CF, Montague PR, Berns GS. Activity in human ventral striatum locked to errors of reward prediction. Nat Neurosci. 2002; 5:97–98. [PubMed: 11802175]

135. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. Neuron. 2003; 38:329–337. [PubMed: 12718865]

136. Knutson B, Westdorp A, Kaiser E, Hommer D. FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage. 2000; 12:20–27. [PubMed: 10875899]

137. Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol. 2000; 84:3072–3077. [PubMed: 11110834]

138. Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron. 2005; 47:129–141. [PubMed: 15996553]

139. Bayer HM, Lau B, Glimcher PW. Statistics of midbrain dopamine neuron spike trains in the awake primate. J Neurophysiol. 2007; 98:1428–1439. [PubMed: 17615124]

140. Seymour B, Daw N, Dayan P, Singer T, Dolan R. Differential encoding of losses and gains in the human striatum. J Neurosci. 2007; 27:4826–4831. [PubMed: 17475790]

141. Daw ND, Kakade S, Dayan P. Opponent interactions between serotonin and dopamine. Neural Netw. 2002; 15:603–616. [PubMed: 12371515]

142. Lohrenz T, McCabe K, Camerer CF, Montague PR. Neural signature of fictive learning signals in a sequential investment task. Proc Natl Acad Sci U S A. 2007; 104:9493–9498. [PubMed: 17519340]

143. Ho T-H, Camerer CF, Chong JK. Self-tuning experience weighted attraction learning in games. Journal of Economic Theory. 2007; 133:177–198.

144. Olsson A, Phelps EA. Social learning of fear. Nat Neurosci. 2007; 10:1095–1102. [PubMed: 17726475]

145. Montague PR, et al. Dynamic gain control of dopamine delivery in freely moving animals. J Neurosci. 2004; 24:1754–1759. [PubMed: 14973252]

## Box 1. Risk modulators of value in the goal-directed system

Many decisions involve the valuation of rewards and costs that occur probabilistically, often called prospects. A central question for the valuation systems is how to incorporate probability in the assignment of value. There are two dominant theories in economics about how this is done. In **expected utility** (EU) **theory**, 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.[3] 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 behavior. 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 outcomes now depend on a reference point r, and they are weighted by a non-linear function $\pi()$ of the objective probabilities.[12, 13] 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 below illustrates the usual assumptions that are imposed in the value and probability functions by the two theories. As shown on the left, 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 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 on the right. 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. The figure on the bottom-right illustrates the version of prospect-theory in which small probabilities are over-weighted and large probabilities are underweighted. PT has been successful in explaining some behavior inconsistent with EU theory in behavioral experiments with humans[13] and monkeys,[23] as well as economic field evidence.[26]

Neuroeconomists make a distinction between prospects involving risk and ambiguity. Risk refers to a situation where all of the probabilities are known. Ambiguity refers to a situation where 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.[28 33 36]

Box 2. 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 one, 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 lower k is associated with less discounting (i.e., more patience). In the second one, known as **exponential discounting**, the corresponding discount factor is γ^t. Note that a value of γ closer to one is associated with a lower discount factor. An important distinction between the two models is illustrated in the left figures below, 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 behavioral studies of goal-directed behavior with adequate statistical power, hyperbolic discount functions always fits observed behavior better than exponential functions.[8] 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 modeling. This property requires that if a reward A is assigned a higher value than B at time t, then the same reward is also assigned a higher value when evaluated at any time $t-k$. Under hyperbolic discounting, in contrast, the relative valuation between the two actions depends on when the choice is made. This is known as **dynamic inconsistency**. The figures on the right illustrate this difference. It depicts 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.

EXPONENTIAL

$V=\delta^t R$

R

R_A

R_B

HYPERBOLIC

$V=R/(1+kt)$

R

R_A

R_B
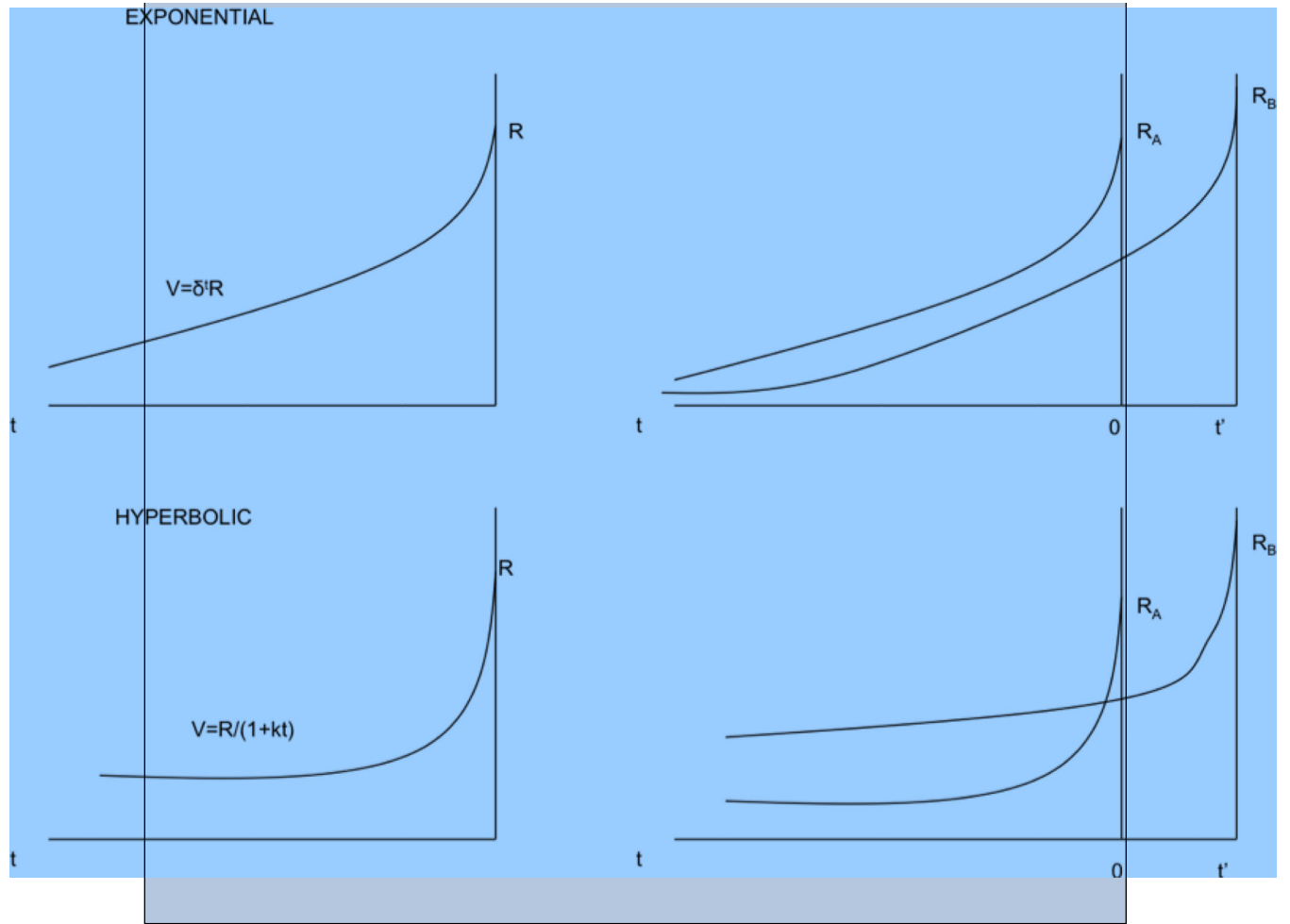
Box 3. Reinforcement learning models action value learning in the habitual system

Several models from computer science have proved useful in modeling how the habitual system learns to assign values to actions.[149] All of these models have the following structure, which is known as a **Markovian decision problem**[10]: (1) the animal can be in a finite set of states and can take a finite set of actions; (2) 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 (3) every time step the animal obtains an action and state-contingent reward $r(a,s)$. A behavioral 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 π, it needs to compute the value of taking every action a in every state s. This is given by

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

where r_{t+k} denotes the reward received at time $t+k$ and $\gamma > 0$ 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 $\hat{Q}(s,a)$ denote the estimate that the system has at any point in time. Equation (1) can be rewritten in recursive form as

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

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**

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

which is a sample measure of how close the estimate is to satisfying expression (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:

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

where η is a number between 0 and 1 that determines the speed of learning. This model is known as **Q-learning**[22] and it satisfies one important property: subject to some technical conditions, the estimated action values converge to those generated by the optimal policy.[24] 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[27] and the actor-critic model.[3135] 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 are 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 reward function. Second, they are able to explain a wide range of conditioning behaviors associated with the habitual system, such as blocking, overshadowing, and inhibitory conditioning.[39] 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. An important open question is what are the algorithms used by the Pavlovian and goal-directed systems to update their values based on feedback from the environment.

Box 4. From neuroeconomics to computational psychiatry

In some situations, the brain's decision-making processes function so differently from our societal norms that that we label the ensuing behaviors and perceptions a psychiatric disease. While there are real philosophical issues about what constitutes a mental illness, the medical community recognizes and categorizes them according to well-accepted diagnostic criteria which, so far, have relied almost exclusively on agreed-upon collections of behavioral features.[2] Despite the almost exclusive emphasis on the behavioral categorization of mental disease, neuroscientists have accumulated a substantial amount of neurobiological data that impinges directly on psychiatric illness.[5] 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.[6, 7] Thus, there is a substantial body of biological data at one end paired with detailed descriptions of the behavioral outcomes at the other end. However, there is precious little in-between. This situation is exemplified by the array of psychotropic drugs that act on known neuromodulatory systems, and produce known changes in behavior, but for which we have little understanding about how they alter the brain's decision-making and perceptual mechanisms. This situation presents an opportunity for neuroeconomics and other computationally oriented sciences to connect the growing body of biological knowledge to the behavioral endpoints.

Reinforcement-learning models provide insight into how the habitual system learns to assign value in a wide range of situations, as well as insight into important neuromodulatory systems, such as dopamine, which are perturbed in a range of mental diseases.[5] Thus, computational models of reinforcement learning provide a new language for understanding mental illness, and a starting point for connecting detailed neural substrates to behavioral outcomes. For example, reinforcement-learning models predict the existence of valuation malfunctions where a drug, disease, or developmental event perturbs the brain's capacity to assign appropriate value to behavioral acts or mental states.[14–17]

Disorders of decision-making can also arise at the action-selection stage, especially when there are conflicts among the valuation systems. This presents the posibility 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, OCD, and obesity. These issues relate directly to the idea of executive control and the way that it is affected by mental disease. While older ideas about executive control have been useful in guiding a description of the phenomenology of control, it is our opinion that future progress will require more computational approaches because only through such models can competing ideas be clearly differentiated. Such efforts are already well underway and a variety of modeling efforts have been applied to executive control and decision-making in humans.[19–21]

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 role in mood disorders such as depression and in Parkinsons' disease.[2532]
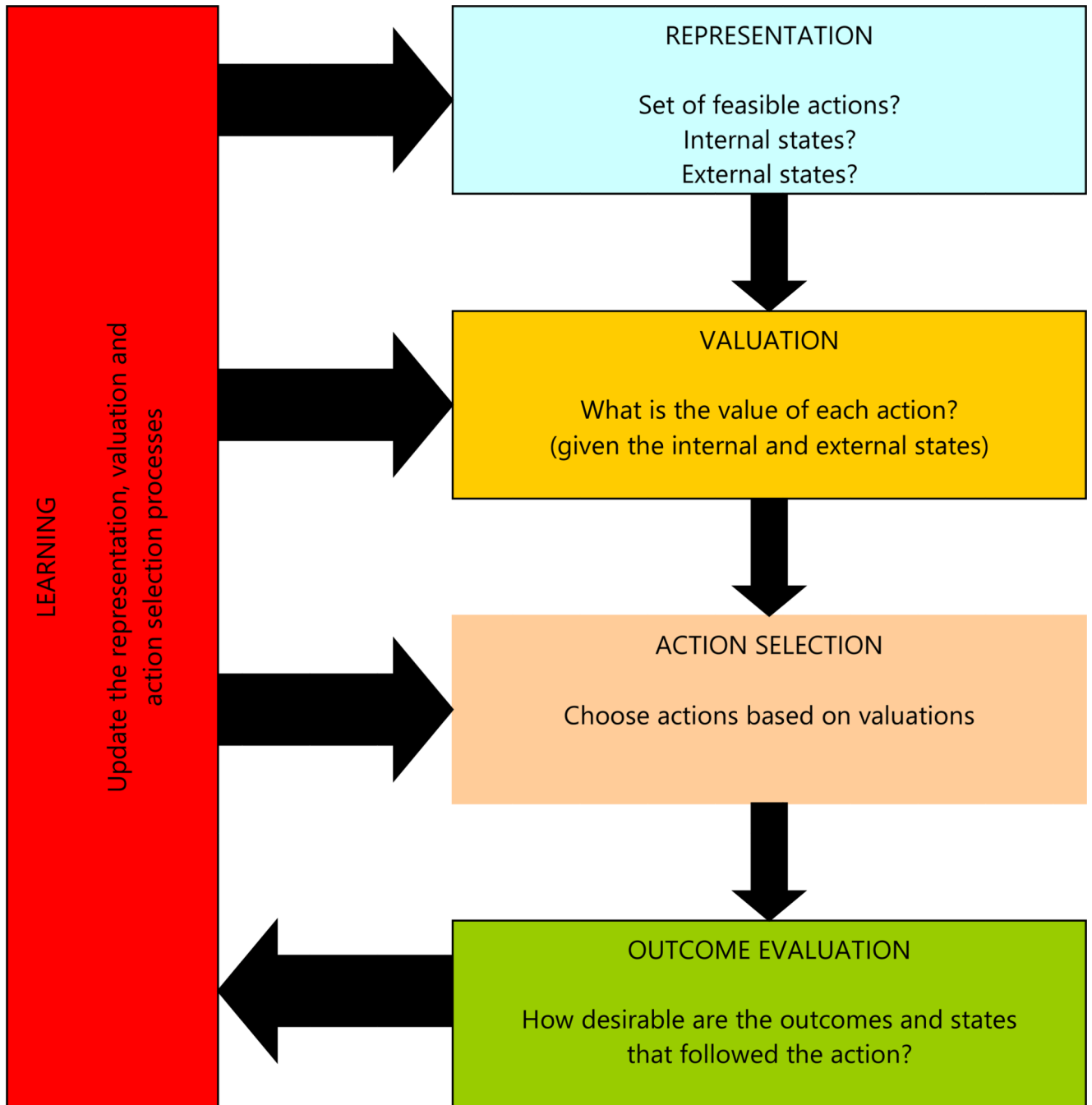
Figure 1. Basic computations involved in making a choice
Value-based decision-making can be broken down into five basic processes: first, the construction of a representation of the decision problem, which entails identifying internal and external states as well as potential courses of action; second, the valuation of the different actions under consideration; third, the selection of one of the actions based on their valuations; fourth, after implementing the decision the brain needs to measure the desirability of the outcomes that follow; finally, the outcome evaluation is used to update the other processes in order to improve the quality of future decisions.

	Pavlovian	Habitual	Goal directed
Pavlovian	Ex: hungry animal presented with food and electric shock simultaneously appetitive Pavlovian system: high value for food, low value for escape behaviors avoidance Pavlovian system: high value for escape behaviors, low value for food	Ex: animal trained to receive a reward for running away from food appetitive Pavlovian system: high value for running towards food avoidance habitual system: high value for running away from food	Ex: individual in a diet considering eating a tasty but high caloric food appetitive Pavlovian system: high value for the food health goal-directed system: low value for the food
Habitual		Ex: animal trained to run towards lever in response to a sound, and away from lever in response to a light, when presented with both stimuli approach habitual system: high value to lever approach avoidance habitual system: high value to lever avoidance	Ex: 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			Ex: 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 may lead to poor decision-making. This figure provides examples of conflict among different valuation systems and of conflict among different value signals of the same type.