74 The Neuroeconomics of Simple Goal-Directed Choice

ANTONIO RANGEL

ABSTRACT This paper reviews what is known about the computational and neurobiological basis of simple goal-directed choice. Two features define this type of choice. First, individuals make decisions between stimuli that are associated with different outcomes or rewards. Second, the brain solves the decision problem by (1) computing the distribution of outcomes associated with each stimulus, (2) assigning a value to each stimulus equal to the expected reward generated by those outcomes, and (3) selecting the stimulus with the highest computed value. A typical example of simple goal-directed choice is given by the problem of choosing a meal from a buffet table.

Neuroeconomics studies the computational and neurobiological basis of animal and human decision making. Its goal is to understand how the brain solves the multitude of choice problems that organisms face every moment of their existence. One important complication in addressing these problems is that decision-making situations come in many different flavors, and it is likely that the brain uses different computations and systems to solve them. Compare, for example, the problem of a lion chasing a gazelle with the problem of a typical consumer deciding which of two cereal boxes to purchase. Both organisms are engaged in decision making, but their problems are very different. The problem of the lion is to select a direction of movement every instant to increase the probability of catching the gazelle. This entails a simple goal ("catch the gazelle"), but a series of action choices. In contrast, the consumer faces a complicated choice between goals ("which cereal box has the best taste-health-price combination?"), but once that decision has been made, the choice over actions is trivial ("pick the motor plan that grabs the chosen cereal box"). Given this complexity, an important task for neuroeconomics is the construction of a neurally relevant taxonomy of choice tasks that can be used to guide the research and to organize the findings.

Another difficulty in neuroeconomics is that there does not seem to be a simple one-to-one mapping between decision-making situations and the neural processes used to make choices. Instead, a sizable and rapidly growing body of animal and human evidence suggests that there are at least three conceptually and neurally separable behavioral controllers at work in most decision-making situations: a Pavlovian system, a habitual system, and a goal-directed system (Balleine, Daw, & O'Doherty, 2008; Daw, Niv, & Dayan, 2005; Dayan, 2008; Dickison & Balleine, 2002; Rangel, Camerer, & Montague, 2008). Although this topic is just beginning to be explored, the existing evidence suggests that the relative importance of the systems changes with the details of the decision-making situation. Given these two complications, it is unlikely that we will find a simple neuroeconomic theory of decision making that covers all types of choice situations.

In order to deal with these two complications, research in neuroeconomics typically focuses its attention on a subset of the behavioral controllers and a well-defined subclass of choice problems. In this chapter we review what is known about a class of problems that has received considerable attention in neuroeconomics and behavioral neuroscience: How does the goal-directed system make choices among sets of stimuli associated with different rewards? We refer to this problem as simple goal-directed choice. The decision-making situations of interest resemble the example of the consumer who has to choose one type of cereal among several options. The consumer cares about which choice he makes because the different stimuli are associated with different combinations of outcomes (or rewards). For example, one cereal box might be tastier and cheaper than another. It is important to emphasize that animals also engage in this type of choice. As an example, consider the problem of a rat that has to press a left or a right lever on a Skinner box in order to obtain one of two rewards, or the problem of a hungry lion confronted with several gazelles.

The review has several goals. First, we show that several choice tasks that have been used in the neuroscience and animal learning literatures are special cases of simple goal-directed choice. Second, we provide a mathematical description of the computations that define the goal-directed system. Third, we use the computational framework as a way to organize what is known about the neurobiology of

---

ANTONIO RANGEL Humanities and Social Sciences, Computational and Neural Systems, California Institute of Technology, Pasadena, California.

the goal-directed system and what are some of the most important open questions.

Simple binary stimulus choice: A behavioral paradigm to study simple goal-directed choice

In a simple binary stimulus-choice task, individuals make repeated choices between pairs of stimuli that are presented to them, one on the left and one on the right. Individuals care about their choice because the stimuli are associated with different outcomes (or rewards) that affect their well-being. They indicate their choice by executing one of two different actions associated with each of the stimuli (e.g., a left or a right button push, a left or a right saccade, etc.). The actions are such that the costs and effort required to execute them are as similar as possible.

A typical example of such a task is depicted in figure 74.1*A* (Karjbich, Armel, & Rangel, 2008). Individuals are shown pairs of high-resolution pictures of familiar snack-food items in a computer monitor and have to choose which one they would like to consume at the end of the experiment by pressing either a left or a right button. There are 70 different such stimuli that are randomly assigned into pairs in 100 different trials. At the end of the experiment, one of the trials is selected at random, and the subject eats the food depicted in the picture that he chose in that trial. Another example is shown in figure 74.1*B* (Baxter & Murray, 2002). During an initial training phase, monkeys are exposed to 120 objects of different shape and color, two at a time. Importantly, 60 of the objects are associated with a food reward (half with a cherry and half with a peanut) that is placed below the object, whereas the other 60 objects are associated with no such reward. The goal of this phase is for animals to learn to associate the 30 cherry objects with the consumption of a cherry and the 30 peanut objects with the consumption of a peanut. During a second training phase, the animals are repeatedly presented with one cherry object and one peanut object, and are taught to make a choice by lifting only one of the objects and grabbing the reward underneath. The location of the objects is fully randomized. After the animals are fully trained, they are tested in one of three conditions: (1) sessions that are preceded by feeding to satiety with cherries, (2) sessions that are preceded by feeding to satiety with peanuts, and (3) sessions with no prefeeding.

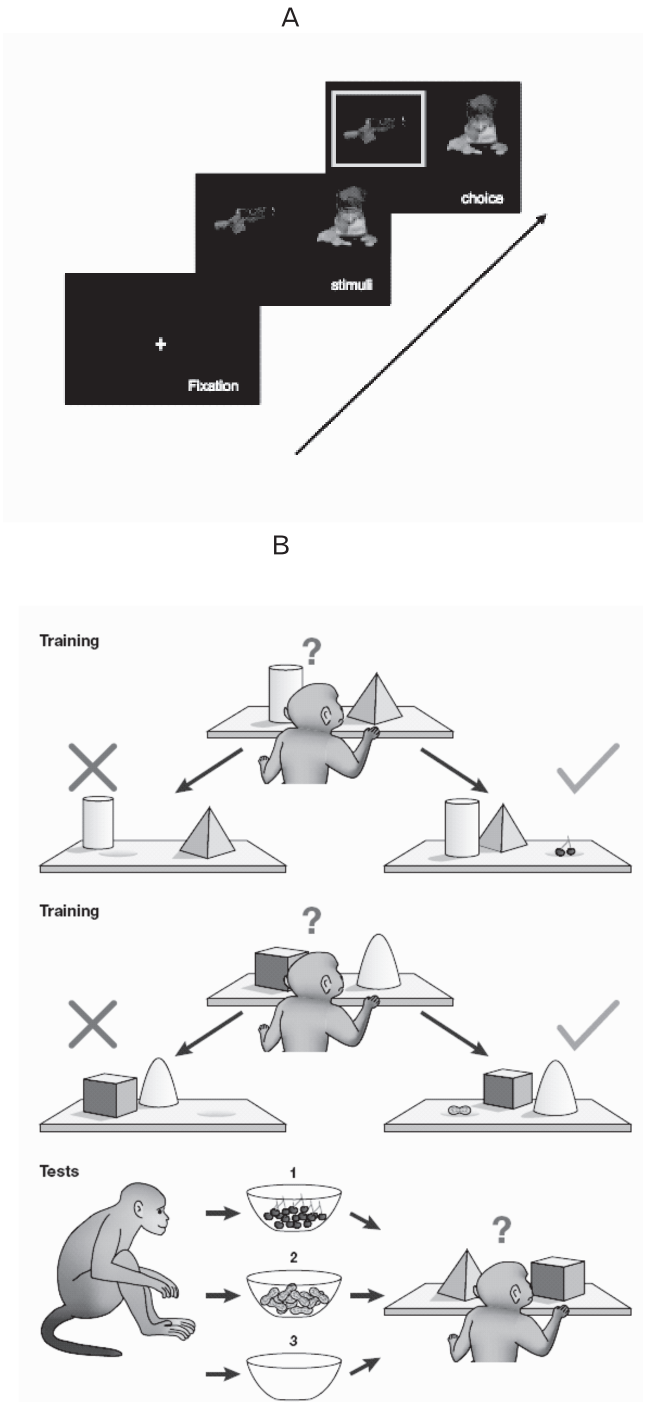

Figure 74.1 Examples of simple binary stimulus choice tasks. (*A*) Binary food choice from Karjbich, Armel, and Rangel (2008). (*B*) Devaluation choice task from Izquierdo, Suda, and Murray (2004). (With permission from Baxter & Murray, 2002.) 7

Note some of the central elements of the simple binary stimulus-choice task. First, there are at least 2 stimuli that the subjects choose from, although the set may be much larger. For later reference, let S denote the set of stimuli and s denote a typical element. Second, each stimulus is associated with a probability over outcomes. Let O denote the set of potential outcomes, o denote a typical outcome, and $p(o|s)$ denote the probability that the subject gets outcome o if stimulus s is chosen. The outcomes can be appetitive (e.g., food) or aversive (e.g., a shock). A stimulus might be paired with multiple outcomes. In the simple tasks we have described, the stimulus-outcome associations are degenerate and time-invariant probability distributions, but we need the more general notation to accommodate other tasks of interest. For example, in reversal-learning tasks the stimulus-outcome associations change with time, and thus we have to write $p^t(o \mid s)$. Third, the mapping from stimuli to action, denoted by $a^t(s)$, changes from trial to trial. As a result, there is not a fixed mapping between actions and stimuli, or actions and outcomes. This approach captures the fact that in the real world the actions required to obtain a particular stimulus often change over time. Fourth, all the actions required to make or implement the choice entail approximately the same costs to the individual. An example would be a Skinner box with two levers that have equal tension, symmetric location, and so on. The property would be violated if one of the levers is more difficult to pull. Note that in order to keep things simple we make an explicit distinction between the potential aversive outcomes associated with a stimulus (which occur at the time of outcome consumption) and the costs associated with taking the action necessary to obtain the stimulus (which occur at the time of choice).

Another important feature of the binary stimulus-choice task is previous experience consuming all the possible outcomes in all the states of the world that might be induced by the experimenter. To understand why this is important, consider the devaluation experiment depicted in figure 74.1*B*. Here monkeys are asked to make choices among stimuli in three different states of the world: a cherries-satiation condition, a peanut-satiation condition, and a no-satiation condition. To qualify as a simple binary stimulus-choice task, monkeys must have had extensive experience consuming the cherries and peanuts in the three states of the world prior to the actual experiment. As we will see, the choices made by the goal-directed system will depend on how it values the outcomes associated with each stimulus *given the state of the world*. If the subjects have not had experience consuming the outcomes in a particular state, they might need to learn how to evaluate them, a phenomenon that Balleine has called incentive learning (Balleine & Dickinson, 1998). This might result in unstable choices across the experiment. The simple binary stimulus-choice paradigm removes this complication by requiring that subjects have extensive prior experience with all the outcomes in all the relevant states of the world.

The other details of the task are not important and can take many different forms. For example, the stimuli could be pictures on a computer screen, or physical objects with different shape and color, or cards with printed photographs or verbal descriptions of rewards, or even real exposure to the actual outcomes. Subjects might get a reward after every decision, or might get rewarded only for a random subset of the choices that they made at the end of the trial. There are also no constraints on the actions associated with choosing a stimulus, as long as they satisfy the equal-cost property. Thus subjects might indicate their choice through an eye movement and then get the chosen liquid delivered to their mouths, or they may indicate their choice through the act of reaching for one of the stimuli in order to consume it.

The binary stimulus-choice paradigm outlined here covers as special cases several tasks that have been used in the neuroscience and animal learning literatures. First are the type of simple binary choices described in figure 74.1*A* (Kable & Glimcher, 2007; Karjbich et al., 2008; Padoa-Schioppa & Assad, 2006, 2008; Tom, Fox, Trepel, & Poldrack, 2007; Wallis & Miller, 2003). Second are the devaluation choice tasks described in figure 74.1*B* (Izquierdo, Suda, & Murray, 2004; Wellman, Gale, & Malkova, 2005). The main differences from the previous set of tasks are that subjects indicate their choice by lifting an object, instead of pressing a button or executing a saccade, and that the value of the outcomes is manipulated by feeding the subject to satiation on some of the foods. Third are reward preference tasks (Izquierdo et al.). The key difference from the previous task is that subjects are exposed to the actual rewards, instead of stimuli associated with them, and they indicate their choice by reaching for the chosen outcome. Fourth are reversal learning tasks (Hampton, Bossaerts, & O'Doherty, 2006). In a typical version of these tasks there are two stimuli and one potential outcome. In every trial the probability of obtaining the outcome is high for one of the stimuli and low for the other, and it evolves over time either through either an exogenously specified process or as a function of the history of choices.

It is important to emphasize that, as general as it is, this behavioral paradigm does not cover many decision situations that have also been used in the literature to study the goal-directed system. It rules out the case of multistimulus (nonbinary) choice. It also rules out a popular odor discrimination task from the rat literature (Schoenbaum, Chiba, & Gallagher, 1998) in which rats decide whether or not to drink a liquid from a single location based on the odor that they receive in an odor port (some odors predict rewards like sugar, others punishers like quinine). Note that instead of a choice between stimuli, this task entails choice between motor plans with fixed state-dependent action-outcome relationships. More generally, the paradigm also rules out any tasks in which there is a constant mapping between actions and stimuli or outcomes at each state of the world. It also rules out instrumental paradigms in which animals engage in free rates of responding, even if they have a choice among multiple responses (Balleine & Dickinson, 1998; Dayan & Balleine, 2002). This last class of decision tasks is substantially more complicated as animals need to decide not only what to choose, but also about when to take action. In contrast, in the binary stimulus-choice paradigm the timing of decision making is controlled by the experimenter.

Multiple behavioral controllers: What is goal-directed choice?

As mentioned at the beginning of the chapter, a growing body of evidence suggests that the brain might deploy different behavioral controllers in parallel in many decision-making situations. In this section we provide a brief review of the computational differences between the three main systems that have been identified: a habitual, a goal-directed, and a Pavlovian system. For more detailed reviews see Balleine and colleagues (2008), Dayan (2008), and Rangel and colleagues (2008).

GOAL-DIRECTED SYSTEM The defining feature of the goal-directed system is that it makes choices over stimuli using model-based computations of value. These are carried out in two steps. First, a value is assigned to each stimulus by identifying the distribution of outcomes associated with it and computing the expected values of those outcomes in the current state of the world. Second, the computed values are compared in order to select one of the stimuli.

Note several key properties of this system. First, its goal is to make choices over stimuli, not actions. Second, it assigns value to stimuli by anticipating the outcomes to which they might lead and then computing their expected reward. Third, this computation is based on stimulus-outcome associations and beliefs about the reward that those outcomes are likely to generate in the current state of the world. It follows that the computation is model based (sometimes called forward looking) and is not based on the historical level of payoff generated by the different stimuli. This last property gives enormous flexibility to the system, since it allows it to rapidly update the value that it assigns to stimuli based on either a change on the stimulus-outcome associations or a change in the state of the world that affects their expected value. Fourth, this flexibility comes at the cost of computational complexity. The brain needs to store or compute stimulus-outcome associations and state-dependent value functions, and then needs to carry out expected value computations online.

HABITUAL SYSTEM The defining feature of the habit system is that it makes choices over actions based on the historical level of rewards that they have generated. This is also done in two steps. First, a value for each of the available actions is retrieved from memory. Second, the retrieved values are compared in order to select one of the stimuli.

Note several key properties of this system. First, choice is made over actions, not over stimuli. Second, the values used to make choices are retrieved from memory, not computed online. Third, the values assigned to actions depend on the level of rewards that they have generated in the past. Multiple studies have shown that relatively simple reinforcement learning algorithms approximate well the process of value learning for this system (Montague, Dayan, & Sejnowski, 1996; Niv & Montague, 2008; Schultz, Dayan, & Montague, 1997; Sutton & Barto, 1998). Fourth, the fact that the learning can be state sensitive leads to the use of state-dependent action values by the habit system. Fifth, the computations made by the habitual system at the time of choice are simpler than those of the goal-directed system, since values are retrieved from memory instead of computed online. Sixth, this computational simplicity comes at the cost of some behavioral flexibility. Although with enough experience the habitual system is able to make optimal decisions in environments that are sufficiently stable, it cannot do so when the action-outcome contingencies are rapidly changing (as, for example, in the simple experiment described in figure 74.1*A*).

PAVLOVIAN SYSTEM In contrast to the previous two systems, which are able to assign values to any stimulus or action, the Pavlovian system assigns values to a small set of actions that are evolutionarily appropriate responses to particular environmental stimuli. Typical examples include preparatory behaviors (such as approaching cues that predict the delivery of food) and consummatory responses to a reward (such as pecking at a food magazine).

Although many Pavlovian behaviors are "hardwired" responses to specific predetermined stimuli, with sufficient experience animals can also learn to deploy them in response to other stimuli. For example, rats and pigeons learn to approach lights that predict the delivery of food. At first glance, Pavlovian behaviors look like automatic, stimulus-triggered responses, and not like instances of value-based choice processes. However, since Pavlovian responses can be interrupted, they must be assigned something akin to a "value" so that they can compete with the actions that are favored by the other valuation systems.

The computational and neurobiological basis of the Pavlovian system is much less well understood than that of the habitual and the goal-directed systems. For recent reviews see Dayan and Seymour (2008) and Rangel and colleagues (2008). This lack of understanding is due, in part, to the fact that there might be multiple Pavlovian controllers, some responsible for triggering outcome-specific responses (e.g., pecking at food or licking at water) and others responsible for triggering more general valence-dependent responses (e.g., approaching positive outcomes and withdrawing from negative ones). Nevertheless, since a wide range of human behaviors with important economic consequences might be controlled by the Pavlovian system (from overeating to the harvesting of immediate smaller rewards at the expense of larger delayed rewards), a detailed understanding of this system is an important open question for neuroeconomics.

Coexisting and Competing Valuation Systems All these behavioral controllers can potentially be active at the same time even in the case of simple binary stimulus choice. Consider, for example, the experiment in figure 74.1*B*. Since some of the stimuli covering the food rewards have been associated with appetitive outcomes, they might trigger Pavlovian approach responses that could influence which of the two objects the monkey lifts first. Similarly, since the execution of the choice entails two constant motor actions (reach for the left object or reach for the right object) and the monkeys receive extensive experience in the task, the habitual system might use historical action values to influence the choice that is made. Finally, the goal-directed system could also bias the monkey's actions by assigning a higher value to the actions associated with the higher-value stimulus. This possibility leads to a very important open question in neuroeconomics about which next to nothing is known: How does the brain assign control to the three different systems? Although some simple computational models have been proposed (Daw et al., 2005; Dayan, Niv, Seymour, & Daw, 2006), to date no experiments have been performed to study how the systems interact and compete at the neural level in simple binary stimulus choice.

In this review we focus on the computations of the goal-directed system during the simple binary stimulus-choice task. We do so not because the effects of the other systems in this type of situations are unimportant, but because much more is known about the role of the goal-directed system. We emphasize, however, that a full understanding of simple stimulus choice will require the study of how the other two systems are deployed in this type of task and of how the allocation of control is resolved.

Computational basis of goal-directed choice in the simple binary stimulus-choice paradigm

In this section we provide a mathematical description of the computations that the goal-directed system needs to make in simple binary stimulus-choice situations.

Representation of the Choice Problem The first problem that the system needs to solve is to identify the parameters of the decision-making problem: What are the potential stimuli that could be chosen? What are the actions required to obtain each stimulus? What external and internal state variables might affect the desirability of the different stimuli and actions? Let e be a summary of the internal and external variables determining the state of the world.

This part of the choice process is often ignored in decision-making models by implicitly assuming that the brain always computes these variables correctly. But given the complexity of the world, it is likely that the brain relies on computational shortcuts. Consider, for example, the problem of a shopper in a modern supermarket aisle that contains thousands of different products. When confronted with such complexity, the brain only evaluates and compares a small subset of the possible items. Since an item is chosen only if it is considered, the representation step has a large impact on the choice that is eventually made. Given the large number of external and internal variables that can impact the choice situation, similar issues are likely to arise in the identification of the relevant states of the world. The algorithms and neural processes at work, as well as the limitations on choice performance to which they lead, are just beginning to be understood (Reutskaja, Pulst-Korenhberg, Nagel, Camerer, & Rangel, under review). Basic open questions include the following: 2 How does the brain determine which actions to assign values to and which actions to ignore? Is there a limit to the number of options that animals can consider at a time? How are internal and external states computed?

Stimulus Valuation As we saw before, the goal-directed system makes choices by assigning values to the different stimuli based on the expected value of the outcomes associated with them. Let $V(s|e)$ denote the value of stimulus s given the state of the world e. In order to compute this value, the system needs two pieces of information: (1) the stimulus-outcome associations, which are summarized by the function $q(o|s)$ specifying the probability that every potential outcome o occurs as a function of the stimulus s, and (2) the value function $v(o|e)$ specifying the value of each outcome given the state of the world.

Note several things about this notation. First, there is a difference between the $p(o|s)$ function that describes the objective mapping between stimulus and outcomes and the $q(o|s)$ function that describes the beliefs of the subject about that relationship. Second, by assumption, the stimulus-outcome associations do not depend on the state of the world, and, for simplicity, we assume that the subjects always know this fact. Third, the value function $v(o|e)$ is the goal-directed system's belief about the reward that it will experience if the outcome occurs, which is a different signal than the level of reward that actually occurs at the time of consumption. Fourth, the value function $v(o|e)$ does not depend on the stimulus. The reason is that there is a conceptual distinction between the (positive or negative) outcomes generated by an stimulus and the costs of taking the action necessary to get that stimulus.

The value assigned to a stimulus is simply the expected value of the outcomes to which it might lead. This is given by

$$V(s|e)=\sum_{o\in O} p(o|s)v(o|e)$$

STIMULUS CHOICE The brain uses the net-value information to make a choice between the stimuli. A sizable amount of behavioral evidence suggests that the maximization process is stochastic and well approximated by a *soft-max* process in which the probability of choosing stimulus s is given by

$$\frac{e^{\tau V(s|e)}}{\sum_{t\in S} e^{\tau V(t|e)}}$$

where τ is a coefficient measuring the sensitivity of the choices to the stimulus values (when $\tau = 0$ each alternative is chosen with equal probability regardless of the values, and for sufficiently large τ almost all of the probability falls on the item with the highest value).

The soft-max model is a reduced-form model of limited use for neuroeconomics, since it describes how the probability of making a choice changes with the net values, but not how the choice is actually made. A large research effort is devoted to this problem (for recent reviews see Bogacz, 2007; Busemeyer & Johnson, 2004; Ditterich, 2006; Gold & Shadlen, 2007; Rangel, 2008). Most of the models that have been proposed are versions of a race-to-barrier diffusion process. A simple version of the model for the case of two alternatives is depicted in figure 74.2*A*. The model has several components. First, there are circuits that compute the value of each of the items. The value assigned to the items is assumed to fluctuate stochastically from instant to instant. Every instant, the two value signals are subtracted to produce a relative-value signal that is then fed to an integrator circuit that computes the value of item 1 minus the value of item 2, thus keeping track of the accumulated relative signal. A decision is made when this relative-value signal becomes sufficiently large ("choose item 1") or sufficiently negative ("choose item 2").

This class of models has several attractive features. First, they predict a logistic choice function similar to the one generated by the soft-max model. Second, they predict that the time required to make a choice should be larger when items have similar values than when the values are far apart.

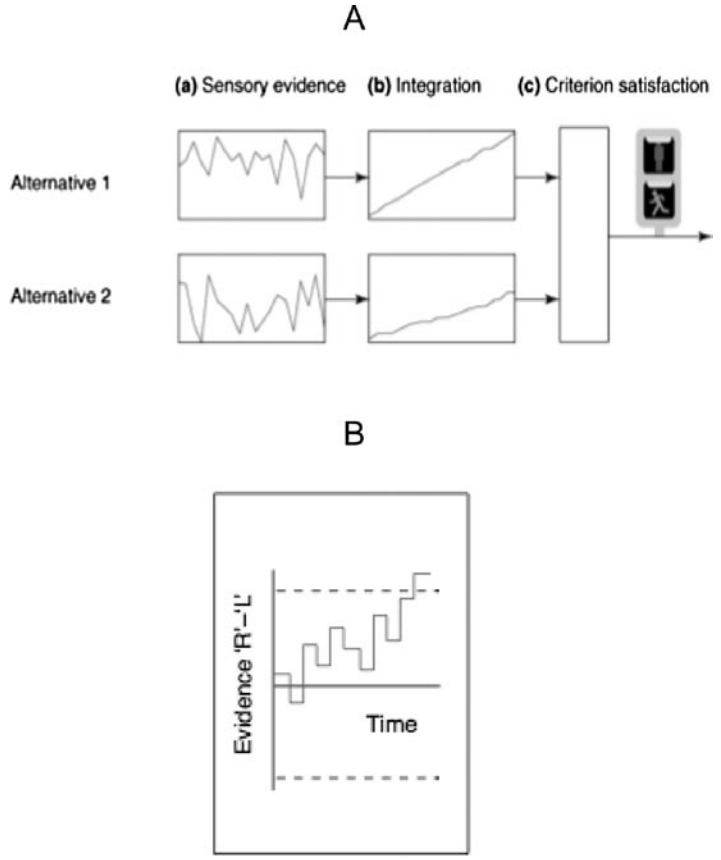

FIGURE 74.2 Models of the value comparison process. (*A*) Illustration of the main components of the race-to-barrier models. (Adapted with permission from Bogacz, 2007.) (*B*) A typical run of the random walk model. The step function represents the accumulated relative value of the "right" target. The process starts at a middle point and stops the first time this variable crosses one of the thresholds (depicted by the bracketed horizontal lines). "Right" is chosen when it crosses the upper threshold; "left" is chosen when it crosses the lower one. Time advances in discrete steps. The size of every step is given by a Gaussian distribution with a mean that is proportional to the true direction of motion. This noise is meant to capture the variability in the valuation processes.

Both predictions are consistent with a large body of behavioral data. Finally, the model makes useful predictions about which kind of computations should be implemented by the brain: there should be circuits computing the value of each stimulus, circuits computing relative values, an integrator circuit, and a circuit that triggers a choice when a barrier is crossed.

These models assume that a choice is first made over stimuli and that the choice is then implemented by deploying the action that leads to that stimulus. We refer to these types of models as *stimulus-based choice.* Another a priori equally plausible theory (Glimcher, Dorris, & Bayer, 2005) specifies that the brain uses the stimulus values to assign a value to every feasible action, and that it then makes the choice through a process of competition over action plans. We refer to this possibility as *action-based choice.* It is difficult to compare these two views on theoretical grounds, since the race-to-barrier models apply well to both types of choice. Thus novel experiments are needed to address this issue. The question is important because the neural systems involved in making the choice are likely to be different in the case of stimulus- and action-based choice.

Learning In some versions of the simple binary stimulus-choice paradigm, subjects receive an outcome after every choice. This provides them with feedback that can be used to update their estimate of the stimulus-outcome associations. Here we propose a simple algorithm that subjects can use to carry out this type of learning.

We assume that the experimental task is structured as follows. Every experimental trial t begins with the revelation of the current state of the world (e^t). A stimulus s^t is then chosen that leads to the set of outcomes O^t and a level of reward r^t.

We assume that learning takes place in two stages. In the first stage a prediction error is computed for every possible outcome in the set O. These prediction errors are given by

$$\delta^t(o) = I_O - q^t\left(o \middle| s^t\right)$$

where I_0 is an indicator function taking a value of 1 if the outcome in question occurs and a value of zero otherwise. Note that positive prediction errors measure the degree to which the occurrence of an outcome was surprising, and negative prediction errors measure the extent to which the nonoccurrence of the other outcomes was surprising. In the second stage the prediction errors are used to update the state-outcome probability function for that stimulus by

$$q^{t+1}\left(o \middle| s^t\right) = q^t\left(o \middle| s^t\right) + \lambda\delta^t(o)$$

where λ is a learning rate between 0 and 1 that affects the speed of learing.

Note that this formulation assumes that only the beliefs for the stimulus that was chosen are updated. This approach assumes a very strong form of discrete learning, an assumption which is plausible in environments where there are a small number of highly dissimilar stimuli but not in domains in which "similar" stimuli have "similar" stimulus-outcome associations. In the later case, the outcome in one state can provide information about the stimulus-outcome associations for other states. The extent to which the goal-directed system engages in this type of generalization is largely unknown.

Neurobiological basis of goal-directed choice in the simple binary stimulus-choice paradigm

In this section we review some of what is known about how the brain implements the computations described in the previous section and highlight some important open questions. For alternative recent reviews see Balleine and colleagues (2008), Rangel (2008), Rangel and colleagues (2008), and Wallis (2007).

Representation Unfortunately, next to nothing is known about this important step in the decision-making process. Open questions of particular interest include the following. How does the brain know when to activate the goal-directed evaluation and comparison circuitry? How does the brain decide which stimuli to evaluate at any given moment? Which aspects of the state of the world are measured, and how are they encoded by the goal-directed evaluation circuitry? The first question is important because organisms are exposed to potential choice stimuli continuously, but the goal-directed choice might only engage in the process of choice sporadically. The second question is important because often there are many potential stimuli and the system might not have the capacity to evaluate all of them fully. Think, for example, of a consumer in a modern supermarket aisle.

Stimulus Valuation Several papers have found neural correlates of the stimulus-value signal ($V(s|e)$). Plassmann, O'Doherty, and Rangel (2007) investigated the neural correlates of stimulus valuation by the goal-directed system in humans using fMRI. They showed pictures of desirable snacks to hungry subjects who had to place bids for the right to eat them at the end of the experiment. The size of the bids was a measure of the value assigned by the brain to each stimulus at the time of choice and positively correlated with BOLD activity in the mOFC and the dorsolateral prefrontal cortex (DLPFC). (For related fMRI findings see Arana et al., 2003; Erk, Spitzer, Wunderlich, Galley, & Walter, 2002; Hare, O'Doherty, Camerer, Schultz, & Rangel, 2008; Paulus & Frank, 2003; Tom et al., 2007; Valentin, Dickinson, & O'Doherty, 2007).

A related study used single-unit electrophysiology in nonhuman primates to look for activity in the orbitofrontal cortex that correlates with stimulus values (Padoa-Schioppa

& Assad, 2006, 2008). Every trial, thirsty animals were given a choice between two stimuli associated with small magnitudes of two different juices. After a period of deliberation, the animals indicated their choice with a left-or-right eye movement. The action associated with each stimulus varied from trial to trial. The authors estimated a logistic-choice model to compute a measure of value for each juice-amount combination that was then correlated with the neural signals. They found a large population of neurons encoding the value of the stimulus associated with each juice independently of the action that it took to get it. They did not find an equivalent population encoding the value of the actions. A closely related study recorded simultaneously from monkey OFC and DLPFC and found neurons encoding for the value of stimuli in both areas, although the value signal arose in DLPFC with a delay of approximately 100 ms (Wallis & Miller, 2003).

The previous studies looked for stimulus-value signals in the case in which animals made choices between appetitive items. An important question is whether the brain uses the same networks to evaluate stimuli associated with aversive items (e.g., choosing which of two undesirable risks to take). Plassmann, O'Doherty, and Rangel (2008) used an experimental design similar to the one we have described to study this question. Subjects were shown pictures of undesirable food items (e.g., canned vegetables) and had to bid to avoid having to eat them. The bids were a measure of the extent to which they disliked the foods. Interestingly, no areas exhibited a positive and significant correlation with this measure of stimulus value. Instead, the study found that activity in the mOFC and the DLPFC were negatively correlated with the bids. This finding suggests that these two structures play a role in the valuation of both appetitive and aversive items, in the appetitive case through increased activity and in the aversive case through decreased activity.

Interestingly, given that the $V(s|e)$ function is a forecast of the actual value of consuming the objects associated with the stimuli, activity in the OFC has also been shown to be correlated with the value of expected outcomes in the absence of choice. For example, Gottfried, O'Doherty, and Dolan (2003) presented subjects with visual stimuli that were paired with different odors and used a devaluation procedure to manipulate the value of some of the odors. Using human fMRI, they found that activity in amygdala and OFC was consistent with the encoding of the expected odor value at the time of cue presentation (prior to the actual odor delivery). (For related human fMRI studies see Gottfried, O'Doherty, & Dolan, 2002; Nobre, Coull, Frith, & Mesulam, 1999; O'Doherty, Deichmann, Critchley, & Dolan, 2002). These findings, together with the ones for goal-directed choice described previously, suggest that the OFC might be involved in the computation of different types of value signals at different stages of the choice process and in different types of tasks.

In all the previous experiments, there were no costs associated with choosing an item. Hare and colleagues (2008) studied a simple choice paradigm in which subjects had to make a decision about whether or not to buy a food snack at a given price. In this case, acquiring the stimulus entailed a cost equal to a loss of money given by the price. Consistent with the studies described before, they found that the value of the foods correlated with activity in the medial OFC, but that the price was not encoded in this area. Instead, a "consumer surplus" signal, equal to the value of the item minus its price, was found in the central OFC. These results suggest that the medial OFC might be involved in the encoding of stimulus value but is not responsive to the costs of acquiring the item.

A difficulty in identifying areas where stimulus values might be encoded is that these signals are most likely positively correlated with other signals that are not part of the goal-directed-system valuation process. Consider several examples that have caused some confusion in the literature. First, exposure to stimuli with very positive or very negative stimulus values might induce an increase in arousal in systems associated with motor preparation. If the experimental condition only includes appetitive items, the arousal and stimulus-value signals will be perfectly correlated, and thus one might misattribute one type of signal for the other. As proposed by Roesch and Olson (2004), one way of dissociating the two signals is to include both appetitive and aversive items in the experiment: neural value signals increase linearly with stimulus value, whereas arousal signals are correlated with the absolute value of the stimulus value. Using this logic in a monkey electrophysiology experiment, Roesch and Olson found that activity in OFC reflected the stimulus value, whereas activity in premotor cortex reflected an arousal-type variable. Second, similar to the case of arousal, exposure to stimuli with very positive or very negative stimulus values might induce an overall increase in attention. Third, in many choice paradigms, goal values and reward prediction errors are positively correlated (even if the design includes both appetitive and aversive items). Hare and colleagues (2008) show that prediction errors and stimulus values can be dissociated by introducing a random monetary prize in every trial that is independent of the choices made by the subjects. Using this experimental trick, they found that BOLD activity in the medial OFC, but not ventral striatum, was correlated with the stimulus values, whereas activity in the ventral striatum was most consistent with the prediction error signal.

Why is medial OFC involved in the computation of stimulus values? Some authors have argued that this area of the prefrontal cortex might be in a unique position to integrate information about stimuli and states of the word into a value

(Schoenbaum, Roesch, & Stalnaker, 2006; Wallis, 2007). This favorable position is due to its multiple connections with limbic areas such as the thalamus, amygdala, and striatum (Carmichael & Price, 1996; Ongur & Price, 2000).

So far we have focused on the neural basis of the stimulus-value signal. This signal represents only the output of the valuation process. As described in the previous section, these values are constructed by either retrieving or computing stimulus-outcome associations [the $p(o|s)$ functions], by retrieving or computing the value associated with each of those outcomes in the current state of the world [the $v(o|e)$ function], and by integrating them into an expected value signal [the $V(s|e)$ stimulus value]. This analysis gives rise to the following important questions: How and where are the stimulus-outcome associations represented? How and where is the $v(o|e)$ valuation function represented, and how does the state of the world modulate its value? How and where are the two of them integrated into the stimulus value signal? The answers to these questions are largely unknown and constitute one of the most important open problems in neuroeconomics.

Stimulus Choice Although several proposals have been made about how the brain compares options in simple stimulus-choice situations (Glimcher et al., 2005; Wallis, 2007), next to nothing is known about this is actually done. Understanding how the goal-directed systems compare the stimulus values to make a choice is another important open problem in neuroeconomics. Other open questions include the following: Does the brain make choices by implementing a race-to-barrier model? If so, is the choice made over actions or stimuli? How are the barriers chosen and implemented? How does the slope of the integrators relate to the strength of the stimulus-value signal encoded in medial OFC? Are there other inputs to the comparison process besides the medial OFC signal? How and where does the brain incorporate information about the cost of acquiring the different stimuli? How does the system go from stimulus choices to motor responses?

Learning There is a large literature in neuroeconomics showing that reward prediction errors are encoded in the ventral striatum in the context of Pavlovian (nonchoice) and habitual choice paradigms (for a comprehensive review see Niv & Montague, 2008). Unfortunately, this literature is not very informative about the learning that takes place in the goal-directed system during the simple stimulus-choice task. The reason is that the prediction errors required here measure how surprising the occurrence of individual outcomes is, as opposed to prediction errors of reward that measure the amount of unexpected reward received at the time of consumption. These are two very different types of learning and are likely to be implemented by different networks. Understanding the computational and neurobiological basis of how the goal-directed system learns stimulus-outcome associations is another important open question for neuroeconomics.

Conclusions

The goal-directed system provides organisms with a flexible and adaptive tool to make decisions. This is based on its ability to assign values to stimuli based on beliefs about the outcomes that they are likely to generate and the value of those outcomes in the current state of the world. The system might be particularly powerful through its interactions with other higher cognitive processes that might allow it to use analytical and memory processes to improve its characterization of the stimulus-outcome associations. Given that the system is thought to play a large role in human decision making, understanding its computational and neurobiological basis is central to understanding the essence of human nature.

This review has emphasized the use of simple mathematical models to describe the computations that the goal-directed system needs to carry out in order to make a choice. These models are useful because they lay down precise descriptions of the computational nature of the problem ("what needs to be encoded") and guide the search for the neural instantiation of the process at work. We believe that the use of these types of models is critical to the rapid advancement of the field.

REFERENCES

Arana, F. S., Parkinson, J. A., Hinton, E., Holland, A. J., Owen, A. M., & Roberts, A. C. (2003). Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. *J. Neurosci.*, *23*(29), 9632–9638.

Balleine, B. W., Daw, N., & O'Doherty, J. (2008). Multiple forms of value learning and the function of dopamine. In P. W. Glimcher, E. Fehr, C. Camerer, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision-making and the brain*. New York: Elsevier.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology*, *37*(4–5), 407–419.

Baxter, M. G., & Murray, E. A. (2002). The amygdala and reward. *Nat. Rev. Neurosci.*, *3*(7), 563–573.

Bogacz, R. (2007). Optimal decision-making theories: Linking neurobiology with behaviour. *Trends Cogn. Sci.*, *11*(3), 118–125.

Busemeyer, J. R., & Johnson, J. G. (2004). Computational models of decision making. In D. Koehler & N. Narvey (Eds.), *Handbook of judgment and decision making* (pp. 133–154). New York: Blackwell.

Carmichael, S. T., & Price, J. L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *J. Comp. Neurol.*, *371*, 179–207.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, *8*(12), 1704–1711.

Dayan, P. (2008). The role of value systems in decision making. In C. Engel & W. Singer (Eds.), *Better than conscious? Implications for performance and institutional analysis*. Cambridge, MA: MIT Press.

Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, *36*(2), 285–298.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehavior of value and the discipline of the will. *Neural Netw.*, *19*(8), 1153–1160.

Dayan, P., & Seymour, B. (2008). Values and Actions in Aversion. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain*. New York: Elsevier.

Dickison, A., & Balleine, B. W. (2002). The role of learning in the operation of motivational systems. In C. Gallistel (Ed.), *Learning, motivation & emotion*, vol. 3 of *Stevens' handbook of experimental psychology* (3rd ed., pp. 497–533). New York: John Wiley & Sons.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Netw.*, *19*(8), 981–1012.

Erk, S., Spitzer, M., Wunderlich, A. P., Galley, L., & Walter, H. (2002). Cultural objects modulate reward circuitry. *Neuroreport*, *13*(18), 2499–2503.

Glimcher, P. W., Dorris, M. C., & Bayer, H. M. (2005). Physiological utility theory and the neuroeconomics of choice. *Games Econ. Behav.*, *52*, 213–256.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*, 535–574.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2002). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. *J. Neurosci.*, *22*(24), 10829–10837.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*, *301*(5636), 1104–1107.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J. Neurosci.*, *26*(32), 8360–8367.

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.*, *28*(22), 5623–5630.

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.*, *24*(34), 7540–7548.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nat. Neurosci.*, *10*(12), 1625–1633.

Karjbich, I., Armel, C., & Rangel, A. (2008). *The role of visual attention in the computation and comparison of economic values*. Unpublished manuscript, Pasadena.

3 Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.*, *16*(5), 1936–1947.

Niv, Y., & Montague, P. R. (2008). Theoretical and empirical studies of learning. In P. W. Glimcher, E. Fehr, C. Camerer, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision-making and the brain*. New York: Elsevier.

Nobre, A. C., Coull, J. T., Frith, C. D., & Mesulam, M. M. (1999). Orbitofrontal cortex is activated during breaches of expectation in tasks of visual attention. *Nat. Neurosci.*, *2*(1), 11–12.

O'Doherty, J. P., Deichmann, R., Critchley, H., & Dolan, R. J. (2002). Neural responses during anticipation of a primary taste reward. *Neuron*, *33*, 815–826.

Ongur, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex*, *10*(3), 206–219.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226.

Padoa-Schioppa, C., & Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.*, *11*(1), 95–102.

Paulus, M. P., & Frank, L. R. (2003). Ventromedial prefrontal cortex activation is critical for preference judgments. *Neuroreport*, *14*(10), 1311–1315.

Plassmann, H., O'Doherty, J., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.*, *27*(37), 9984–9988.

Plassmann, H., O'Doherty, J., & Rangel, A. (2008). *Aversive goal* 4
values are negatively encoded in the medial orbitofrontal cortex at the time of decision making. Unpublished manuscript.

Rangel, A. (2008). The computation and comparison of value in goal-directed choice. In P. W. Glimcher, C. F. Camerer, E. Fehr, & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain*. New York: Elsevier.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.*, *9*(7), 545–556.

Reutskaja, E., Pulst-Korenhberg, J., Nagel, R. M., Camerer, 5
C. F., & Rangel, A. (under review). The brain uses a random search and pairwise comparison algorithm to make multi-item choices under time pressure.

Roesch, M. R., & Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science*, *304*(5668), 307–310.

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat. Neurosci.*, *1*(2), 155–159.

Schoenbaum, G., Roesch, M. R., & Stalnaker, T. A. (2006). Orbitofrontal cortex, decision-making and drug addiction. *Trends Neurosci.*, *29*(2), 116–124.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*, 515–518.

Valentin, V. V., Dickinson, A., & O'Doherty, J. P. (2007). Determining the neural substrates of goal-directed learning in the human brain. *J. Neurosci.*, *27*(15), 4019–4026.

Wallis, J. D. (2007). Orbitofrontal cortex and its contribution to decision-making. *Annu. Rev. Neurosci.*, *30*, 31–56.

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.*, *18*(7), 2069–2081.

Wellman, L. L., Gale, K., & Malkova, L. (2005). GABAa-mediated inhibition of basolateral amygdala blocks reward devaluation in macaques. *J. Neurosci.*, *25*(18), 4577–4586.

6