Imaging Valuation Models in Human Choice

P. Read Montague,[1,2] Brooks King-Casas,[1] and Jonathan D. Cohen[3]

[1] Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030

[2] Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, Texas 77030

[3] Department of Psychology, Center for the Study of Brain, Mind, and Behavior, Green Hall, Princeton University, Princeton, New Jersey 08544

Key Words

reward, reinforcement learning, dopamine, ventral striatum, fictive learning signal

Abstract

To make a decision, a system must assign value to each of its available choices. In the human brain, one approach to studying valuation has used rewarding stimuli to map out brain responses by varying the dimension or importance of the rewards. However, theoretical models have taught us that value computations are complex, and so reward probes alone can give only partial information about neural responses related to valuation. In recent years, computationally principled models of value learning have been used in conjunction with noninvasive neuroimaging to tease out neural valuation responses related to reward-learning and decision-making. We restrict our review to the role of these models in a new generation of experiments that seeks to build on a now-large body of diverse reward-related brain responses. We show that the models and the measurements based on them point the way forward in two important directions: the valuation of time and the valuation of fictive experience.

Contents

INTRODUCTION

Decision-making can be difficult. Choosing the best decision in the face of incomplete information is a notoriously hard problem, and the quest to find neural and psychological mechanisms that guide human choice remains in its early stages. Despite decades of research on how decisions are represented and evaluated—both in psychology and in behavioral neuroscience—many fundamental questions remain unanswered, ranging from the concrete to the abstract. For example, How does the human brain represent basic tasks like finding food, mates, and shelter? How are potentially competing desires represented? Which aspects of the current state of affairs do we consider when making a decision? How does the brain choose a motor plan best suited for a particular task? Such questions, and many related ones, have generated large areas of specialized research. These are difficult questions because they require good guesses about the exact problems that the human brain is trying to solve, as well as good guesses about the mathematical formalisms and computational algorithms that best capture the solutions the brain implements.

In the simplest terms, human decision-making can be framed as an energetic problem that pits an organism's investment for each choice against the immediate and long-term returns expected. This trade-off is a fundamental one and has long acted as a selective constraint shaping the evolution of biological decision-making mechanisms. Consequently, we should expect the mechanisms that estimate the value of decisions to be crafty and efficient. Two supporting players lie behind every decision: representation and valuation. To make efficient choices, the brain must represent the available choices and calculate the differential value of each, including both near-term and distal future rewards.

One particularly promising approach to these issues is the use of disciplined theoretical work to guide carefully designed imaging experiments. Accordingly, we have chosen to structure this review around a family of computational models from the field of machine learning that has been used to design experiments, interpret results, and predict neural responses that underlie valuation during decision-making. In this synthesis, we (*a*) provide a selective review of neuroimaging work that identifies a consistent set of neural responses to reward delivery in which rewards range from juice squirts to preferences for cultural artifacts; (*b*) describe models of the relationship between value and reward;

(*c*) illustrate the biological plausibility of these models through a review of work that links neural responses to specific parametric features of these models; and (*d*) extend the models to address computations that value time and fictive experience.

NEURAL RESPONSES TO REWARD

Human decision-making provides a natural behavioral domain in which to probe the neural substrates of valuation; however, the lack of good neural probes in humans forced most of the early work on human choice into the theoretical domain (von Neumann & Morgenstern 1944, Bush & Mosteller 1955, Simon 1955, Luce & Raiffa 1957). This early theoretical work fell on the shared boundary of mathematical psychology, economics, and what is now called behavioral economics (Camerer 2003). In contrast, early neural work on valuation focused on its close cousin: reward. The difference between the two is simple but critical. Reward refers to the immediate advantage accrued from the outcome of a decision (e.g., food, sex, or water). In contrast, the value of a choice is an estimate about how much reward (or punishment) will result from a decision, both now and into the future. Thus, value incorporates both immediate and long-term rewards expected from the decision. So reward is more like immediate feedback, whereas value is more like a judgment about what to expect.

Behaviorally, reward can be easily measured and quantified. To an experimental psychologist or behavioral neuroscientist, a reward is simply a positive reinforcer, some external event that makes a target behavior more likely in the future. Neurally, early work on reward processing identified brain regions in mammals (mainly rodents) that, when stimulated, appeared to be a neural analogue of external rewards to a behaving animal (Olds & Milner 1954, Olds 1958, Olds 1962, Phillips & Olds 1969, Brauth & Olds 1977). In recent years, this seminal work has been cast

more explicitly in the language of decision-making (Herrnstein & Prelec 1991, Shizgal 1997, Gallistel et al. 2001, Gallistel 2005).

Until recently, however, work on reward-processing and decision-making did not make direct contact with neural mechanisms of valuation in humans—that is, the computation of value by humans (see Shizgal 1997 for review). Three developments have recently changed this situation. The first was detailed electrophysiological work on reward-processing in behaving monkeys during tasks involving learning (Ljungberg et al. 1992, Quartz et al. 1992, Schultz et al. 1993, Montague et al. 1996, Schultz et al. 1997, Hollerman & Schultz 1998, Schultz 2000, Schultz & Dickinson 2000, Waelti et al. 2001, Bayer & Glimcher 2005, Tobler et al. 2005) and decision-making (Platt & Glimcher 1999, Gold & Shadlen 2001, Shadlen & Newsome 2001, Glimcher 2002, Gold & Shadlen 2002, Huk & Shadlen 2003, Glimcher 2003, McCoy et al. 2003, Dorris & Glimcher 2004, Sugrue et al. 2004, Rorie & Newsome 2005, Sugrue et al. 2005). This electrophysiology work helped lay the foundation for more recent work that has exploited the second major development: the advent of modern human neuroimaging techniques that can be used to measure the physical correlates of neural activity during learning and decision-making (Posner et al. 1988; Ogawa et al. 1990a,b, 1992, 1993; Belliveau et al. 1990, 1991). The third development was the importation of formal algorithms for reward-learning from computer science and engineering; in particular, the field of machine learning (Sutton & Barto, 1998; Dayan & Abbott 2001). These models provide a theoretical framework for interpreting the monkey electrophysiological studies, and more recently have begun to guide reward expectancy experiments in human subjects using noninvasive neuroimaging techniques. A new generation of reward-learning and decision-making experiments now profits from a growing connection to these formal models of valuation and learning. In this review, we focus on this interplay

between modeling work and neuroimaging studies in humans. For context, we begin by briefly reviewing the rapidly expanding set of imaging results on reward-processing in humans.

A large and ever growing number of neuroimaging studies have examined brain responses to rewarding stimuli. This work began by using primarily appetitive stimuli as reward probes (for a review see O'Doherty 2004) but has progressed to more abstract rewards such as money (Breiter et al. 2001, Elliot et al. 2003, Knutson et al. 2003), cultural rewards such as art and branded goods (Erk et al. 2002, Kawabata & Zeki 2004, McClure et al. 2004, O'Doherty et al. 2006), and even social rewards such as love and trust (Bartels & Zeki 2004, King-Casas et al. 2005, Delgado et al. 2005). Despite the diversity of these rewards, experiments have consistently identified a common set of neural structures that activate to these stimuli, including the orbitofrontal cortex (OFC), ventral striatum, and ventromedial prefrontal cortex (VMPFC). These structures comprise a ventral valuation network (VVN) consistently activated across an array of rewarding dimensions. For example, the orbitofrontal cortex (OFC) has been implicated in hedonic experience across all sensory modalities (Rolls 2000), including gustatory (Zald et al. 2002, Kringelbach et al. 2003), olfactory (O'Doherty et al. 2000, Rolls et al. 2003a), auditory (Blood & Zatorre 2001, Blood et al. 1999), somatosensory (Francis et al. 1999, Rolls et al. 2003b), and visual stimuli (Lane et al. 1999, Royet et al. 2000). Similarly, activation in regions of the striatum and OFC has been observed in response to a wide range of rewards. At the same time, these constituents of the VVN are sensitive to different aspects of rewarding stimuli: Areas in the striatum and OFC are particularly responsive to rewards that change, accumulate, or are learned over time (Koepp et al. 1998, Delgado et al. 2000, Berns et al. 2001, Elliott et al. 2003, Knutson et al. 2003, Galvan et al. 2005, Sugrue et al. 2005), whereas activity in VMPFC scales with reward value (Knutson et al. 2001, O'Doherty et al. 2003).

Recently, imaging studies of reward responses have extended beyond appetitive stimuli to include cultural objects and social stimuli. Two recent studies investigating neural responses to aesthetic stimuli, such as art and bodily beauty (Aharon et al. 2001, Kawabata & Zeki 2004), have implicated OFC, whereas two other studies have also associated preference judgments to art with activity in prefrontal cortex (Cela-Conde et al. 2004) and striatum (Vartanian & Goel 2004). The receipt of cultural objects such as art cannot be considered a primary reward, yet these objects nonetheless elicit activity in the same neural structures also activated by receipt of primary reinforcers. Likewise, delivery of brand information of preferred consumer goods also activates reward-processing regions within VMPFC, OFC, and striatum (Erk et al. 2002, Paulus & Frank 2003, McClure et al. 2004). The study of beverage preference by McClure and colleagues demonstrated that behavioral preference, when dissociated from brand preference, scaled linearly with the VMPFC response even when the primary reward (sugared, caffeinated beverage) was kept exactly the same (McClure et al. 2004). Finally, recent work has also found responses in the striatum and VMPFC to the receipt of rewarding social stimuli. For example, images of romantic partners elicit responses in the striatum, particularly during the early stages of a romance (Bartels & Zeki 2004, Aron et al. 2005). Moreover, responses in VMPFC and striatal structures also scale with the humor of jokes (Goel & Dolan 2001, Mobbs et al. 2003, Mobbs et al. 2005).

REINFORCEMENT-LEARNING MODELS

Paralleling the progress on imaging and neurophysiological studies of reward processing, computational accounts of human decision-making capable of connecting with behavioral

and brain responses have matured. Although different in detail, all reinforcement-learning models share some basic assumptions. The first assumption is that organisms (learning agents) possess goals. In reinforcement-learning models, agents learn to achieve goals under the guidance of reinforcement signals. This process is made explicit by representing the learning problem as a space of states through which the agent moves either using actual motor actions or using internal changes in state (like mental rehearsal). With each movement, or state transition, the agent receives a reinforcement signal (a signed scalar value) that combines two sources of information: (*a*) information about immediate rewards (or punishments), and (*b*) information about the long-term opportunities for reward (or cost) associated with the state change. Consequently, reinforcement signals combine short-term feedback from immediate rewards with longer-term judgments about likely future rewards to yield an assessment (a number) that ranks the relative worth of the agent's state. So these signals are equipped to act reasonably like advice, combining immediate feedback with the best guess about how good the future is likely to be based on past experience. Theoretical work on reinforcement learning is vast and extends far beyond the scope of this review (for accessible reviews, see Sutton & Barto 1998, Kaelbling et al. 1996, Dayan & Abbott 2001). Daw (2003) gives an excellent overview of almost all the neurobiologically relevant uses of reinforcement learning.

Two fundamental components of any reinforcement-learning model are the representation of the problem faced by the agent as a state space (the representation piece) and a value associated with each state in the space (the valuation piece). These are illustrated in **Figure 1**. The value of a state represents the reward that can be expected from that state averaged over all time points from now into the distant future. These values are silent, stored numbers. To probe these silent values, experiments must extract them indirectly through
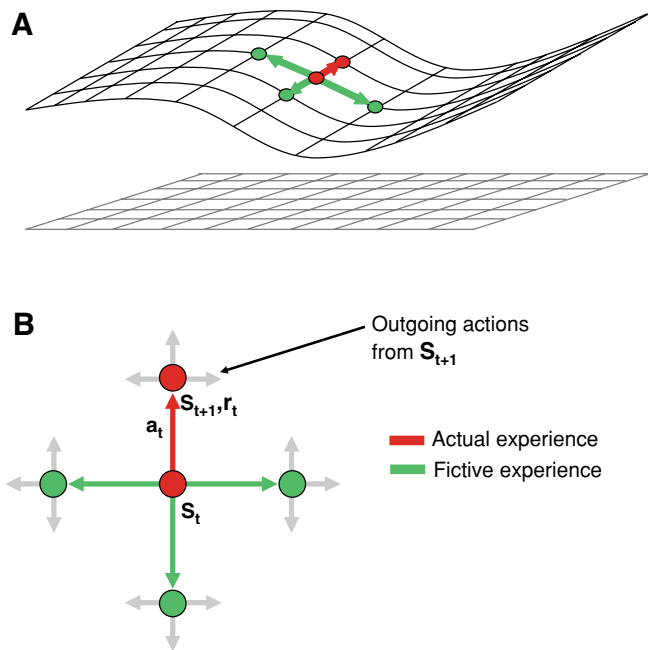

Figure 1

Representation of states, values, and actions. *A*: States are represented by vertices of the flat underlying grid. The current estimated value of each state is represented by the corresponding height above the plane. These values must be estimated from experience and stored in the brain. *B*: Actions (a_t) are chosen, and this results in the agent observation of a new state (S_{t+1}) and a reward (r_t). Learning is more efficient if both experiential learning signals (generated by red actions) and fictive learning signals (generated by information about green actions) influence changes in the estimated values of each state. Recent neuroimaging experiments have uncovered neural correlates of counterfactuals, an observation consistent with the existence of fictive error signals (Coricelli et al. 2005).

observation of the agent's actions in combination with some knowledge of the underlying state space. An important feature of this approach is that it does not assume any specific structure for the underlying state space. There is no natural way to read out values directly unless we know ahead of time that the agent will always act to choose states with the highest value.

Another equally fundamental element of reinforcement-learning models is the mechanism that guides the update of the values of states on the basis of experience. This update is guided by reinforcement signals. Reinforcement signals are constructed by combining information about immediate reward and

changes in value that occur owing to changes in the agent's state. In the popular temporal differences (TD) error formulation, the guidance signal, commonly called "the critic" (Barto et al. 1983; Sutton & Barto 1998), is written as

reward prediction error (TD error)
$$= r(S_t) + \gamma V(S_{t+1}) - V(S_t), \qquad 1.$$

where S_t is the current state and S_{t+1} is the next state to which the animal has transitioned after taking action a_t. The parameter t can be thought of as time, or any variable that provides a way to order the states visited. V is the value function that reflects the long-term value of each state (discounted by the parameter $0 < \gamma < 1$), and r is the immediately experienced reward associated with a given state (Dayan 1992, 1993, 1994a,b; Daw 2003).

Reinforcement signals like the TD error can be used to criticize (i.e., evaluate) the change in state and can also be used to direct the selection of new actions (i.e., following a behavioral policy). More specifically, a behavioral policy can be described as a function $P_{a,s}$ that maps states s to actions a and depends on the system having some estimate of the value function V. Suppose that these estimates of the value function are available and stored in the nervous system as "weights" $W_{a,s}$, which represent the value of taking action a given that the system is in state s. A common stochastic policy used in TD learning is to let $P_{a,s}$ be a logit or "softmax" function that prescribes the probability of taking action a in state s,

$$P_{a,s} = \frac{e^{\mu W_{a,s}}}{\sum_k e^{\mu W_{k,s}}}. \qquad 2.$$

We are all familiar with evaluation functions but may not realize it. Recall the chess-playing program Deep Blue, which beat world chess champion Gary Kasparov in a match in 1997 (Hsu 2002). The heart of Deep Blue was its evaluation function, its analog to the value functions that we have been describing. In this case, the evaluation function evaluated different moves (actions) given the current layout of the chess board (the current state of the chess board). For each possible chess move, the evaluation function produced a different number ranking the relative value of that move. Now chess is an extremely complex game, so the evaluation function could not look through every possible move. Instead, it had tricks built in to let it estimate primarily those chess moves (actions) with the highest value given the current layout of the chess board (the state). Of course Deep Blue had another trick as well: the capability of looking up ~200,000,000 different chess moves per second for any given board layout. And although Kasparov's brain could not explicitly look up chess moves at that breakneck rate, he was still able to keep the match close (3.5–2.5), so his brain must have discovered a more efficient representation of the problem allowing him to find quickly the high-value actions (chess moves). That's exactly the spirit of reinforcement learning's use of state space (board layout), action (chess move), and policy (the rule that picks the next chess move given the board layout) descriptions here. And it is also important to keep in mind that we are describing the simplest form of a reinforcement-learning system. More complex reinforcement-learning models are now being applied to problems of behavioral control in uncertain environments—models that should be able to guide imaging experiments in human subjects (Daw et al. 2005; see below).

The TD algorithm has been used effectively to address a variety of neurobiological and psychological phenomena, from the firing pattern of dopamine neurons in the ventral tegmental area in response to cues that predict rewards (Quartz et al. 1992, Montague et al. 1993, Montague & Sejnowski 1994, Friston et al. 1994, Houk et al. 1995, Montague et al. 1996, Schultz et al. 1997; for review see Montague et al. 2004) to the patterns of neural activity observed for striatal structures in human neuroimaging experiments (e.g., O'Doherty et al. 2003; McClure et al. 2003; also see King-Casas et al. 2005). However, one

problem with the simplest TD model is that it can be brittle. Its success depends heavily on having an appropriately rich representation of the state space, and a behavioral policy (a mechanism for selecting actions) that is properly matched to the value function over that space (see Bertsekas & Tsitsiklis 1996, Dayan & Abbott 2001). A learning system, including the human brain, must know about the problem for the TD algorithm to be a useful learning method.

An important variant of TD learning, called Q-learning, addresses some of these problems and provides a natural framework for biological experiments where actions and states are closely linked (Watkins 1989, Watkins & Dayan 1992). First, it learns a value function (called the Q function) not just over states, but over state-action pairs. Second, it is less dependent on the exact behavioral policy than is the TD algorithm sketched above. It is more forgiving of exact choices made as long as a decent policy is followed on average. As long as the states are visited with adequate frequency (i.e., the system has sufficiently broad experience), Q-learning can converge to optimal or near-optimal solutions (Watkins & Dayan 1992; see Daw 2003). Q-learning was proposed by Watkins in the late 1980s and unified separate approaches to reinforcement learning. The interested reader should consult some of the original sources or reviews, especially the now-classic book by Sutton & Barto (1998). The Q-learning algorithm, with an estimate $\hat{Q}$ of the Q function, may not initially be very accurate (i.e., has little initial information about the comparative value of the various action-state pairs). However, with experience, the algorithm refines this value function through repetition of the three basic steps outlined below.

1. Select action a on the basis of an initial value function Q and current state s.
2. Observe reward r and new state s'.
3. Update estimate $\hat{Q}$ of Q and set $s = s'$.

As we mentioned, Q-learning will learn the optimal value function for a range of policies.

However, the best policy is the one that picks the action with the largest value, just like a chess move with the highest rank in Deep Blue's evaluation function. That is, "in state s, take the action a that yields the largest Q":

$$policy(s) = \max_a \hat{Q}(s, a); \qquad 3.$$

"max over a" means pick the a that gives the biggest estimate $\hat{Q}$ of Q. So in Q-learning, the model depends on the underlying states only indirectly.

The algorithm used to update the estimate $\hat{Q}$ of the optimal value function Q follows the same basic prescription as the TD algorithm (see Watkins 1989, Dayan 1992). The optimal Q function over state-action pairs satisfies a recursive equation (Bellman equation)

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \max_{\tilde{a}} Q(s_{t+1}, \tilde{a}), \quad 4.$$

where γ is a time discount parameter, and the max (maximum) function as before picks the action with the greatest estimated value from all the actions $\tilde{a}$ available in state s_{t+1}. This expression holds strictly only for the optimal Q. The agent's problem is to learn this optimal Q. The Q-learning algorithm does this by assuming that the agent maintains some estimate $\hat{Q}$ of the optimal Q and then updates this estimate iteratively using the reinforcement signal r that it receives for a given action a_t taken from a given state s_t in a manner very similar to the TD learning algorithm:

$$\Delta \hat{Q}(s_t, a_t) = \lambda_t \lfloor r(s_t, a_t) \\ + \gamma \max_{\tilde{a}} \hat{Q}(s_{t+1}, \tilde{a}) - \hat{Q}(s_t, a_t) \rfloor, \quad 5.$$

where λ_t is the learning rate. Brain responses that correlate with this kind of experienced reward prediction error signal have been found in the striatum during conditioning experiments in humans and are reviewed shortly. Furthermore, such signals shift from one part of the striatum to another depending on whether a task is passive (no action is required to receive reward) or active (i.e., one or more actions are required to receive reward).

Despite the greater robustness to behavioral strategies, Q-learning also suffers from

a limitation shared with TD learning: it is influenced only by rewards and actions that are actually experienced. As a consequence, an animal using this algorithm would learn only from its own actions and experienced rewards. It is well known, however, that adaptive organisms ranging from bees to humans use counterfactual (fictive) information. That is, they learn from what might have been (Roese & Summerville 2005) as well as what has come to pass.

AN ABUNDANCE OF CRITICS: THE NEED FOR FICTIVE LEARNING SIGNALS

When foraging for food, animals explicitly and implicitly share information about what might have been. For example, consider explicit information sharing. Honeybees live in hives and forage for nectar and pollen as sources of raw material for the hive (Oster & Wilson 1978). Once a forager returns to the hive, she shares information about her experience with those that gather around her. In so doing, she is communicating what could have been to other foragers or would-be foragers had they sampled the flower patches that she has already experienced. How should the recipients of this information respond? First, we must note that the recipients of this information must already possess the capacity to respond to counterfactual information appropriately. Given that this capacity is intact, the recipient bees should update their estimates of these unvisited patches (provided that they consider the communicator to be telling the truth). In a hive, bees have a strong genetic incentive to communicate the truth to one another. This explicit information sharing allows the entire hive to learn at a much higher rate than any individual forager could. It does not matter exactly how this sharing is implemented in each bee's brain; the important issue is the tendency of the hive members to treat another individual's experience as though it was their own. The same kind of information sharing can also take place inci-

dentally during foraging of groups of animals that are not genetically related. One animal sees another animal finding grub worms under specific kinds of rocks and updates its own estimate of the value of those rocks as predictors of rewards. In this case, the recipient of the information must have an adequate model of itself and the other animal to understand how to use the information to update its internal value function on the basis of what it might have had.

These examples highlight the fact that adaptive learners, including humans, can benefit by sensing what might have been, comparing it with what actually was the case, and using this comparison to influence learning and subsequent decision-making. Q-learning can similarly exploit signals related to actions not taken if the following simple modification is made:

$$\sum_{a \in \{\text{all actions}\}} \Delta \hat{Q}(s_t, a) = \Delta \hat{Q}(s_t, a_t) + F(\text{actions not taken from state } s_t). \quad 6.$$

Here, the update in the estimates of the Q value function depends not just on differences between expectations and experience (first term on right; red action-states in **Figure 1**) but also on some function F of fictive actions not taken (second term on right; green action-states in **Figure 1**). Lohrenz and colleagues have called this extended form "counterfactual Q-learning" and hypothesized two distinct fictive learning signals (errors) contributing linearly to the function F (T. Lohrenz, personal communication):

$$\sum_{a \in \{\text{all actions}\}} \Delta \hat{Q}(s_t, a) = \Delta \hat{Q}(s_t, a_t) + \sum_{\substack{\text{fictive} \\ \text{actions } \tilde{a}_t}} \Delta \hat{Q}(s_t, \tilde{a}_t) + \Delta \hat{Q}(s_t, \text{best } \tilde{a}_t)$$

$$= \text{experience critic} + \text{fictive critic} + \text{supervisor critic}. \quad 7.$$

The first error signal is the experiential error, that is, the ongoing temporal difference between expected rewards and experienced

rewards. This term is the direct analog to the TD error described above. The second term is a fictive critic signal. It is an error signal that speeds learning by using any available information about other actions that might have been taken (e.g., through observation of others' actions in a comparable state), along with estimates of the associated rewards. The second term expresses the fact that the learner should use information about other possible actions and outcomes because it is a cheaper way to learn about the world; a lot of information return on a small investment (e.g. just a brief glance). At first glance, the fictive error signal might be thought of as reflecting regret and/or relief. Such emotions are associated with a constellation of other internal states and behavioral responses that extend well beyond the capacities built into the fictive error signals we have characterized here. Consequently, we propose that the fictive error signals outlined here probably form only one part of the complex emotion of regret/relief; however, our review should illustrate that even emotions could in principle be captured by computational descriptions. The third term represents an ongoing comparison between the best action that could have been taken in terms of reward outcomes and the one actually taken. This term is equivalent to the supervised actor-critic architecture proposed by Rosenstein & Barto (2004), which uses an internal supervisor for reinforcement. This architecture was proposed to remedy problems that simple reinforcement-learning models have with complicated decision problems that may require a task to be supervised from different levels. The third term of Equation 7 provides a natural basis for such a supervisor critic by informing the system after the fact, whether an action was as good as it could have been when compared with the best action.

In summary, we have reviewed reinforcement-learning models central to the design and interpretation of reward expectancy experiments, in particular, the TD model. We note that the TD model is too simple to account for the range of reward learning that a real organism carries out, and we have reviewed a more sophisticated family of reinforcement-learning models called Q-learning. Both frameworks possess the implicit limitation that they learn only via the actual experience of the learner, and so we outlined a generalization of Q-learning to counterfactual Q-learning. This generalization predicts the existence of two different fictive learning signals that we refer to as the fictive critic and supervisor critic, respectively. The former provides a natural way to speed learning, and the latter can act as an internal supervisor for reinforcement learning. These models emerged originally from work in machine learning literature, where they have been successfully used to design increasingly sophisticated autonomous agents that can learn about and adapt to their environment. Their use in neuroscience, however, has been to provide a formal framework within which to describe, in semiquantitative terms, the valuation processes that drive learning and decision-making in natural organisms. Recently, this framework has been put to use in predicting patterns of neural activity that should be associated with valuation, decision-making, and learning processes in the human brain if they are accurately described by the reinforcement-learning mechanisms outlined above. In the following sections, we review empirical studies that have begun to test these predictions.

IMAGING VALUATION AND LEARNING: PREDICTION ERROR SIGNALS

Most applications of reinforcement-learning models to brain function make the assumption that the neuromodulator dopamine implements the reinforcement-learning signal (Dayan 1994, Montague et al. 1996, Schultz et al. 1997, Dayan & Abbott 2001, Montague et al. 2004; also see Montague et al. 1995). Dopamine has long been associated with the idea of reward but historically was assumed

to mediate directly the reward value of an event. However, reinforcement-learning models have helped differentiate the concept of reward from a reward-dependent learning signal, and accumulating empirical evidence—both from neurophysiological studies as well as more recent neuroimaging studies—strongly support the idea that dopamine release is associated with the learning signal rather than (or at least in addition to) the reward itself. There is now overwhelming evidence, both from direct measurements of dopaminergic spike trains and from fast electrochemical measurements of dopamine transients, that phasic changes in dopamine delivery carry a reward prediction error signal (for reviews see Schultz & Dickinson 2000, Montague et al. 2004; also see Lavin et al. 2005).

Evidence also shows that phasic changes in dopamine delivery carry information about novelty to target structures; however, this is not surprising. Both novelty and prediction errors are needed by an adaptive learning system that must forage in an uncertain environment for food and mates, a fact highlighted by theoretical efforts to deal directly with the computational meaning of novelty responses (Dayan et al. 2000, Kakade & Dayan 2002, Yu & Dayan 2005). The most compelling of these findings demonstrate that, although phasic dopamine release is observed in response to the delivery of an unpredictable reward, dopamine release is greater in response to a cue that reliably predicts subsequent reward than in response to the reward itself. In general, dopamine signals are observed to be strongest in response to unpredictable events that reliably signal reward delivery and can even be suppressed when a reward is expected but not delivered. This pattern of responses conforms precisely to the behavior of the TD error, described formally by Equation 1.

The reward prediction error model of phasic dopamine delivery has motivated a growing number of functional magnetic resonance imaging (fMRI) studies of reward expectancy in humans (Elliot et al. 2000;

Knutson et al. 2000, 2001; Berns et al. 2001; Breiter et al. 2001; Pagnoni et al. 2002; Seymour et al. 2004; McClure et al. 2003; O'Doherty et al. 2003, 2006). In neural structures receiving strong dopaminergic inputs, these experiments consistently reveal responses possessing timing and polarity consistent with a reward prediction error signal. (We label these semiquantitative predictions because timing and signal polarity do not represent a fully rigorous dynamical model.) These regions include dorsal and ventral striatum, nucleus accumbens, and portions of medial frontal, orbitofrontal, and cingulate cortex. Most of these results have used standard contrastive analyses (using a general linear model) to identify regions that exhibit statistically significant hemodynamic responses under conditions predicted by a reward prediction error (Friston et al. 1995a,b). However, more detailed and direct assessments of the time course of hemodynamic measurements within individual trials also show a signal that faithfully tracks the predicted time course of negative and positive reward prediction error signals (McClure et al. 2003).

One exciting finding has been the discovery that the location of activation in the striatum (dorsal versus ventral) in response to reward prediction errors depends on whether a task requires an action for reward to be received (as in instrumental conditioning) or whether no action is required for reward (as in classical conditioning). For action-contingent rewards, reward prediction error signals are detected in dorsal and ventral striatum (O'Doherty et al. 2003), whereas for passive reward presentation, prediction errors signals are detected only in ventral striatum (McClure et al. 2003, O'Doherty et al. 2003). Another domain in which reinforcement-learning models make interesting, testable predictions is sequential decision-making under risk (Montague et al. 1996, Egelman et al. 1998, Elliot et al. 2000, Montague & Berns 2002). Here, reinforcement-learning models can be used to make quantitative, trial-by-trial predictions about the biases that

decision-makers will have for particular choices, using their past experience of the rewards associated with each choice. For example, O'Doherty et al. (2006) used a TD model to fit the behavior of each individual subject, and this best-fit TD model was then used as a linear regressor in the analysis of the fMRI-measured hemodynamic changes. As in earlier studies, this revealed activity in classic dopamine-projection areas, including the ventral striatum. These studies are significant for methodological as well as substantive reasons: They represent one of the few applications of a formally explicit, computational model of a psychological function to guide the design and analysis of a neuroimaging experiment. Furthermore, the findings provide direct evidence for the neural implementation of the hypothesized, psychological function that converges closely with findings from direct neuronal recordings in nonhuman species. These findings were made possible because of the specificity of the predictions, which in turn was possible only because the psychological function was specified in a mechanistically explicit, computational form.

SOCIAL ECONOMICS: VALUATION DURING HUMAN EXCHANGE

In humans, one of the most important classes of value computations is evoked by social exchanges with other humans, especially social exchanges with shared economic outcomes. For the past two decades, three types of economic games (and their variants), contrived to emulate human exchange (trade), have been used by psychologists and behavioral economists to probe social computations underlying fairness (ultimatum game; Camerer 2003, Güth et al. 1982), trust (trust game; Camerer & Weigelt 1988, Berg et al. 1995), and cooperation (Prisoner's Dilemma; Axelrod 1984, Rapoport & Chammah 1965). These three games probe fundamental psychological mechanisms including those that (*a*) detect and respond to fairness (Güth et al.

1982, Forsythe et al. 1994), (*b*) punish unfairness (Fehr & Gächter 2002), and (*c*) build and respond to models of the partners participating in the exchange (Kreps et al. 1982, Camerer & Weigelt 1988).

These games are representative of a broader class of behavioral economic probes used to test social interactions ranging from competitiveness to the influence of groups and time on human decision-making (Camerer 2003, Camerer et al. 2003). These experimental approaches to human exchange take their theoretical foundation and mathematical structures from economics and game theory, and in so doing explicitly represent social rewards and social valuations (Kagel & Roth 1995, Camerer 2003). Just as reinforcement-learning theory brings from machine learning a formal framework for studying the influence of valuation on learning and decision-making, so game theory brings a class of formal models to the study of psychological processes involved in social exchange and their neural underpinnings. Initial forays in this direction have begun to produce promising results.

IMAGING SOCIAL EXCHANGE: UNFAIRNESS PUNISHED AND COOPERATION REWARDED

One of the simplest probes of fairness is the one-round ultimatum game (Güth et al. 1982), which might appropriately be renamed "take it or leave it." The game is played between two players. The pair is given some endowment, say $100. The first player proposes a split of this money to the second player, who can respond by either accepting the proposal (take it) or rejecting it (leave it). If the proposal is rejected, neither player receives any money, thereby hurting both players' outcomes, and this hurt is shared, although not necessarily equally. Findings from this game reveal a disparity between the rational agent model of human exchange and the way that humans actually behave. According to the rational agent model, proposers should offer as little as possible, and responders should

accept whatever they are offered because something is better than nothing. Thus, an offer of $1, or even 1¢, should be accepted. However, this is not what happens when the experiment is run. Responders routinely reject offers less than about 20% of the endowment, even when this means foregoing considerable sums of money, in some cases as much as a month's pay (Henrich et al. 2001). Apparently, this rejection rate results from a prior representation of the likely rejection rate of the average player because proposers usually offer significant amounts, presumably to avoid unequal splits that they anticipate will be rejected (Fehr & Schmidt 1999).

These behaviors are observed even in circumstances in which partners interact only once and the interaction is confidential, suggesting that they are driven by strong, highly ingrained behavioral mechanisms and fairness instincts. That is, the findings suggest that humans possess fairness instincts that are exercised even when they incur considerable financial expense. It is important to note that for such fairness instincts to be adaptive more generally, participants must possess mechanisms that engender complementary behaviors on each side of the exchange. Proposers must possess a reasonably accurate prior model of what the responder is likely to reject, and the responders must be willing to reinforce such models by enforcing substantial rejection levels, thus insuring fair (higher) offers from proposers. The problem of how such interaction instincts evolved is a fascinating evolutionary problem, which has been explained in terms of the value of reputation to the individual (Nowak & Sigmund 2005), as well as the value of altruistic punishment in stabilizing social cohesion against environmental challenges (Fehr & Simon 2000, Fehr & Fischbacher 2003). Until recently, however, it has been difficult to adjudicate between hypotheses about the mechanisms driving behavior in such interactions. One approach to this challenge has been to use neuroimaging methods to identify neural structures engaged in such tasks and associated with specific be-

havioral outcomes (e.g., Sanfey et al. 2003, de Quervain et al. 2004).

In one study, Sanfey & colleagues (2003) used fMRI to monitor brain responses while subjects played one-round ultimatum games against human and computer partners. A separate condition was also presented to each subject to control for monetary reinforcement outside the context of the social exchange. Three interesting findings emerged from this study. First, consistent with earlier behavioral work on this game (Roth 1995, Camerer 2003), not only did responders consistently reject offers made by human partners that were less than 20% of the initial endowment but, interestingly, for any given offer level, rejection rates were higher for human partners than for computer partners. These observations are consistent with the view that the ultimatum game and similar tasks probe mechanisms designed for mediating exchange with other humans rather than merely probe economic mechanisms designed to harvest rewards efficiently from the world whether or not these rewards derive from humans (social exchange) or from some other source (e.g., food sources in a field). Second, the experiment identified neural responses that correlated with the degree of fairness of the offers, a neural response that was, again, larger for human partners than for computer partners. Third, these responses were most prominent in anterior insula and correlated with behavioral response. For any given level of unfairness, the response in the anterior insula to unfair human offers was greater than the response to unfair computer offers. Furthermore, for offers in the unfair range, the degree of insula activity correlated positively with the likelihood that the responder would reject the offer. From other work, we know that responses in this region of the brain correlate with negative emotional states that attend pain, hunger, thirst, anger, and physical disgust (Derbyshire et al. 1997, Denton et al. 1999, Tatarani et al. 1999, Calder et al. 2001). Thus, physical disgust, and what one might interpret as moral disgust in the

ultimatum game, seems to deploy the same brain systems. This suggests that these two types of valuations may share common computational components.

The findings from this study suggest that violations of fairness elicit what can be interpreted as an emotional response (disgust) that, at least under the laboratory conditions in which it was elicited (confidential, one-shot interactions), seems to violate the dictums of the rational agent model of economic behavior. However, as noted above, this behavior may reflect a rational computation, over an evolutionary time scale, to circumstances involving (and relying on) repeated social interactions. In recent years, there have been numerous demonstrations that humans are capable of enforcing social norms through explicit punishment at a cost to the punisher (Fehr & Gächter 2002, Boyd et al. 2003, Fehr & Rockenbach 2004).

From an evolutionary perspective, it may appear odd (at least on the surface) that humans will undertake costly punishment of a transgressor of social norms without any expectation of direct returns—a pattern of behavior often referred to as altruistic punishment (Axelrod 1986, Henrich & Boyd 2001). However, the work of Fehr and colleagues has produced a substantial set of findings that support this point (Fehr & Rockenbach 2004). They suggest that such behaviors provide evolutionary advantages to social groups by preserving social cohesion in times of environmental stress, when pressures mount for individuals to act selfishly in their own interest. Although it is beyond the scope of this review, it is worth noting that this group has described a formal model of how altruistic punishment can evolve on the basis of these principles—a model that might be exploited to develop a richer, quantitative understanding of the factors that engage such altruistic behaviors. At the same time, more precise knowledge of the mechanisms that drive such behavior may help inform the theory. The findings from the Sanfey et al. (2003) study suggest one set of mechanisms that responds

aversively to violations of fairness. However, such behavior may also employ endogenous brain mechanisms of positive reinforcement to enforce these and other behavioral algorithms (see King-Casas et al. 2005 and section below for direct evidence to this effect in a related game). Recent work by Fehr and his colleagues addresses this possibility.

Using PET imaging, de Quervain and colleagues (2004) tested the hypothesis that the anticipation and/or execution of punishment that enforced a social norm would be associated with activity in reward-processing brain structures. Participants in their experiment had the opportunity to punish other participants who chose not to reciprocate in an economic exchange game. They found that punishing a defector correlated with activation in regions of the striatum similar to those that have been observed in response to other sources of reward. Moreover, they also found that activation in the striatum, as well as medial prefrontal cortex, correlated with the anticipation of the satisfaction that attended the delivery of the punishment. This latter finding is consistent with other reward expectancy experiments, generalizing them to an inherently social signal: punishment of a transgressor of social norms.

Neuroimaging studies have demonstrated the engagement of brain reward systems by other forms of social interaction, including cooperation and trust. For example, Rilling & colleagues (2004) used fMRI to probe neural activity during performance of the iterated Prisoner's Dilemma, a game commonly used as a model for studying serial cooperation (in a context in which reputations can form). They observed that activity in reward-processing structures were associated with cooperative interactions, consistent with the hypothesis that these responses reflected the operation of the mechanism responsible for incentivizing cooperation and discouraging nonreciprocation. Robust activations during cooperation were found in nucleus accumbens (ventral striatum), caudate nucleus (dorsal striatum), ventromedial prefrontal cortex,

and rostral anterior cingulate cortex. All these regions receive strong dopaminergic projections (known to carry reward, saliency, and reward prediction error signals). Thus, like altruistic punishment, cooperative social behavior may also engage central reward-processing mechanisms that can be measured using neuroimaging methods.

The findings described above provide evidence that neural mechanisms involved in valuation are engaged by social as well as economic exchange, which suggests that the brain relies on common valuation mechanisms to guide decision-making in diverse domains. This also suggests an important way in which neuroimaging methods—as a means of directly querying underlying mechanisms of valuation—may be particularly valuable as a complement to behavioral methods (i.e., preferences revealed strictly by observing behavioral choice) for studying valuation in the social domain. Whereas the valuation of an isolated good available in the physical environment may, and optimally should, map directly onto a corresponding response, this may not be so when the good appears in an environment populated by other agents whose actions can anticipate, respond to, and influence one's own. Indeed, such displays can sometimes be devastating, exposing the indiscriminant advertiser to exploitation (e.g., at the most primitive level, indicating to a predator the most effective direction of pursuit). Consequently, a strong incentive exists to keep such valuation functions (or parts of them) private, an incentive that is likely to have exerted considerable evolutionary pressure. In particular, this may have driven behaviors associated with social exchanges in competitive environments to remain as opaque to competitors as possible with respect to underlying valuation mechanisms, while remaining cooperative enough to garner profitable exchange with potential partners. In other words, the mechanisms driving behavior in such circumstances have likely evolved to keep private exactly how valuable one perceives particular elements of a social

exchange to be. Under such conditions, neuroimaging methods may provide a valuable means of peering behind the behavioral curtain and more directly observe the function of proximal mechanisms of valuation.

IMAGING SOCIAL LEARNING: REPUTATION, RECIPROCITY, AND TRUST

The accumulating evidence that similar neural mechanisms are engaged in the valuation of social as well as other types of rewards has produced another potentially valuable dividend: the promise that quantitative models used to understand reward-learning processes can also be used to understand important aspects of social valuation processes. One example of progress being made in this direction involves the use of another economic exchange game that probes fairness, cooperation, and reputation building: the trust game. This game was proposed initially in a simple form by Camerer & Weigelt (1988) and can be seen as a modified ultimatum game. The singular feature of the exchange was labeled "trust" by Berg et al. (1995), a group who put the game into its modern form used worldwide today. Like the ultimatum game and Prisoner's Dilemma, the trust game involves an exchange between two players in which cooperation and defection can be parametrically encoded as the amount of money sent to one's partner. On each exchange, one player (the investor) is endowed with an amount of money. The investor can keep all the money or decide to invest some amount, which is tripled and sent to the other player (the trustee) who then decides what fraction to send back to the investor.

Single round versions of this game have been played across a wide range of cultural groups and the modal results are consistent: The investor almost never keeps all the money, but instead makes substantial offers to the trustee that could be considered close to fair splits (Camerer 2003). This move is typically met with a reasonable (close to fair) return from the trustee. The initial act of

trust on the part of the investor entails some likelihood of a loss (risk), and can be viewed as a cooperater signal, reflecting a (modally correct) assumption that the trustee will respond in kind. This game can be played personally (nonanonymously) or anonymously, and the results are basically the same. The risk involved in a unilateral trusting offer is apparently mitigated by other response mechanisms in the average human's brain. Just like observed behavior in the ultimatum game, the rational agent model in its simple form would not have predicted this experimental result. Berg et al. (1995) recognized this game as an embodiment of the problem of trust and the risk that trusting another human entails; most modern forms of the game are modifications of their formulation. One early neuroimaging study used a single-round version of this game to assess neural responses during such an exchange (McCabe et al. 2001) and found differences between playing a computer and a human.

A recent large-scale fMRI study using a modified version of this trust game imaged neural responses simultaneously from both partners interacting in a multiround version of the task (King-Casas et al. 2005). The multiround version (in this case, 10 rounds) allowed reputations (models) to form between investor and trustee, and simultaneous measurements from the two brains (Montague et al. 2002) allowed the investigators to observe correlations in neural activity across the pair of interacting brains, as these models improved across rounds. Two notable findings emerged in this study.

First, in a region rich with dopaminergic input (ventral head of the caudate nucleus), neural signals correlated directly with deviations in tit-for-tat reciprocity. This neural finding was significant because reciprocity was the major behavioral signal that explained changes in the willingness to increase or decrease trust (money sent to one's partner) on the next move.

The second finding of interest also emerged from the caudate nucleus in exactly

the region identified as responding strongly to large changes in reciprocity. Here, in the trustee's brain, the time series of neural responses were examined near the moment when the investor's response was revealed to both players' brains. In particular, King-Casas and colleagues (2005) separated neural responses in this region of the trustee's brain according to whether the trustee was going to increase or decrease money sent (level of trust) on their next move, yielding a simple but operational neural correlate of the trustee's "intention to change their level of trust" on that next move. Recall that in this game, trust was stripped of its normally rich social meaning and operationalized as the amount of money sent to one's partner. As shown in **Figure 2**, the most remarkable finding was that the neural correlate of the intention to increase trust could be modeled directly as a reinforcement-learning signal (a reward prediction error signal). This signal possessed all the features seen in simple conditioned experiments like those summarized in **Figure 2A**.

Notice in **Figure 2** that the intention-to-increase-trust response began (in early rounds) by reacting to the revelation of the investor's decision and, through learning, underwent a temporal transfer to an earlier point in time as the reputation of the investor built up in the brain of the trustee. The timing, contingencies, and change through learning exhibited by this signal possess all the important features of a reward prediction error signal similar to that measured in simpler conditioning experiments (e.g., like those shown in **Figure 3**). One important point from these results is that even abstractions like the intention to trust can engage reward-predicting and machinery of the midbrain and striatum, and thereby act themselves as primary rewards. This suggests one way that ideas in general may gain the behavioral power to guide goal-directed behavior; to the rest of the brain, they acquire the properties of an important control signal, a primary reward (food, water, sex).
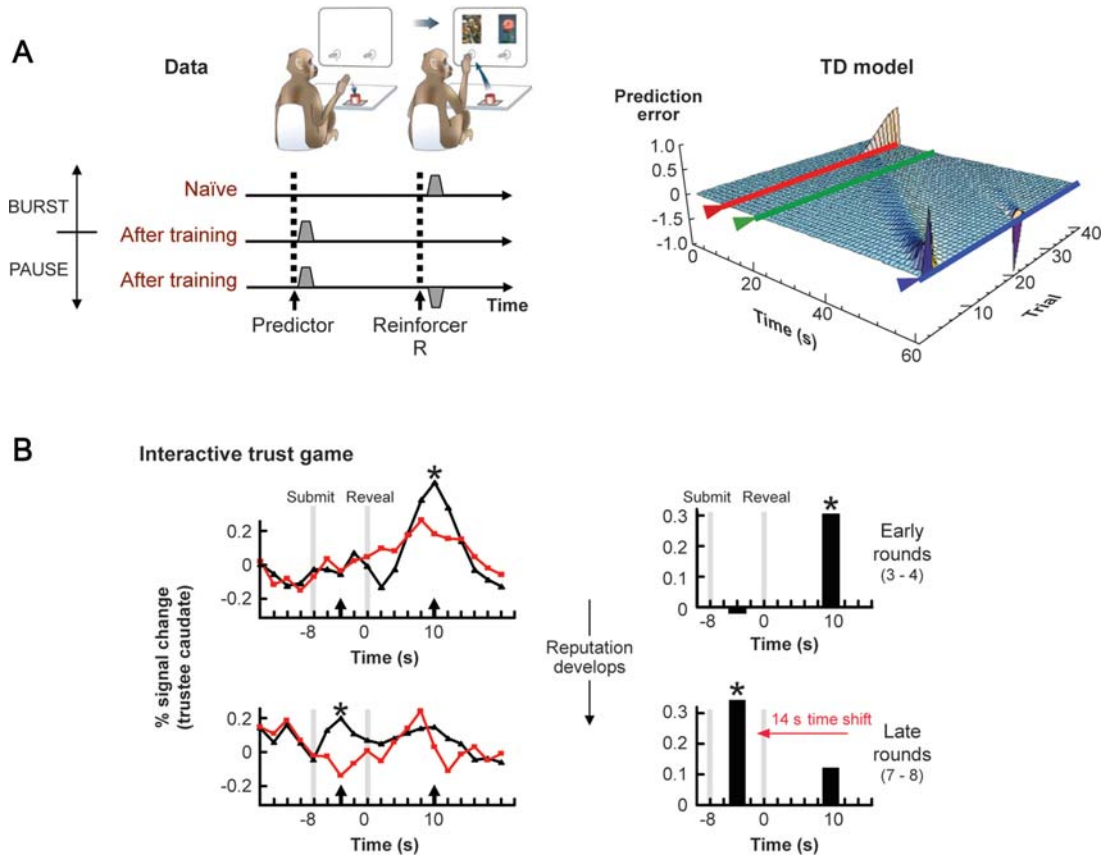
A: Data — Naïve / After training / After training; BURST / PAUSE; Predictor, Reinforcer R; Time. TD model — Prediction error; Time (s); Trial.

B: Interactive trust game — % signal change (trustee caudate); Submit, Reveal; Time (s). Reputation develops. Early rounds (3–4); Late rounds (7–8); 14 s time shift.

Figure 2

Reinforcement-learning signals during social exchange. *A*: (*Left*) Summary of burst and pause responses in monkey dopamine neurons during reward-learning tasks. The pattern is the same across a range of experiments. Surprising receipt of reward causes burst responses in a naïve animal's dopamine neurons. The presence of a temporally consistent predictor of reward induces two changes: (*a*) response to formerly surprising reward disappears, and (*b*) response grows to the earliest predictor. Notice the pause response if the reward is omitted at the expected time. (*Right*) This same behavior is seen in a reinforcement-learning model, plotted here as a TD error signal. Two cues (*red and green*) consistently predict reward delivery (*blue*). After training, the burst response to reward (*blue*) disappears and shifts to the earliest consistent predictor of reward (*red*) *B*: (*Left*) Average hemodynamic response measured in the caudate nucleus of trustee brain in a 10-round trust game. Response time series are shown near the time that investor decisions are revealed and are divided according to whether the trustee increases trust (money sent) on their next move (*black*) or decreases its trust on his/her next move (*red*). The difference between these curves is significant initially following revelation of investor's decision (reactive) and, after reputations form, shifting to time before the investor's decision is even revealed. (*Right*) Summary of hemodynamic responses in left panel at the times indicated by the arrows (adapted from Schultz et al. 1997, Hollerman & Schultz 1998, King-Casas et al. 2005).

Other explanations for these data are possible. One possibility is that the strial signal reflected the monetary rewards associated with trust. There was a weak, positive correlation between increases in trust by the

trustee and changes in the next investment by the investor. However, in the context of this game, trust may simply reduce exactly to a proxy for such an expectation. In fact, it may be that trust always reduces in this way,
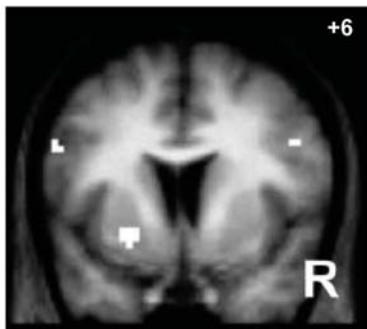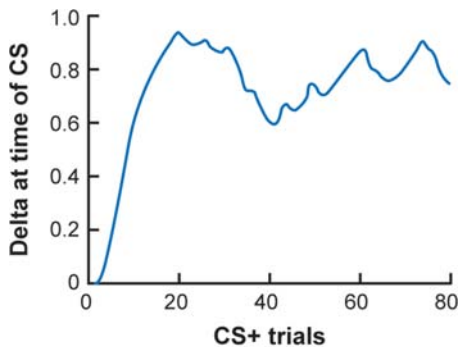
Figure 3

Imaging computational processes using a reinforcement-learning model regressor. (*Top*) Temporal difference error regressors computed for all trials in a reward-dependent classical conditioning experiment. This shows the predicted TD-error signal expected at the time of the conditioned stimulus and plotted as a function of trial number during which reward was delivered. (*Bottom*) Statistically significant brain activation for the regressor computed in the top panel (p < 0.001) (adapted from O'Doherty et al. 2003).

although the internal currency in different kinds of exchanges may become substantially more complex (Montague & Berns 2002). Nevertheless, the larger point remains: Humans can frame on this economic exchange, build models of what to expect from their partner, develop models of their partner's likely behavior, and in so doing generate physical signals in dopamine-rich brain regions that mimic the properties of a reward prediction error.

INTERTEMPORAL CHOICE, EXPLORATION, AND EXPLOITATION

Virtually all valuation and reinforcement-learning mechanisms discount the value of a good with time. That is, a reward available sooner is considered to be more valuable than one that is available later. This captures the intuitive sense that sooner delivery is more certain but also concurs with more abstract considerations (for example, a sum of money delivered now can be invested and generate more returns than a sum of money delivered later). In the models we have discussed, this idea is expressed as the γ parameter, which implements an exponential form of discounting. This form of discounting concurs with standard economic models. There, the use of exponential discounting is motivated by the fact that it ensures consistency of intertemporal choice—that is, in the relative preference for goods available at different points in time—which is consistent with (and in fact, a requirement of) the rational agent model (Koopmans 1960, Ainslie 1975, Frederick et al. 2002). For example, if one prefers $1000 today over $1100 in a week, this indicates that one is discounting the value of the offer by at least $100 for the delay of a week. Therefore, one should equally prefer $1000 to be delivered in a year over $1100 to be delivered in a year and a week. However, people often do not conform to this prediction, expressing a preference for the $1100 option in the latter case. This and an overwhelming number of other behavioral findings concerning intertemporal choice suggest that humans (and other animals) exhibit much steeper devaluation over the near term than is predicted by exponential discounting. Such behavior is observed almost universally and has been used to account for a wide variety of pathologies in human decision-making, from susceptibility to marketing, to drug abuse, obesity, and the failure of Americans to save adequately for retirement (Angeletos et al. 2001, Gruber & Botond 2001, Laibson 1997,

O'Donoghue & Rabin 1997, Loewenstein & Thaler 1989).

One way to explain this pervasive phenomenon is to assume that the discount function is hyperbolic rather than exponential (Herrnstein 1997, Rachlin 2000). This predicts much steeper discounting in the near term than in the long term, which in turn can explain preference reversals (see **Figure 2A**). However, this explanation begs several important questions. Why should exponential discounting, as it is expressed in reinforcement-learning models, account adequately for the variety of valuation and learning behaviors we have reviewed here? In some cases, models based on exponential discounting provide quantitatively accurate descriptions of both behavior and neural function. A second more fundamental question is, How does one justify hyperbolic discounting—where it is observed—in terms of the rational agent model favored by standard economic theory? One answer to these questions is to assume that hyperbolic discounting reflects the operation of more than a single valuation mechanism. The simplest version of this view suggests that there are two canonical mechanisms: one that has a very steep discount function (in the limit, valuing only immediate rewards), and one that treats rewards more judiciously over time with a shallower discount function (**Figure 2B**) (e.g., Shefrin & Thaler 1988, Loewenstein 1996, Laibson 1997, Metcalfe & Mischel 1999).

Recently, neuroimaging evidence has been brought to bear on the debate over a single discounting mechanism versus multiple discounting mechanisms. In one study, McClure et al. (2004) used fMRI to measure neural activity in participants as they made intertemporal choices similar to the one in the example above (although with considerably smaller sums at stake). When they compared choices involving the option for an immediate reward (e.g., a $10 **www.amazon.com** gift certificate today or one worth $11 in two weeks) with choices involving only delayed rewards (e.g., a $10 gift certificate in two weeks vs one worth $11 in four weeks), they found neural activity in many of the same regions rich in the dopamine projections we have discussed here, including ventral striatum and medial prefrontal cortex. In contrast, other regions, including dorsolateral prefrontal and posterior parietal cortex, were activated by all decisions. These are areas commonly associated with more deliberative cognitive processes (e.g., calculation, problem-solving, and reasoning) (Duncan 1986, Stuss & Benson 1986, Shallice & Burgess 1991, Dehaene et al. 1998, Koechlin et al. 1999, Miller & Cohen 2001). Furthermore, for choices that engaged both sets of mechanisims (i.e., involving an option for immediate reward), the relative strength of activity in the two systems predicted behavior, with greater activity in the prefrontal-parietal system predicting choices for the later-but-greater reward. These findings suggest that the brain houses at least two distinguishable valuative mechanisms, one of which exhibits properties consistent with a steep discount function and another that is more sensitive to the value of future rewards.

Other recent findings have generalized these results to primary forms of reward such as the delivery of small quantities of fluid to thirsty participants (McClure et al. 2005a). Such studies have not yet fully characterized the form of the discount function used by each mechanism. However, one reasonable hypothesis is that the system that exhibits a steep discount function, and involves areas rich in dopamine projections, reflects the operation of the same valuation mechanisms that we have discussed in the context of reinforcement-learning models. If this is confirmed, it would provide a natural point of contact between the neuroscience literature on reinforcement learning and economic treatments of time discounting. The marriage of these two rich theoretical traditions, coupled with methods to test quantitative predictions about the proximal mechanisms underlying valuation, make this a promising area of emerging research at the intersection of

neuroscience, psychology, and economics. See **Figure 4**.

Exploitation versus Exploration

Up to this point, our consideration of valuation and learning has assumed that the intrinsic value of a reward is stable, meaning that its value is the same whenever it is encountered. However, this is of course not true in the real world. One may care more about food when he/she is hungry than when he/she has just eaten. Similarly, the environment may change, which will require a change in one's valuation function. For example, although one may be hungry and wish to look for food, he/she may know that all the available food has been eaten and therefore should divert one's behavior toward some other goal (i.e., place greater value on some other reward). These observations reveal a fundamental challenge for real-world valuation functions and goal-directed behavior: the trade-off between exploitation and exploration.

Exploitation refers to behavior that seeks to maximize a particular form of reward (i.e., achieve a particular goal). Reinforcement-learning algorithms promote exploitation in the sense that they favor and progressively strengthen those behaviors that generate the most reward.

However, environments, and even an organism's own needs, change over time, often as a product of the organism's own behavior. Exploiting an environment for a resource may deplete the environment of that resource, or satisfy the organisms need for it, eventually diminishing its reward value. When either the availability of reward or its value changes in this way, then reinforcement-learning algorithms run into trouble. That is, they are designed to optimize behavior in a stationary environment—one that does not change, or changes very little, over time. To address this fact, an adaptive system must be able to change its behavior to explore new environments and identify new sources of reward. Insofar as changes in the environment, or the organism's internal state, increase in likelihood with time, the tension between exploitation and exploration aligns with different time scales of adaptation: Exploitation is about optimizing performance in temporally local stationarities, whereas exploration is about adapting to longer-term time scales, over which change is likely.

Although the tension between exploitation and exploration has been recognized in both the animal learning literature (e.g., Krebs et al. 1978; Krebs and Kacelnik 1984) as well as work in machine learning (Kaelbling et al. 1996), relatively little work has addressed the neural mechanisms that regulate the balance between exploitation and exploration. Recently, however, both neurophysiological and formal modeling work has begun to suggest that, like reinforcement learning, neuromodulatory systems play a central role in this function. In particular, work by Aston-Jones and colleagues (Aston-Jones et al. 1994, Usher et al. 1999, Aston-Jones & Cohen 2005) has suggested that the brainstem nucleus locus coeruleus (LC), which is responsible for most of the norepinephrine released in the neocortex, may be critically involved in regulating the balance between exploitation and exploration. This work suggests that the LC functions in two modes: one in which phasic LC responses facilitate context-congruent behavioral responses (exploitation), and another in which such phasic responses are absent but an increase in tonic LC firing facilitates the execution of a broader class of behavioral responses (exploration).

A recent theoretical model proposed by Yu & Dayan (2005) is consistent with this hypothesis, which suggests that tonic norepinephrine release may code for unexpected forms of uncertainty, favoring a shift in the behavioral set. Most recently, these ideas have been extended by McClure et al. (2005) to propose a model in which cortical valuation mechanisms (in orbitofrontal and anterior cingulate cortex) provide input to the LC, which in turn adaptively regulates its mode of function to optimize the operation
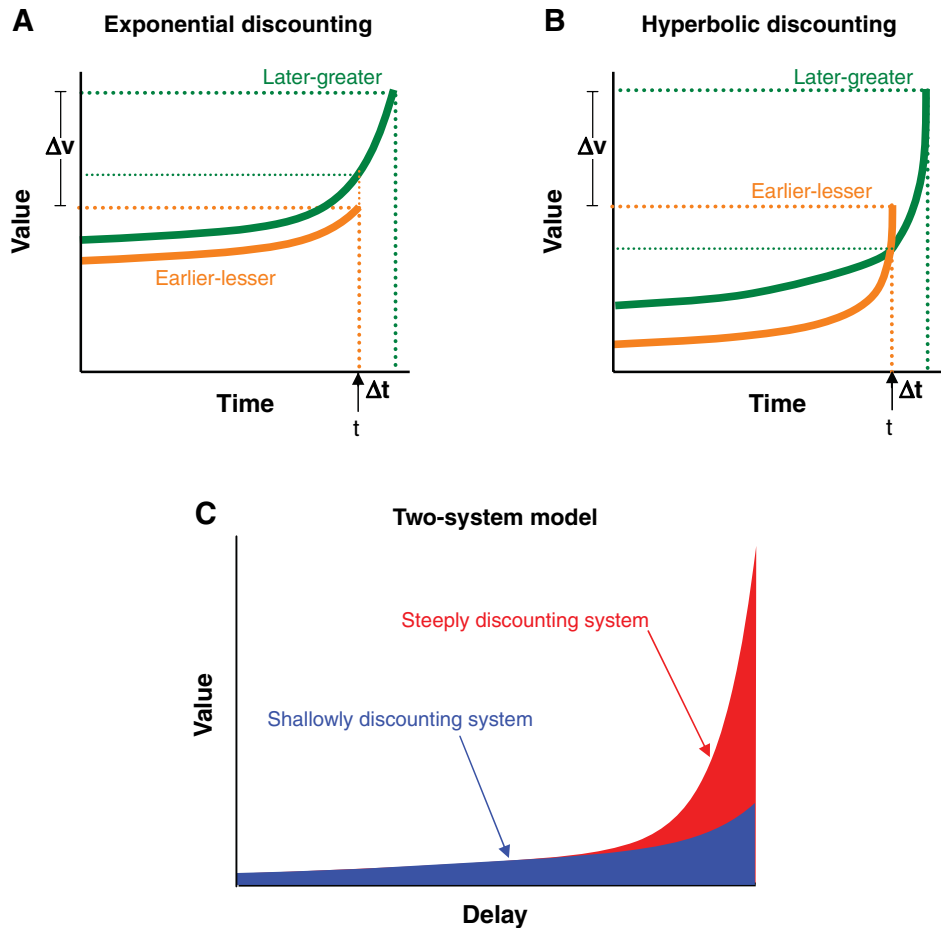
A Exponential discounting

Later-greater

Earlier-lesser

Value

Δv

Δt

t

Time

B Hyperbolic discounting

Later-greater

Earlier-lesser

Value

Δv

Δt

t

Time

C Two-system model

Steeply discounting system

Shallowly discounting system

Value

Delay

Figure 4

Comparison of different discounting functions. *A*: The discounted value of two goods available in the
future plotted as a function of time from availability. One is intrinsically more valuable but will be
available at a later time (later-greater) than the other (earlier-lesser). The heavy dotted lines show the
discounted value of each good at the time that the earlier ones becomes available, whereas the lighter
dotted line shows the value of the later good at that same time. Note that, at all times, the value of the
later-greater good (even after discounting for its later availability) is always more than the earlier-lesser
good, including at the time of delivery of the earlier good. *B*: Value of the same two goods in Panel *A*
(with the same differences in value ⟨1⟩Δv⟨/1⟩ and same temporal offsets in availability ⟨1⟩Δt⟨/1⟩) but now
computed using a hyperbolic rather than an exponential discount function. Note that for most time prior
to when the earlier good is available, the value of the later good remains greater than that of the earlier
good, as it does for exponential discounting. However, at the time the earlier good becomes available, its
value exceeds that of the later good. This is due to the steep near-term discounting of the hyperbolic
discount function and explains preference reversals seen commonly in human behavior (see text).
C: A discount function composed from two exponential functions, one with a large time constant (*red*,
"steeply discounting function") and another with a smaller time constant (*blue*, "shallowly discounting
function"). This function has properties similar to the hyperbolic functions illustrated in Panel *B*.

of a dopamine-based reinforcement-learning mechanism.

CONTEXT-DEPENDENT MODULATION OF VALUATION SIGNALS AND COMPETITION IN DECISION-MAKING

Our review of neuroimaging studies has focused on a specific set of mechanisms involved in valuation and decision-making: those that seem to be well characterized by formal models of reinforcement learning. In part, this is because this area of research represents an exciting and promising convergence of formal modeling and empirical research involving both neuroimaging as well as neurophysiological methods. However, the review would be seriously incomplete if we did not consider, at least briefly, how the mechanisms involved in reinforcement learning are influenced by and interact with other mechanisms responsible for valuation and decision-making in the brain. Indeed, in many of the studies described above, the valuation signals observed in classic reward-processing areas, such as the striatum and paralimbic cortex (e.g., anterior insula and medial prefrontal cortex), have been observed to be modulated by context. That is, they are not isolated determinants of behavior.

For example, in a recent trust game, Delgado and colleagues (2005) found that the response to partner feedback in caudate nucleus was modulated by a form of social priming. The investigators systematically manipulated the social identity of one's partner by telling tales about the partner. In some games the partner was "praiseworthy," in some the partner was "neutral," and in others the partner was "morally suspect." The presumed learning signal evident in caudate was observed in the neutral condition (consistent with King-Casas et al. 2005) but was lacking in both the positive and negative priming conditions. This result suggests that learning parameters are susceptible to modulation by explicit (and presumably also implicit) biases, which have been detailed in the social cognitive neuroscience literature (e.g., Lieberman 2005).

Such computations are likely to be sensitive to other noncognitive variables as well. For example, in a recent report, Kosfeld and colleagues (2005) actively manipulated subjects' oxytocin levels, a neuropeptide thought to be critical in prosocial approach behavior in human and nonhuman mammals (Insel & Young 2001, Uvnas-Moberg 1998). Specifically, Kosfeld et al. examined the effects of intranasal administration of oxytocin on behavior expressed within a trust game similar to the ones used by King-Casas et al. (2005) and Delgado et al. (2005). As expected, willingness to trust was enhanced with oxytocin administration. The specific nature of these modulatory influences, and their impact on reinforcement-learning mechanisms, remains to be explored. However, the collection of results emerging from the social neuroscience literature promises to provide valuable constraints on the development of formal theory in this domain.

Finally, a full account of human valuation and decision-making behavior must go well beyond the basic mechanisms of reinforcement learning upon which we have focused here. Many of the studies discussed above provide evidence that overt behavior is influenced by interactions between different valuative and decision-making mechanisms. For example, the studies of intertemporal choice revealed that the outcome of decisions between an immediate-but-lesser versus later-but-greater reward was correlated with the relative activity of mechanisms closely associated with reinforcement learning (such as the ventral striatum and medial prefrontal cortex) versus higher-level cortical mechanisms commonly associated with more deliberative forms of decision-making (such as the dorsolateral prefrontal and posterior parietal cortex). Similarly, the decision to accept or reject an unfair offer in the ultimatum game was correlated with the relative degree of activity in insular cortex versus dorsolateral prefrontal cortex. These findings suggest that, in

these circumstances, the outcome of behavior was determined by a competition between different valuative mechanisms that produced differing appraisals of the same stimuli and therefore favored different behaviors. Similar findings have been observed in other domains of decision-making, such as moral reasoning (Greene et al. 2001, Greene & Cohen 2004).

One theme that runs through this work, and more broadly in fields such as cognitive psychology and behavioral economics, is that human decision making behavior reflects the convergent influences of multiple subsystems (e.g., Posner & Snyder 1975, Kahneman 2003). Some of these systems support higher-level cognitive processes capable of complex, deliberative forms of evaluation, whereas others involve more primitive or specialized and automatic mechanisms such as reinforcement learning. Critically, these are likely to use different value functions. Thus, although they may work synergistically to govern behavior under many (or even most) circumstances, under others they may produce different appraisals and therefore favor different behavioral dispositions.

Competition between different subsystems reflects a fundamental constraint on valuation and decision-making mechanisms in the brain: There is only one body, and therefore only a limited number of actions can be executed at any time. This poses a primary challenge for a valuative mechanism, which is to prioritize the value of competing actions in any given setting. The constraint of a single body may pose an even greater challenge for the brain as a whole, requiring it to prioritize different valuative and decision-making mechanisms, when these mechanisms produce different appraisals of the same circumstance and thereby favor different actions. Recent modeling work has sought to formalize these ideas and give structure to hypotheses concerning interactions between different types of valuation and decision-making systems (e.g., Daw et al. 2005). Such work is critical to achieving a more rigorous understanding of the complexity of such interac-

tions. However, even now, insight provided by work in this area has begun to reshape thinking about human behavior in a broad range of disciplines, from economics to moral philosophy, and may even have consequences for broader issues regarding social policy and the law (e.g., Greene & Cohen 2004, Cohen 2005).

SUMMARY

The neuroimaging findings regarding valuation, decision-making, and learning reviewed here can be summarized with the three following general observations: (*a*) There is striking consistency in the set of neural structures that respond to rewards across a broad range of domains, from primary ones such as food, to more abstract ones such as money and social rewards (including reciprocity, fairness, and cooperation). (*b*) Similarly, reinforcement-learning signals within these domains (evoked by errors in reward prediction) generate responses in the same neural structures as those engaged in simple conditioning tasks. (*c*) The dynamics of these signals, both within and across learning trials, conform to predictions made by formal models of reinforcement learning. These findings reveal a remarkable degree of conservation in both the structure and function of reinforcement-learning mechanisms across different domains of function in the human brain. At the same time, the function of these mechanisms is clearly modulated by other systems, and such interactions are an important, active, and exciting area of current exploration. Finally, and most important, reinforcement-learning mechanisms represent only one valuation and decision-making system within the brain. A deeper, more precise understanding of how other valuation and decision-making mechanisms are implemented promises to generate important new insights into who we are and why we feel and act the way we do. Such work will impact a wide range of disciplines concerned with human behavior, from clinical research on drug addiction, to fields, such as economics and the

law, that have traditionally been far removed from neuroscience. We have illustrated that current applications of neuroimaging methods, coupled with the development and use of formally rigorous theories to guide the design and interpretation of neuroimaging experiments, have already begun to provide a solid foundation for work in this area. We look forward to the rapid progress that is likely to ensue in the coming years.

LITERATURE CITED

Ainslie G. 1975. Specious reward: a behavioral theory of impulsiveness and impulse control. *Psychol. Bull.* 82:463–96

Angeletos GM, Laibson D, Repetto A, Tobacman J, Weinberg S. 2001. The hyperbolic consumption model: calibration, simulation, and empirical evaluation. *J. Econ. Perspect.* 15:47–68

Aston-Jones G, Cohen JD. 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28:403–50

Aston-Jones G, Rajkowski J, Kubiak P, Alexinsky T. 1994. Locus coeruleus neurons in monkey are selectively activated by attended cues in a vigilance task. *J. Neurosci.* 14(7):4467–80

Axelrod RM. 1984. *The Evolution of Cooperation*. New York: Basic Books

Aharon I, Etcoff N, Ariely D, Chabris CF, O'Connor E, Breiter HC. 2001. Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* 32:537–51

Aron A, Fisher H, Mashek DJ, Strong G, Li H, Brown LL. 2005. Reward, motivation, and emotion systems associated with early-stage intense romantic love. *J. Neurophysiol.* 94:327–37

Axelrod R. 1986. An evolutionary approach to norms. *Am. Pol. Sci. Rev.* 80:1095–11

Axelrod R, Hamilton WD. 1981. The evolution of cooperation. *Science* 211:1390–96

Bartels A, Zeki S. 2000. The neural basis of romantic love. *Neuroreport* 11(17):3829–34

Bartels A, Zeki S. 2004. The neural correlates of maternal and romantic love. *Neuroimage* 21:1155–66

Barto AG, Sutton RS, Anderson CW. 1983. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans. Sys. Man Cybernet.* 13:834–46

Bayer HM, Glimcher PW. 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47:129–41

Belliveau J, Kennedy D, McKinstry R, Buchbinder B, Weisskoff R, et al. 1991. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* 254:716–19

Belliveau J, Rosen B, Kantor H, Rzedzian R, Kennedy D, et al. 1990. Functional cerebral imaging by susceptibility-contrast NMR. *Magn. Reson. Med.* 14:538–46

Berg J, Dickhaut J, McCabe K. 1995. Trust, reciprocity, and social history. *Games Econ. Behav.* 10:122–42

Berns GS, Chappelow J, Zink CF, Pagnoni G, Martin-Skurski ME, Richards J. 2005. Neuro-biological correlates of social conformity and independence during mental rotation. *Biol. Psychiatry* 58:245–53

Berns GS, Cohen JD, Mintun MA. 1997. Brain regions responsive to novelty in the absence of awareness. *Science* 76:1272–75

Berns GS, McClure SM, Pagnoni G, Montague PR. 2001. Predictability modulates human brain response to reward. *J. Neurosci.* 21:2793–98

Bertsekas DP, Tsitsiklis JN. 1996. *Neuro-Dynamic Programming*. Belmont, MA: Athena Sci.

Bischoff-Grethe A, Martin M, Mao H, Berns GS. 2001. The context of uncertainty modulates the subcortical response to predictability. *J. Cogn. Neurosci.* 13:986–93

Blood AJ, Zatorre RJ. 2001. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc. Natl. Acad. Sci. USA* 98:11818–23

Blood AJ, Zatorre RJ, Bermudez P, Evans AC. 1999. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat. Neurosci.* 2:382–7

Boyd R, Gintis H, Bowles S, Richerson PJ. 2003. The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. USA* 100:3531–35

Brauth SE, Olds J. 1977. Midbrain activity during classical conditioning using food and elec-trical brain stimulation reward. *Brain Res.* 134:73–82

Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P. 2001. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30:619–39

Bush RR, Mosteller F. 1955. *Stochastic Models for Learning*. New York, NY: Wiley

Calder AJ, Lawrence AD, Young AW. 2001. Neuropsychology of fear and loathing. *Nat. Rev. Neurosci.* 2:352–63

Camerer CF. 1999. Behavioral economics: Reunifying psychology and economics. *Proc. Natl. Acad. Sci. USA* 96:10575–77

Camerer CF. 2003. *Behavioral Game Theory*. Princeton, NJ: Princeton Univ. Press

Camerer CF, Loewenstein G, Rabin M. 2003. *Advances in Behavioral Economics*. Princeton, NJ: Princeton Univ. Press

Camerer CF, Fehr E. 2006. When does "economic man" dominate social behavior? *Science* 311:47–52

Camerer C, Weigelt K. 1988. Experimental tests of the sequential equilibrium reputation model. *Econometrica* 56:1–36

Camerer C, Thaler RH. 1995. Ultimatums, dictators and manners. *J. Econ. Perspect.* 9:209–19

Camerer C. 2003. Dictator, ultimatum, and trust games. In *Behavioral Game Theory: Experiments in Strategic Interaction*. New York: Russell-Sage

Cela-Conde CJ, Marty G, Maestu F, Ortiz T, Munar E, et al. 2004. Activation of the prefrontal cortex in the human visual aesthetic perception. *Proc. Natl. Acad. Sci. USA* 101:6321–25

Coricelli G, Critchley HD, Joffily M, O'Doherty JP, Sirigu A, Dolan RJ. 2005. Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. Neurosci.* 8(9):1255–62

Cohen JD. 2005. The vulcanization of the human brain: a neural perspective on interactions between cognition and emotion. *J. Econ. Perspect.* 19:3–24

Dalton KM, Nacewicz BM, Johnstone T, Schaefer HS, Gernsbacher MA, et al. 2005. Gaze fixation and the neural circuitry of face processing in autism. *Nat. Neurosci.* 8:519–26

Davidson RJ, Irwin W. 1999. The functional neuroanatomy of emotion and affective style. *Trends Cogn. Sci.* 3:11–21

Davidson RJ, Putnam KM, Larson CL. 2000. Dysfunction in the neural circuitry of emotion regulation—a possible prelude to violence. *Science* 289:591–94

Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8:1704–11

Daw N. 2003. *Reinforcement learning models of the dopamine system and their behavioral implications.* PhD thesis. Carnegie Mellon Univ. Dep. Comp. Sci., Pittsburgh, PA.

Dayan P. 1992. The convergence of TD(lambda) for general lambda. *Machine Learn.* 8:341–62

Dayan P. 1993. Improving generalisation for temporal difference learning: The successor representation. *Neural Comput.* 5:613–24

Dayan P, Sejnowski TJ. 1994. TD(lambda) converges with probability 1. *Mach. Learn.* 14:295–301

Dayan P. 1994. Computational modelling. *Curr. Opin. Neurobiol.* 4:212–17

Dayan P, Abbott LF. 2001. *Theoretical Neuroscience.* Cambridge, MA: MIT Press. 576 pp.

Dayan P, Kakade S, Montague PR. 2000. Learning and selective attention. *Nat. Neurosci.* 3:1218–23

Dayan P, Watkins CJCH. 2001. Reinforcement learning. *Encyclopedia of Cognitive Science.* London, UK: MacMillan

Dehaene S, Dehaene-Lambertz G, Cohen L. 1998. Abstract representation of numbers in the animal and human brain. *Trends Neurosci.* 21:355–61

Delgado MR, Frank RH, Phelps EA. 2005. Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat. Neurosci.* 8:1611–18

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. 2000. Tracking the hemodynamic responses to reward and punishment in the striatum. *J. Neurophysiol.* 84:3072–77

Denton D, Shade R, Zamarippa F, Egan G, Blair-West J, et al. 1999. Neuroimaging of genesis and satiation of thirst and an interoceptor-driven theory of origins of primary consciousness. *Proc. Natl. Acad. Sci. USA* 96:5304–9

de Quervain DJ, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, et al. 2004. The neural basis of altruistic punishment. *Science* 305:1254–58

Derbyshire SW, Jones AK, Gyulai F, Clark S, Townsend D, Firestone LL. 1997. Pain processing during three levels of noxious stimulation produces differential patterns of central activity. *Pain* 73:431–45

Dorris MC, Glimcher PW. 2004. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44:365–78

Duncan J. 1986. Disorganisation of behavior after frontal lobe damage. *Cogn. Neuropsychol.* 3:271–90

Egelman DM, Person C, Montague PR. 1998. A computational role for dopamine delivery in human decision-making. *J. Cogn. Neurosci.* 10:623–30

Elliott R, Friston KJ, Dolan RJ. 2000. Dissociable neural responses in human reward systems. *J. Neurosci.* 20:6159–65

Elliott R, Newman JL, Longe OA, Deakin JF. 2003. Differential response patterns in the striatum and orbitofrontal cortex to financial reward in humans: a parametric functional magnetic resonance imaging study. *J. Neurosci.* 23:303–7

Erk S, Spitzer M, Wunderlich AP, Galley L, Walter H. 2002. Cultural objects modulate reward circuitry. *Neuroreport* 13:2499–503

Fehr E, Fischbacher U. 2003. The nature of human altruism. *Nature* 425:785–91

Fehr E, Fischbacher U. 2004. Social norms and human cooperation. *Trends Cogn. Sci.* 8:185–90

Fehr E, Rockenbach B. 2004. Human altruism: economic, neural, and evolutionary perspectives. *Curr. Opin. Neurobiol.* 14:784–90

Fehr E, Gächter S. 2002. Altruistic punishment in humans. *Nature* 415:137–40

Fehr E, Schmidt KM. 1999. A theory of fairness, competition, and cooperation. *Q. J. Econ.* 71:397–404

Fehr E, Simon G. 2000. Fairness and retaliation: the economics of reciprocity. *J. Econ. Perspect.* 14(3):159–81

Frederick S, Loewenstein G, O'Donoghue T. 2002. Time discounting and time preference: a critical review. *J. Econ. Literat.* 40:351–401

Forsythe R, Horowitz J, Savin NE, Sefton M. 1994. Fairness in simple bargaining experiments. *Games Econ. Behav.* 6:347–69

Francis S, Rolls ET, Bowtell R, McGlone F, O'Doherty J, et al. 1999. The representation of pleasant touch in the brain and its relationship with taste and olfactory areas. *Neuroreport* 10:453–59

Friston KJ, Tononi G, Reeke GN, Sporns O, Edelman GM. 1994. Value-dependent selection in the brain: simulation in a synthetic neural model. *Neuroscience* 59:229–43

Friston KJ, Frith CD, Turner R, Frackowiak RSJ. 1995a. Characterising dynamic brain responses with fMRI: a multivariate approach. *NeuroImage* 2:166–72

Friston KJ, Holmes AP, Worsley K, Poline JB, Frith CD, Frackowiak RSJ. 1995b. Statistical parametric maps in functional brain imaging: a general linear approach. *Hum. Brain. Mapp.* 2:189–210

Fulton S, Woodside B, Shizgal P. 2000. Modulation of brain reward circuitry by leptin. *Science* 287:125–28

Gallistel CR. 2005. Deconstructing the law of effect. *Games Econ. Behav.* 52:410–23

Gallistel CR, Mark TA, King AP, Latham PE. 2001. The rat approximates an ideal detector of changes in rates of reward: implications for the law of effect. *J. Exp. Psych. Anim. Behav. Process.* 27:354–72

Galvan A, Hare TA, Davidson M, Spicer J, Glover G, Casey BJ. 2005. The role of ventral frontostriatal circuitry in reward-based learning in humans. *J. Neurosci.* 25:8650–56

Glimcher PW. 2002. Decisions, decisions, decisions: choosing a biological science of choice. *Neuron* 36:323–32

Glimcher PW. 2003. *Decisions, Uncertainty, and the Brain. The Science of Neuroeconomics.* Cambridge, MA: MIT Press

Glimcher PW. 2003. The neurobiology of visual-saccadic decision making. *Annu. Rev. Neurosci.* 26:133–79

Glimcher PW, Rustichini A. 2004. Neuroeconomics: the consilience of brain and decision. *Science* 306:447–52

Goel V, Dolan RJ. 2001. The functional anatomy of humor: segregating cognitive and affective components. *Nat. Neurosci.* 4:237–38

Gold JI, Shadlen MN. 2001. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* 5:10–16

Gold JI, Shadlen MN. 2002. Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* 36:299–308

Gottfried JA, O'Doherty J, Dolan RJ. 2002. Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J. Neurosci.* 22(24):10829–37

Greene JDC, Cohen JD. 2004. For the law, neuroscience changes nothing and everything. *Philos. Trans. R. Soc. London Ser. B* 359:1775–85

Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293:2105–8

Gruber J, Botond K. 2001. Is addiction 'rational?' Theory and evidence. *Q. J. Econ.* 116:1261–305

Güth W, Schmittberger R, Schwarze B. 1982. An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Org.* 3:367–88

Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, et al. 2001. Cooperation, reciprocity and punishment in fifteen small-scale societies. *Am. Econ. Rev.* 91:73–78

Henrich J, Boyd R. 2001. Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *J. Theor. Biol.* 208:79–89

Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, et al. 2004. *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies.* Oxford, UK: Oxford Univ. Press

Herrnstein RJ. 1997. *The Matching Law: Papers in Psychology and Economics*, ed. RJ Herrenstein, H Rachlin, DI Laibson, p. 334. Cambridge, MA: Harvard Univ. Press

Herrnstein RJ, Prelec D. 1991. Melioration: a theory of distributed choice. *J. Econ. Perspect.* 5:137–56

Hoffman E, McCabe K, Shachat K, Smith V. 1994. Preferences, property rights, and anonymity in bargaining games. *Games Econ. Behav.* 7:346–80

Hollerman JR, Schultz W. 1998. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* 1:304–9

Houk JC, Adams JL, Barto AG. 1995. A model of how the basal ganglia generate and use neural signals that predict reinforcement. In *Models of Information Processing in the Basal Ganglia*, ed. JC Houk, JL Davis, DG Beiser, pp. 249–70. Cambridge, MA: MIT Press

Hsu FH. 2002. *Behind Deep Blue: Building the Computer that Defeated the World Chess Champion.* Princeton, NJ: Princeton Univ. Press

Huk AC, Shadlen MN. 2003. The neurobiology of visual-saccadic decision making. *Annu. Rev. Neurosci.* 26:133–79

Insel TR, Young LJ. 2001. The neurobiology of attachment. *Nat. Rev. Neurosci.* 2:129–36

Kaelbling LP, Littman ML, Moore AW. 1996. Reinforcement learning: a survey. *J. Artif. Intellig. Res.* 4:237–85

Kagel JH, Roth AE. 1995. *The Handbook of Experimental Economics.* Princeton, NJ: Princeton Univ. Press

Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–91

Kahneman DA. 2003. Perspective on judgment and choice: mapping bounded rationality. *Am. Psychol.* 58(9):697–720

Kakade S, Dayan P. 2002. Dopamine: generalization and bonuses. *Neural Netw.* 15:549–59

Kawabata H, Zeki S. 2004. Neural correlates of beauty. *J. Neurophysiol.* 91:1699–705

King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR. 2005. Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308:78–83

Knutson B, Fong GW, Adams CM, Varner JL, Hommer D. 2001. Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* 12:3683–87

Knutson B, Fong GW, Bennett SM, Adams CM, Hommer D. 2003. A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *Neuroimage* 18:263–72

Knutson B, Westdorp A, Kaiser E, Hommer D. 2000. FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage* 12:20–27

Koechlin E, Basso G, Pietrini P, Panzer S, Grafman J. 1999. The role of the anterior prefrontal cortex in human cognition. *Nature* 399:148–51

Koepp MJ, Gunn RN, Lawrence AD, Cunningham VJ, Dagher A, et al. 1998. Evidence for striatal dopamine release during a video game. *Nature* 393:266–68

Koopmans TC. 1960. Stationary ordinal utility and impatience. *Econometrica* 28 2:287–309

Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E. 2005. Oxytocin increases trust in humans. *Nature* 435:673–76

Krebs JR, Kacelnik A. 1984. Time horizons of foraging animals. *Ann. N.Y. Acad. Sci.* 423:278–91

Krebs JR, Kacelnick A, Taylor P. 1978. Test of optimal sampling by foraging great tits. *Nature* 275:27–31

Kreps D, Wilson R, Milgrom P, Roberts J. 1982. Rational cooperation in the finitely repeated Prisoners' Dilemma. *J. Econ. Theory* 27:245–52

Kringelbach ML, O'Doherty J, Rolls ET, Andrews C. 2003. Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. *Cereb. Cortex* 13:1064–71

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, et al. 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. USA* 89:5675–79

Laibson D. 1997. Golden eggs and hyperbolic discounting. *Q. J. Econ.* 112:443–77

Lane RD, Chua PM, Dolan RJ. 1999. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. *Neuropsychologia* 37:989–97

Lane RD, Reiman EM, Bradley MM, Lang PJ, Ahern GL, et al. 1997. Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia* 35:1437–44

Lavin A, Nogueira L, Lapish CC, Wightman RM, Phillips PEM, Seamans JK. 2005. Mesocortical dopamine neurons operate on distinct temporal domains using multimodal signaling. *J. Neurosci.* 25:5013–23

Lieberman MD. 2005. Principles, processes, and puzzles of social cognition: an introduction for the special issue on social cognitive neuroscience. *Neuroimage* 28:745–56

Ljungberg T, Apicella P, Schultz W. 1992. Responses of monkey dopamine neurons during learning of behavioral reactions. *J. Neurophysiol.* 67:145–63

Loewenstein G. 1996. Out of control: visceral influences on behavior. *Organ. Behav. Hum. Decis. Process.* 65:272–92

Loewenstein GF, Thaler RH. 1989. Anomalies: intertemporal choice. *J. Econ. Perspect.* 3:181–93

Luce RD, Raiffa H. 1957. *Games and Decisions*. New York: Wiley

McCabe K, Houser D, Ryan L, Smith V, Trouard T. 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl. Acad. Sci. USA* 98:11832–35

McClure SM, Daw N, Montague PR. 2003. A computational substrate for incentive salience. *Trends Neurosci.* 26(8):423–28

McClure SM, Berns GS, Montague PR. 2003. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38(2):339–46

McClure SM, Li J, Tomlin D, Cypert KS, Montague LM, Montague PR. 2004. Neural correlates of behavioral preference for culturally familiar drinks. *Neuron* 44:379–87

McClure SM, Laibson DI, Loewenstein G, Ericson K, McClure KD, Cohen JD. 2005a. Neural mechanisms of time discounting for primary reward. *Soc. Neuroecon.* Abstr.

McClure S, Gilzenrat M, Cohen J. 2005b. An exploration-exploitation model based on norepinepherine and dopamine activity. *Advances in Neural Information Processing Systems*, Vol. 18. Cambridge, MA: MIT Press

McCoy AN, Crowley JC, Haghighian G, Dean HL, Platt ML. 2003. Saccade reward signals in posterior cingulate cortex. *Neuron* 40:1031–40

Metcalfe J, Mischel W. 1999. A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psycholog. Rev.* 106:3–19

Miller EK, Cohen JD. 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24:167–202

Mobbs D, Greicius MD, Abdel-Azim E, Menon V, Reiss AL. 2003. Humor modulates the mesolimbic reward centers. *Neuron* 40:1041–48

Mobbs D, Hagan CC, Azim E, Menon V, Reiss AL. 2005. Personality predicts activity in reward and emotional regions associated with humor. *Proc. Natl. Acad. Sci. USA* 102:16502–6

Montague PR, Dayan P, Nowlan SJ, Pouget A, Sejnowski TJ. 1993. Using aperiodic reinforcement for directed self-organization. In *Advances in Neural Information Processing Systems*, 5:969–76. San Mateo, CA: Morgan Kaufmann

Montague PR, Sejnowksi TJ. 1994. The predictive brain: temporal coincidence and temporal order in synaptic learning mechanims. *Learning Memory* 1:1–33

Montague PR, Dayan P, Person C, Sejnowski TJ. 1995. Bee foraging in an uncertain environment using predictive Hebbian learning. *Nature* 376:725–28

Montague PR, Dayan P, Sejnowski T. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16:1936–47

Montague PR, Berns GS. 2002. Neural economics and the biological substrates of valuation. *Neuron* 36:265–84

Montague PR, Berns GS, Cohen JD, McClure SM, Pagnoni G, et al. 2002. Hyperscanning: simultaneous fMRI during linked social interactions. *Neuroimage* 16:1159–64

Montague PR, Hyman SE, Cohen JD. 2004. Computational roles for dopamine in behavioural control. *Nature* 431:760–67

Niv Y, Duff MO, Dayan P. 2005. Dopamine, uncertainty and TD learning. *Behav. Brain Funct.* 1:6

Nowak M, Sigmund K. 2005. Evolution of indirect reciprocity. *Nature* 437:1291–98

Ochs J, Roth AE. 1989. An experimental study of sequential bargaining. *Am. Econ. Rev.* 79:355–84

Ochsner KN, Bunge SA, Gross JJ, Gabrielli JD. 2002. Rethinking feelings: an fMRI study of cognitive regulation of emotion. *J. Cogn. Neurosci.* 14:1215–29

O'Doherty JP. 2004. Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* 14(6):769–76

O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ. 2003. Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia* 41(2):147–55

O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F, et al. 2000. Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. *Neuroreport* 11:399–403

O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ. 2002. Neural responses during anticipation of a primary taste reward. *Neuron* 33(5):815–26

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. 2003a. Temporal difference models and reward-related learning in the human brain. *Neuron* 38(2):329–37

O'Doherty JP, Critchley H, Deichmann R, Dolan RJ. 2003b. Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J. Neurosci.* 23:7931–39

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–54

O'Doherty JP, Buchanan TW, Seymour B, Dolan RJ. 2006. Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron* 49(1):157–66

O'Donoghue T, Rabin M. 1999. Addiction and self control. In *Addiction: Entries and Exits*, ed. J Elster, pp. 169–206. New York: Russell Sage

Ogawa S, Lee TM, Nayak AS, Glynn P. 1990. Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn. Reson. Med.* 14:68–78

Ogawa S, Lee TM, Kay AR, Tank DW. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA* 87:9868–72

Ogawa S, Menon RS, Tank DW, Kim SG, Merkle H, et al. 1993. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. *Biophys. J.* 64:803–12

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, et al. 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* 89:5951–55

Olds J. 1958. Self-stimulation of the brain. *Science* 127:315–24

Olds J. 1962. Hypothalamic substrates of reward. *Physiol. Rev.* 42:554–604

Olds J, Milner P. 1954. Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. Comp. Physiol. Psychol.* 47:419–27

Oster GF, Wilson EO. 1978. *Caste and Ecology in the Social Insects*. Princeton, NJ: Princeton Univ. Press

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

Paulus MP, Frank LR. 2003. Ventromedial prefrontal cortex activation is critical for preference judgments. *Neuroreport* 14:1311–15

Phillips MI, Olds J. 1969. Unit activity: motivation dependent responses from midbrain neurons. *Science* 165:1269–71

Platt ML, Glimcher PW. 1999. Neural correlates of decision variables in parietal cortex. *Nature* 400:233–38

Posner MI, Petersen SE, Fox PT, Raichle ME. 1988. Localization of cognitive operations in the human brain. *Science* 240:1627–31

Posner MI, Snyder CRR. 1975. Attention and cognitive control. In *Information Processing and Cognition*, ed. RL Solso, pp. 55–85. Hillsdale, NJ: Erlbaum

Quartz S, Dayan P, Montague PR, Sejnowski T. 1992. Expectation learning in the brain using diffuse ascending connections. *Soc. Neurosci. Abst.* 18:1210

Rachlin H. 2000. *The Science of Self-Control*. Cambridge, MA: Harvard Univ. Press.

Rapoport A, Chammah AM. 1965. *Prisoner's Dilemma: A Study in Conflict and Cooperation*. Ann Arbor: Univ. Mich. Press

Rilling JK, Gutman DA, Zeh TR, Pagnoni G, Berns GS, Kilts CD. 2002. A neural basis for social cooperation. *Neuron* 35:395–405

Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD. 2003. The neural basis of economic decision-making in the Ultimatum Game. *Science* 300:1755–58

Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD. 2004. The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22:1694–703

Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD. 2004. Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *Neuroreport* 15:2539–43

Robbins TW, Everitt BJ. 1996. Neurobehavioral mechanisms of reward and motivation. *Curr. Opin. Neurobiol.* 6:228–36

Roese NJ, Summerville A. 2005. What we regret most… and why. *Per. Soc. Psychol. Bull.* 31:1273–85

Rolls ET. 2000. *The Brain and Emotion*. Oxford, UK: Oxford Univ. Press

Rolls ET, Kringelbach ML, de Araujo IE. 2003a. Different representations of pleasant and unpleasant odours in the human brain. *Eur. J. Neurosci.* 18:695–703

Rolls ET, O'Doherty J, Kringelbach ML, Francis S, Bowtell R, McGlone F. 2003b. Representations of pleasant and painful touch in the human orbitofrontal and cingulate cortices. *Cereb. Cortex* 13:308–17

Rorie AE, Newsome WT. 2005. A general mechanism for decision-making in the human brain? *Trends Cogn. Sci.* 9:41–43

Rosenstein MT, Barto AG. 2004. Supervised actor-critic reinforcement learning. In *Learning and Approximate Dynamic Programming: Scaling up to the Real World*, ed. J Si, A Barto, W Powell, D Wunsch, pp. 359–80. New York: Wiley

Roth A. 1995. Bargaining experiments. In *Handbook of Experimental Economics*, ed. JH Kagel, AE Roth, pp. 253–348. Princeton, NJ: Princeton Univ. Press

Royet JP, Zald D, Versace R, Costes N, Lavenne F, et al. 2000. Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: a positron emission tomography study. *J. Neurosci.* 20:7752–59

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD. 2003. The neural basis of economic decision making in the ultimatum game. *Science* 300:1755–57

Schultz W. 2000. Multiple reward signals in the brain. *Nat. Rev. Neurosci.* 1:199–207

Schultz W, Apicella P, Ljungberg T. 1993. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* 13:900–13

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. *Science* 275:1593–99

Schultz W, Dickinson A. 2000. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* 23:473–500

Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, et al. 2004. Temporal difference models describe higher-order learning in humans. *Nature* 429:664–67

Shadlen MN, Newsome WT. 2001. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* 86:1916–36

Shallice T, Burgess P. 1991. Higher-order cognitive impairments and frontal lobe lesions in man. In *Frontal Lobe Function and Dysfunction*, ed. HS Levin, HM Eisenberg, AL Benton, pp. 125–38. New York: Oxford Univ. Press

Shefrin HM, Thaler RH. 1988. The behavioral life-cycle hypothesis. *Econ. Inq.* 26:609–43

Shizgal P. 1997. Neural basis of utility estimation. *Curr. Opin. Neurobiol.* 7:198–208

Shizgal P, Fulton S, Woodside B. 2001. Brain reward circuitry and the regulation of energy balance. *Int. J. Obesity* 25:S17–21

Sigmund K, Fehr E, Nowak MA. 2002. The economics of fair play. *Sci. Am.* 286:82–87

Simon H. 1955. A behavioral model of rational choice. *Q. J. Econ.* 69:99–118

Singer T, Seymour B, O'Doherty JP, Stephan KE, Dolan RJ, Frith CD. 2006. Empathic neural responses are modulated by the perceived fairness of others. *Nature* 439:466–69

Stuss DT, Benson DF. 1986. *The Frontal Lobes*. New York: Raven Press

Sugrue LP, Corrado GS, Newsome WT. 2004. Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782–87

Sugrue LP, Corrado GS, Newsome WT. 2005. Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat. Rev. Neurosci.* 6:1–13

Sutton RS. 1988. Learning to predict by the methods of temporal difference. *Mach. Learn.* 3:9–44

Sutton RS, Barto AG. 1998. *Reinforcement Learning*. Cambridge, MA: MIT Press

Tataranni PA, Gautier JF, Chen K, Uecker A, Bandy D, et al. 1999. Neuroanatomical correlates of hunger and satiation in humans using positron emission tomography. *Proc. Natl. Acad. Sci. USA* 96:4569–74

Tobler PN, Fiorillo CD, Schultz W. 2005. Adaptive coding of reward value by dopamine neurons. *Science* 307:1642–45

Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185:1124–31

Tversky A, Kahneman D. 1981. The framing of decisions and the psychology of choice. *Science* 211:453–58

Usher M, Cohen JD, Servan-Schreiber D, Rajkowski J, Aston-Jones G. 1999. The role of locus coeruleus in the regulation of cognitive performance. *Science* 283:549–54

Uvnas-Moberg K. 1998. Oxytocin may mediate the benefits of positive social interaction and emotions. *Psychoneuroendocrinology* 23:819–35

Vartanian O, Goel V. 2004. Neuroanatomical correlates of aesthetic preference for paintings. *Neuroreport* 15:893–97

von Neumann J, Morgenstern O. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press

Waelti P, Dickinson A, Schultz W. 2001. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412:43–48

Watkins CJCH. 1989. *Learning from delayed rewards*. PhD thesis. Cambridge Univ. King's College. Cambridge, UK

Watkins CJCH, Dayan P. 1992. Q-learning. *Mach. Learn.* 8:279–92

Yu AJ, Dayan P. 2005. Uncertainty, neuromodulation, and attention. *Neuron* 46:681–92

Zald DH, Hagen MC, Pardo JV. 2002. Neural correlates of tasting concentrated quinine and sugar solutions. *J. Neurophysiol.* 87:1068–75

RELATED RESOURCES

Hyman SE, Nestler EJ, Malenka RC. 2006. Reward-related memory and addiction. *Annu. Rev. Neurosci.* 29:565–98

Nader K, Bechara A, van der Kooy D. 1997. Neurobiological constraints on behavioral models of motivation. *Annu. Rev. Psychol.* 48:85–114

Packard MG, Knowlton BJ. 2002. Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.* 25:563–93

Rolls ET. 2000. Memory systems in the brain. *Annu. Rev. Psychol.* 51:599–630

Schultz W. 2006. Behavioral theories and the neurophysiology of reward. *Annu. Rev. Psychol.* 57:87–115

Contents

INDEXES

ERRATA

An online log of corrections to *Annual Review of Neuroscience* chapters (if any, 1977 to
the present) may be found at http://neuro.annualreviews.org/