Brain Circuitry for Social Decision Making in Non-Human Primates

Daeyeol Lee and Michael C. Dorris

INTRODUCTION

Experimental studies of decision making have, for the most part, examined choices with clearly defined probabilities and outcomes in which the decision maker selects between options that have consequences only for them. This is reflected in the fact that the canonical decision tasks involve choices between monetary gambles – for example, participants might be asked whether they prefer a 50% chance of $25, or a certain $10. Though the outcomes and likelihoods are often complex and uncertain, and sometimes ambiguous, these decisions are typically made in socially isolated settings.

However, in our daily life decisions are seldom made in these sterile situations, and indeed many of our everyday decisions and choices are made in the context of a social interaction. We live, work, and play in highly complex social environments, and the decisions we make are often dependent on the concomitant decisions of others, for example, when we are deciding to extend an offer of employment or when we are entering a business negotiation. These decisions have the potential to offer a useful window into more complex forms of decision making, decisions that approximate many of the more interesting choices we make in real-life. This chapter examines a neuroeconomic approach to study this problem in non-human

Neuroeconomics. DOI: http://dx.doi.org/10.1016/B978-0-12-416008-8.00026-7

primates, that is, by directly measuring or manipulating neural signals in monkeys who are engaged in such social decisions.

The nature of decision making may change fundamentally when the outcome of a decision is dependent on the decisions of others, an issue also taken up in Chapters 2, 11, 25, and 27 of this book. Under these kinds of conditions, the standard expected utility computation that underlies many of the existing theories and models of decision making described in Chapter 1 is complicated by the fact that we must also attempt to infer the beliefs of our partner or opponent in attempting to reach the optimal decision, as noted in Chapter 2.

As part of the neuroeconomic approach, several groups of researchers have begun to investigate the psychological and neural correlates of relatively simple social decisions using tasks derived from a branch of experimental economics that focuses on game theory. These tasks, though simple, may require sophisticated reasoning about the motivations of other players in the task. The combination of these tasks and modern neuroscientific methods have the potential to greatly extend our knowledge of both the brain mechanisms involved in social decision making, as well as advancing the theoretical models of how we make decisions in a rich, social environment.

This chapter focuses on the use of invasive electrophysiological techniques in monkeys for studying decision-making processes during game-theoretic tasks. The biggest advantage of this approach is that it allows us to directly measure and manipulate neural signals and circuits with exquisite spatial and temporal resolution during the actual decision-making process. For a number of technical reasons that will be discussed below, this approach has been limited to simple iterative games such as *rock-paper-scissors*. These simpler games are ideal for examining neural processes involved in representing reward, probability, subjective value, choice selection and adaptive learning. The reader is directed to Chapters 11 and 25 that adopt the complimentary neuroeconomic approach of brain imaging in humans during games. The advantage with human brain imaging is that it examines the decision processes in the species we are most interested in, ourselves. Also, more sophisticated games can be employed in humans to examine social preferences and related concepts, such as fairness, reciprocity, and trust that play important roles in challenging social situations. The currently available neuroimaging techniques, however, lack the spatial and temporal richness of direct neurophysiological measurements. Together, these human and non-human primate approaches are providing us with unparalleled access to the process within the "black-box" and the promise of a deeper understanding of how social animals successfully (and sometimes unsuccessfully) interact.

GAME THEORY

In essence, game theory is a collection of rigorous models aimed at explaining situations in which decision makers must interact with one another, and it is the focus of Chapter 2 in this volume. In classical game theory (e.g., von Neumann and Morgenstern, 1944), it is assumed that decision makers have full knowledge not only about each of the alternative actions they can choose, but also know about how the payoff is determined jointly by their actions and actions of other decision makers. The concept of an *equilibrium* plays a central role in understanding these interactions. For example, a *set of strategies* is referred to as a Nash equilibrium (Nash, 1950; also see Chapter 2) when no individual players can increase their payoffs by deviating from such strategies unilaterally. For example, if both players in a game of rock-paper-scissors were choosing between the three options unpredictably and in equal proportions (a *mixed-strategy*) they would be at the Nash equilibrium because neither would have an incentive to change their strategy, conditional upon their belief that their opponent is also behaving rationally in this regard. Such game theoretic equilibria would be accurate models of human or animal decision making, however, only to the extent that real decision makers are capable of making all the inferences necessary to identify and implement such equilibrium strategies. In fact, when the behaviors of humans and animals during various games are systematically studied in laboratory experiments, the results often display similar systematic deviations from the predictions of equilibrium strategies (a point taken up in the preceding chapter and in Camerer, 2003). Typically, decision makers are both less selfish and more willing to consider factors such as reciprocity and equity (Chapter 11), than the classical game theory might predict. In addition, when the same game is played repeatedly, decision makers tend to adjust their strategies gradually to improve the outcomes of their choices. In fact, humans and monkeys display similar dynamics in their choice behaviors during iterative games (Lee, 2008), and in a way that is often not captured by classical game theory. It is also important, however, for the reader to keep in mind that despite these strategic similarities between species, it is unclear whether monkeys performing experiments in laboratories truly understand that they are engaged in a strategic game because they often do not face a live opponent in the laboratory, nor can the researchers provide them with verbal instruction or receive self-reports from the monkeys.

Nonetheless, the well-characterized tasks and formal modeling approach offered by game theory provides a useful foundation for the study of decisions in a social context. From an experimental standpoint, the mathematical framework of game theory provides a common language in which findings from different research groups, and indeed research methodologies, can be compared, and deviations from model predictions quantified. These tasks produce a surprisingly varied and rich pattern of decision making, while employing quite simple rules. The rules for the three iterative, repeated games that have been studied in monkeys – *matching pennies*, the *inspection game*, and *rock-paper-scissors* – are shown in normal form in Figure 26.1. As we describe the results for each below, we will examine the manner in which these tasks have been adapted for neurophysiological experiments in awake, behaving monkeys.

The benefits of combining game theoretic tasks with systems neuroscience techniques, such as single-neuron recording, are twofold. First, as described above, choice patterns in these tasks often do not conform precisely to the predictions of classical game theory, and therefore more precise characterizations of behavior, in terms of the neural process that underlie them, will be important in adapting these models to better fit how decisions are actually made. Second, neuroscience can provide important biological constraints on the processes involved, and indeed research is revealing that many of the processes thought to

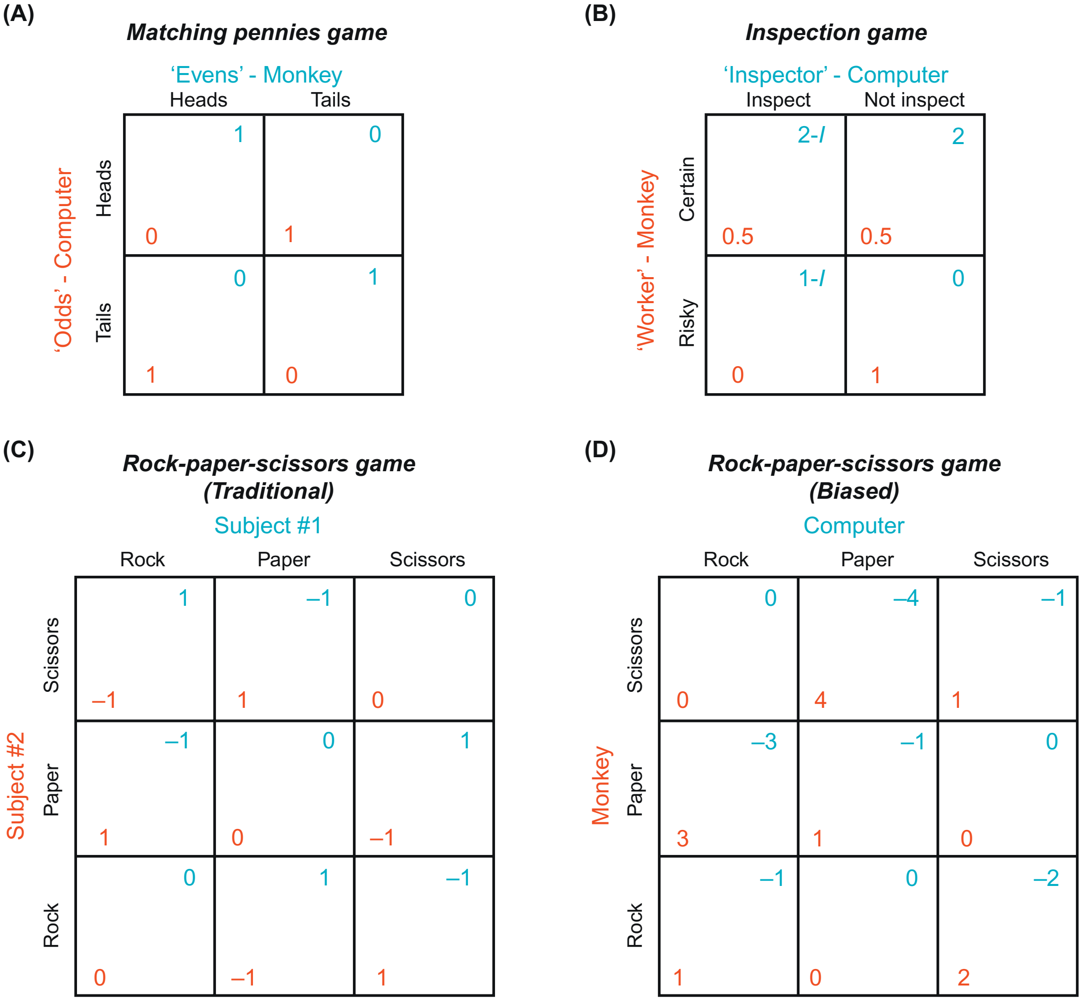

FIGURE 26.1 Payoff matrices for mixed-strategy games used in neurophysiological experiments in non-human primates. Matching pennies task (A), inspection game task (B), traditional rock-paper-scissors task (C), and the biased rock-paper-scissors task (D). The red numbers in each cell refer to the units of liquid reward received by the monkey (or the "virtual" units for the blue numbers in the case of the computer opponent). For the inspection game task (B), *I* refers to the "cost of inspection" for the computer opponent and it ranged from 0.1 to 0.9 across blocks of trials in 0.2 step increments.

underlie this type of complex decision making may overlap strongly with more fundamental brain processes such as reward, punishment and disgust. Knowledge of the "building blocks" of decision making in games will greatly assist in constructing better models of this process.

PRIMATE VISUO-SACCADIC CIRCUITRY AS A MODEL SYSTEM FOR STUDYING THE NEURAL BASIS OF SOCIAL DECISION MAKING

Although use of awake, behaving monkeys has been a mainstay of systems neuroscience research for over 40 years, their use in conjunction with game-theoretic tasks is less than 10 years old. Though still in its infancy, this research has already produced significant insights into the hidden processes that occur within the so-called "black-box" during social interactions. Here we outline the general neurophysiological techniques for the non-neuroscientist and highlight their promise and limitations in providing future insights.

A suitable animal model is required to permit direct access to the neural substrate during decision making in games play. For a number of reasons, the rhesus monkey (*Macaca mulatta*) has been the primary animal model for studying higher-order decision processes. The general organization of their nervous system is similar to that of humans and this complexity allows these non-human primates to learn relatively sophisticated behavioral tasks in the laboratory. The suitability of these non-human primates likely extends into the social domain, a point taken up in Chapter 7. Both species have well-established hierarchical social structures with sophisticated signaling systems for maintaining this structure (Byrne and Whiten, 1989; de Waal, 1990). Across a number of decision-making contexts, including that of mixed-strategy games on which we focus here, monkeys and humans display apparently comparable strategies, suggesting that many of the underlying neural processes may be shared.

For a number of practical reasons, decision-making research in animals has focused primarily, but not exclusively (Kalaska *et al*., 2003; Romo and Salinas, 2003), on the monkey visuo-saccadic system (Glimcher, 2003; Schall and Thompson, 1999). The primate visuo-saccadic system is of critical importance because it allows us to efficiently extract visual information from our environment. This is achieved by aligning the foveae – the central portion of the retinae associated with the highest visual acuity – with targets of interest using ballistic eye movements known as saccades, followed by stable fixation, when visual information is acquired and processed in extra-striate visual areas. Although not traditionally considered "choices," saccades are in fact the behavioral read-out of one of our most common decisions, that of choosing when and where to look.

The neural circuitry underlying visual processing and saccadic control is well understood, which provides a solid foundation for asking questions about the decision processes that link sensation to action. Saccades are particularly simple and stereotyped movements and, unlike other sensory-motor systems, all the circuitry is housed entirely within the head. This last point is important because the head can be restrained from moving during experiments thus providing the stability necessary for recording tiny neurons within an awake and otherwise moving animal. To do so, monkeys are trained to sit in specialized head-restraining chairs while performing experiments. Consequently, social interactions within the neuroscience laboratory have involved directing saccades to visual targets controlled by virtual computer opponents (Figure 26.2) rather than direct, rough-and-tumble interactions between monkeys. Comparable restrictions are incurred when conducting experiments on social decision making within scanners in human experiments (see Chapter 6).

Advantages and Disadvantages of a Systems Neurophysiology Approach

The advantages of the systems neurophysiology approach stem from the direct access to the neural substrate that it seeks to characterize. Neuronal signals can be sampled with exquisite temporal (<1 ms) and spatial (individual neurons) resolution and, with nearly comparable precision, neuronal activity can also be artificially manipulated.

For those not familiar with the methodology associated with neurophysiology in awake, behaving monkeys, we outline it briefly. It is treated in greater detail in Chapter 6. A chamber with a removable cap is fixed over a small opening in the skull and cleaned daily under antiseptic conditions. At the onset of each experiment, a fine metal electrode or needle pierces the membrane (dura) which covers the brain and, with high precision, is slowly lowered to the brain region of interest. These procedures are painless and cause little damage to neural tissue because the brain lacks pain receptors and only very thin probes are used. These latter properties are critical because to obtain accurate experimental results both the animal and brain must be in as natural a state as possible.

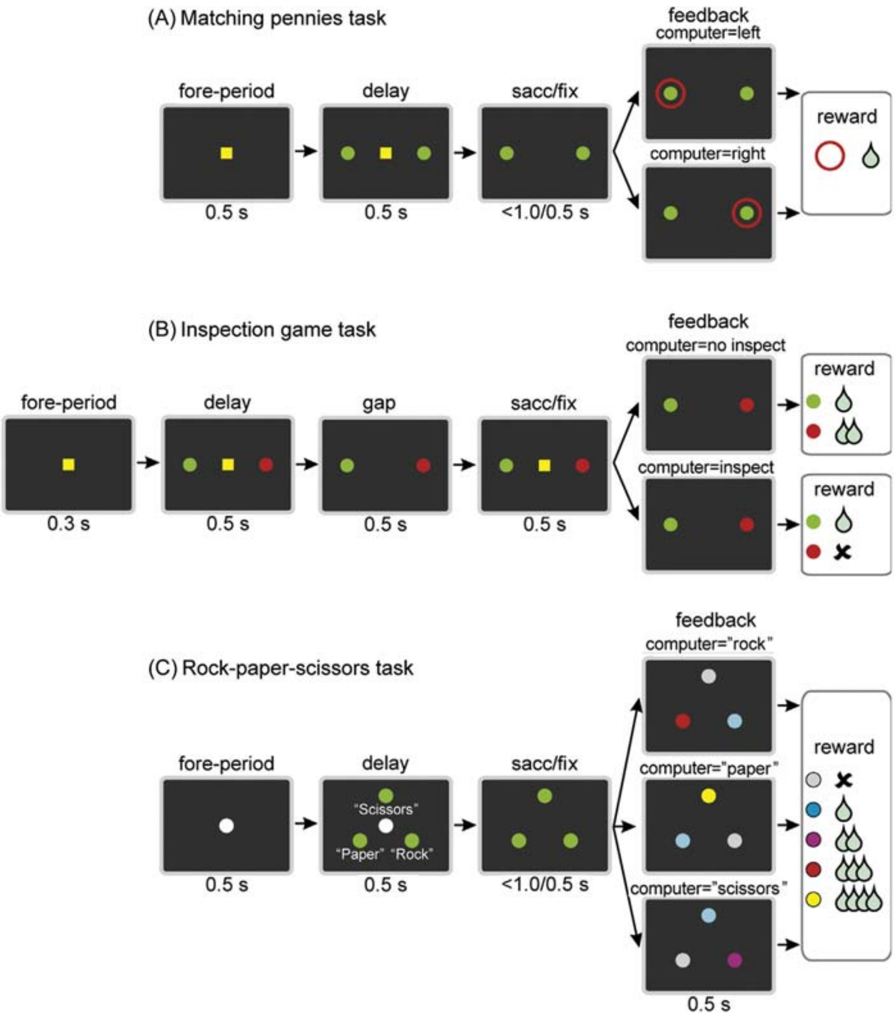

FIGURE 26.2 Matching pennies task (A) inspection game task (B) and rock-paper-scissors task (C) used to study the choice behaviors in monkeys. (A) During the matching pennies task, the animals indicated their choices by shifting their gaze towards one of the two peripheral targets. During the feedback period, the target chosen by the computer opponent was indicated by a red ring, and the animal was rewarded only when it chose the same target as the computer. (B) During the inspection game task, the animals indicated their choices after the gap period by shifting gaze towards the risky red target in the neuron's response field or the certain green target opposite the neuron's response field. The monkey always received 1 unit of juice for choosing the certain target. When the monkey chose the risky target, the monkey received 2 units of juice if the computer opponent did not inspect and zero units of juice if the computer opponent inspected. (C) During the rock-paper-scissors task, the animals were required to shift their gaze towards one of three peripheral targets. During the feedback period, the actual outcome from the chosen target and fictive outcomes from the other unchosen targets, which were determined by a biased rock-paper-scissors game, were indicated by different colors.

It is the action potentials, or electrical pulses originating in one neuron and propagating along axons described in Chapter 5, which are recorded with microelectrodes during these experiments. This neuronal activity can be correlated with features of the sensory instructions, internal variables predicted by economic theory, aspects of the choice response, and the type of reinforcement. Because this neural activation can be measured with millisecond precision, it is the best means for understanding the moment-to-moment computations underlying the decision process.

Artificial manipulation of neuronal activity can provide more direct evidence that a brain region is causally involved in the decision process, complimenting the correlational evidence provided by neuronal recordings. A number of techniques for manipulating neuronal activity exist as described in Chapter 6. This chapter describes the artificial excitation of neuronal activity through electrical micro-stimulation. The temporal precision, spatial extent, and intensity of neuronal activity manipulation can be controlled more precisely than with the techniques currently available for reversibly inactivating regions of the brain.

A number of potential disadvantages exist in using non-human primates to infer the neural processes underlying human social interactions, however. To date, non-human primates have only been trained to perform simple mixed-strategy games during neurophysiological recordings. The reader should refer to Chapters 2, 7, 11 ,and 25 for a discussion of other forms of social interactions and games that non-human primates have been trained to perform, and which will surely be examined with neurophysiological techniques in the near future. Much of the challenge in using non-human primates is assessing whether they share key cognitive abilities with us necessary to perform complex social interactions and, if so, distilling these abstract tasks into formats that monkeys can understand. Moreover, it may be difficult to train animals on game-theoretic tasks without verbal instructions, using only operant conditioning techniques. Even if comparable choice strategies are used during experimental games, we must remember that this is a prerequisite, not proof, that the same neural mechanisms are shared in these two species. That being said, monkeys and humans have displayed remarkably similar strategies under the simple mixed strategy games studied to date (Barraclough *et al.*, 2004; Dorris and Glimcher, 2004; Lee *et al.*, 2004, 2005; Thevarajah *et al.*, 2009, 2010). Although the limits of this animal model have yet to be determined, understanding the neural mechanisms underlying decision making during games in monkeys is important because these may be directly related to our own decision-making mechanisms or, at the very least, they represent the core mechanisms upon which our more sophisticated decision processes rest.

Adapting Games for Non-Human Primates

Neurophysiologists have initially focused their efforts on simple mixed-strategy games primarily because non-human primates can be trained relatively easily on these tasks. We next briefly describe some of these games, and how they have been modified for the neurophysiology laboratory (Figure 26.2).

Nearly all tasks studied to date involve thirsty animals competing against dynamic computer opponents for liquid rewards. Monkeys sit in front of visual displays and indicate their choices by looking to one of several potential choice targets followed by visual feedback on the choice of the computer opponent. At the onset of each experiment, a microelectrode is manipulated, moved back and forth, until the experimenter succeeds in isolating the activity of a single neuron from the background of general brain activity.

A critical concept needed to interpret neurophysiological findings is that of the neuronal response field. In many brain areas involved in vision and eye movement control, each neuron is activated by a particular combination of sensory and motor attributes that together define the neuron's response field. In some structures, such as the visual cortex or the superior colliculus, populations of neurons with similar response fields are organized together into topographic maps of sensory and motor space (Figure 26.3). Sensory attributes include the spatial location of visual stimuli relative to the foveae, the speed and direction of motion, color and shape. Movement-related, or *motor*, attributes include the direction and amplitude of the saccadic eye movement and the timing of the saccadic response. If neurons within a given brain region have response fields with defined sensory and motor attributes, the experimenter determines the neuron's response field properties and tailors the choice targets to engage the neuron under study.

Response fields in various brain regions are further elaborated in two ways that are relevant to the decision-making process. First, response field

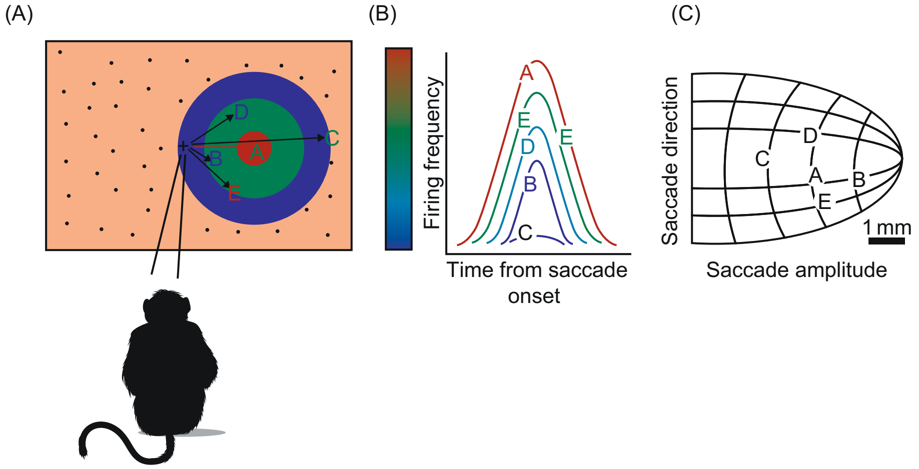

FIGURE 26.3 Example of a response field from an individual neuron and population level topographic maps in brain areas with visual and saccadic responses. Saccadic responses within the superior colliculus (SC) are illustrated here but the principle is similar for visual aligned responses in more visual areas. (A) Once activity from a single neuron is isolated from background noise, the monkey makes saccades from a central location (cross) towards targets placed throughout the visual field (black dots). (B) The amount of neuronal activity elicited by this single neuron is shown for the five saccadic vectors highlighted in panel A. From these vector-associated activities a heat-map is constructed illustrating the neuronal response field. The saccade vector associated with target A is considered the center of this neuron's response field because it elicits the highest firing frequency. (C) The neurons within the intermediate layers of the SC are organized as a topographic map of saccade vectors. The populations of neurons with the highest activation are shown in the left SC associated with the five rightward saccades shown in panel A.

properties evolve as we move from sensory to motor related brain regions; in the sensory cortical areas, response fields encode stimulus properties largely irrespective of the movements or decisions made by the subject and, later on, in motor areas, response fields encode properties of the movements largely irrespective of incoming sensory attributes. This transformation has been well characterized by decades of neuroscience research. Second, response field activation in many brain areas is shaped by cognitive and economic factors even when immediate sensory and motor attributes are fixed. These modulatory processes result from interactions with other brain regions that lack classical response fields such as much of the frontal cortex and parts of the basal ganglia. Neuronal responses from these modulatory regions tend to be heterogeneous and only weakly tuned to sensory and motor attributes of the task. Below we outline neuroeconomic approaches to determine how economic variables such as choices, reward, subjective value and beliefs are represented in higher cortical regions to extend classical sensory-motor response fields to select appropriate social actions.

In a typical neuroeconomic experiment in monkeys, each game *trial*, or *round*, begins with the animal fixating a central visual stimulus. The animal indicates its choice by directing a saccade to one of the peripheral targets upon their presentation, or after a short delay. Whether the animal receives a liquid reward (or its amount, type, etc.) depends on both their own choice and that of the computer opponent. Although computer algorithms vary in their details across studies, during competitive mixed-strategy games all look for patterns in the animal's history of choices and rewards in an effort to predict and counter the animal's upcoming actions so that they can approximate a more natural and biological opponent.

Monkeys have been trained to perform simple zero-sum games such as matching pennies (Barraclough *et al*., 2004; Lee *et al*., 2004; Thevarajah *et al*. 2009, 2010) and rock-paper-scissors (Abe and Lee, 2011; Lee *et al*., 2005) and non zero-sum games such as the inspection game (Dorris and Glimcher, 2004). Another successful means for studying adaptive decision making in nonhuman primates uses foraging tasks that produce results consistent with Herrnstein's matching law. During these tasks, the frequency with which the animal chooses a given target tends to match the fractional reward the animal acquires from that target (Corrado *et al*., 2005; Herrnstein 1961; Lau and Glimcher, 2005; Sugrue *et al*., 2004). Because matching law tasks do not involve interaction with a strategic opponent they technically are not games. Nevertheless, we mention them here because it is unclear whether monkeys can distinguish between these two classes of adaptive tasks. Indeed, matching law experiments and traditional games may well be more similar than has been widely appreciated.

Statistical Analyses of Iterative, Repeated Game Data

A final advantage of studying social decision making in awake-behaving monkeys is that a wealth of choice and neuronal data can be collected from each experimental session. Typically, monkeys will play many hundreds, if not more than a thousand, repeated trials during a single experimental session. This is advantageous for a number of reasons. First, neural signals and choice sequences are highly stochastic, so large data sets are extremely valuable for developing a more accurate representation of a given neuron's contribution to an overall choice strategy. Second, having long sequences of both neuronal signals and choice patterns allows researchers to examine how the history of previous choices and their outcomes affect processing on the current trial. It is particularly important to keep track of such factors as one's own choices and their outcomes, your opponent's choices, and overall reward rate during social decision making. These are critical both for providing accurate estimates of the subjective value of the options to guide the current choice but also are integral to the learning process and adapting to dynamic opponents and conditions. Lastly, such large data sets allow us to perform rigorous comparisons of various statistical models for choice and neural activity. We can ask whether neurons in a particular brain region represent certain variables predicted by economic models or to determine which of the competing models provides the best description of learning, choice behavior, and neural activity.

Given the large amount of choice data that can be obtained from multiple sessions of behavioral experiments in monkeys, a number of studies have compared different learning models to gain insights about the nature of learning that takes place during repeated games. As summarized in the following sections, these studies have also begun to identify the neural signals in multiple brain areas, including the prefrontal cortex and basal ganglia that are likely to play an important role for decision making during social interactions.

REINFORCEMENT LEARNING

Reinforcement Learning in Games

When decision makers are allowed to make decisions repeatedly in a particular game and observe the outcomes of their choices as well as the choices of other players, their behaviors can be described by various learning models more accurately than by the equilibrium predictions of the classic game theory (Camerer, 2003; Camerer and Ho, 1999; Erev and Roth, 1998; Fudenberg and Levine, 1998). The models in reinforcement learning theory (Sutton and Barto, 1998) have successfully provided parsimonious explanations for a wide range of choice behaviors (see Chapters 15 and 16), including those occurring during social interactions (Lee, 2008; Lee *et al*., 2012). Reinforcement learning theory provides a large number of computational algorithms that can be used to discover successful strategies by trial and error. In contrast to the static equilibrium strategies described by traditional economic approaches, these learning models make predictions about the dynamics of trial-by-trial choice behavior. The goal of such algorithms is, of course, to maximize the sum of the future rewards that are usually discounted according to their delays. Perhaps surprisingly, these dynamic models, which seek to maximize reward, often converge towards an approximation of the Nash equilibrium under some conditions.

Algorithms in reinforcement learning theory can be divided into two different categories, depending on how the value functions are updated through experience (Sutton and Barto, 1998; see Chapters 15, 16, 17 and 21). In the simple or so-called *model-free* reinforcement learning algorithms which were the focus of Chapter 15, the value functions for a given decision maker are updated exclusively according to the actual payoffs or rewards resulting from his or her previous actions. By contrast, in the *model-based* reinforcement learning algorithms covered in Chapter 16, behaviors can be adjusted more flexibly according to the decision-maker's knowledge of his or her environment.

One area in which these kinds of models have been extended is in the domain of what rewards a decision maker would have received if he or she had chosen differently. The outcomes from such hypothetical actions are referred to as *fictive outcomes*. Analogous to the reward prediction error of traditional, model-free, reinforcement learning, the difference between fictive outcomes and the outcomes predicted from the current value functions is referred to as a *fictive reward prediction error*. In model-based reinforcement learning, such as the *experience-weighted attraction* (EWA) model of Camerer and Ho (1999), value functions are independently updated according to both real and fictive reward prediction errors simultaneously. Human neuroimaging studies have, in fact, identified signals related to fictive reward prediction errors in the striatum (Daw *et al*., 2011; Lohrenz *et al*., 2007). However, whether dopamine neurons encode fictive reward prediction errors in addition to actual reward prediction errors is not yet known. The activity of individual neurons related to fictive outcomes have, however, been identified in prefrontal cortical areas, including the anterior cingulate cortex (Hayden *et al*., 2009) and orbitofrontal cortex (Abe and Lee, 2011).

Model-Free Reinforcement Learning During *Matching Pennies* Games in Monkeys

In the classic version of the matching pennies game, each of two players chooses from two alternative options, and one of the players (matcher) wins if their two choices "match" and loses otherwise. The payoff to the other player (non-matcher) is opposite, so the sum of the two players' payoffs is zero. When two rational players participate in the matching pennies game, the Nash equilibrium is for each player to choose the two targets with equal probabilities and independently across successive trials. To test whether and how monkeys approximated optimal decision-making strategies in competitive games through experience, a number of studies have examined the choice behavior of monkeys in a computer-simulated matching pennies game (Barraclough *et al*., 2004; Cui and Andersen, 2007; Lee *et al*., 2004; Thevarajah *et al*., 2009; Figure 26.2A). During one of these neurophysiological experiments in monkeys (Barraclough *et al*., 2004), each monkey played the matching pennies game against a computer opponent. The animal was required to begin each trial by fixating a small yellow square presented in the center of a computer screen ("fore-period," Figure 26.2A). Shortly thereafter, two identical green disks were presented along the horizontal meridian, and the animal was required to shift its gaze towards one of the targets when the central fixation target was extinguished. The computer opponent also chose one of these two targets – although that was invisible to the monkey – according to a prespecified algorithm described below. The animal was rewarded only when it chose the same target as the computer.

To investigate how the animal's choice behavior would be affected by the strategy of its opponent, the strategy of the computer opponent was systematically manipulated in a series of experiments by Lee and colleagues (2004). Initially, for several days, the computer opponent chose the two targets with equal probabilities regardless of the animal's choices. This was referred to as *algorithm 0*, and corresponds to the Nash equilibrium strategy of the matching pennies game at the static equilibrium. In this case, then the computer played this static equilibrium, the animal's expected payoff was fixed regardless of what it chose. All three monkeys tested with algorithm 0 displayed a strong bias to choose one of the two targets more frequently (Lee *et al.*, 2004).

In the next stage of the experiment, the computer opponent applied a set of statistical tests to the monkey's choices to determine whether the animal's decisions were randomly divided between the two targets, and whether successive choices were statistically independent. If the animal showed a bias towards one target or non-independence of sequential choices, the computer used this information to adjust its choices so as to maximize the probability that it would win each round. This more dynamic approach was referred to as *algorithm 1*. Importantly, this algorithm did not examine the animal's reward history, and therefore was not sensitive to any bias that the animal might show that arose from using information about previous rewards to determine future choices. When tested with algorithm 1, monkeys chose the two targets more or less equally. In addition, the animal's successive choices were relatively independent, and as a result, the animal's overall reward rate was close to the one that would have been achieved by two players in Nash Equilibrium, a value of 0.5 (Lee *et al.*, 2004). Interestingly, the animals were more likely to choose the same target on the next trial if the choice in the previous trial was rewarded (win-stay) and switch to the other target otherwise (lose-switch). Such *win-stay-lose-switch* (WSLS) strategies were not penalized during the period of algorithm 1, since the information about the animal's reward history was not utilized by the computer opponent. All three animals chose their targets according to the WSLS strategy in substantially more than 50% of the trials.

In the final stage of these behavioral experiments on the matching pennies task, the computer opponent (*algorithm 2*) also exploited the biases in the animal's choice and reward history, including the tendency to use the WSLS strategy. When this was the case, the animals were less likely to obtain reward if they used the WSLS strategy more frequently than 50% of the trials. As expected, this reduced the probability of the WSLS strategy significantly in all animals. However, the WSLS strategy was still used more frequently than 50% in all animals, suggesting that the animals still relied on a reinforcement learning algorithm to approximate the Nash equilibrium strategy during the matching pennies task.

In reinforcement learning models, the probability of choosing a particular action is typically given by a softmax or logistic transformation of the value functions for all actions. When there are only two choices, this reduces to the following equation.

$$\mathrm{P_t(right)} = \frac{e^{\beta Q_t(\mathrm{right})}}{e^{\beta Q_t(\mathrm{right})} + e^{\beta Q_t(\mathrm{left})}} \qquad (26.1)$$

where $P_t(right)$ denotes the probability of choosing the rightward saccade in trial t and $Q_t(x)$ the value function of choosing action x ($x =$ right or left). The parameter β determines the randomness, or stochasticity, of the decision maker. The choices are completely random and unrelated to the value functions, when $\beta = 0$, and become more deterministic as β increases. In the standard model-free reinforcement learning algorithm, the value functions for the actions chosen by the decision maker (listed as the Q-terms above) are adjusted according to the following equation:

$$Q_{t+1}(x) = Q_t(x) + \alpha\,\{\mathrm{reward}_t - Q_t(x)\} \qquad (26.2)$$

where reward_t indicates the reward received by the decision maker (1 and 0 for rewarded and unrewarded trials, respectively) and α the learning rate.

Empirically, Lee and colleagues (2004) found that the choices of monkeys during the matching pennies game were relatively stochastic (Barraclough *et al.*, 2004; Lee *et al.*, 2004). The animal's choices were also well accounted for by the model-free reinforcement learning model. In addition, the fact that the probability of using the WSLS strategy decreased against the more exploitative computer opponent using algorithm 2 suggests that this might be due to a smaller learning rate. Alternatively, this could also be the result of a smaller β, which would have made the animal's choices more stochastic. The parameters estimated for the animal's behavioral data suggest that the changes in the animal's choices related to the different algorithms of the computer opponent were largely due to the changes in the learning rate (Figure 26.4A). These results suggest than when faced with a more exploitative opponent during a competitive game, animals made their choices more stochastic, perhaps by reducing their learning rates. In addition, they provide a nice example of so-called "meta-learning," in which the parameters of a learning model, such as learning rate, are optimized (Schweighofer and Doya, 2003; Soltani *et al.*, 2006).

Hybrid Learning During the *Rock-Paper-Scissors* Game in Monkeys

In the model-free reinforcement learning described above, only the value function for the action chosen by the decision maker in a given trial is updated according to the outcome of that action. In contrast, results from studies on experimental games in humans suggest that people can also adjust the value functions for unchosen actions, according to the

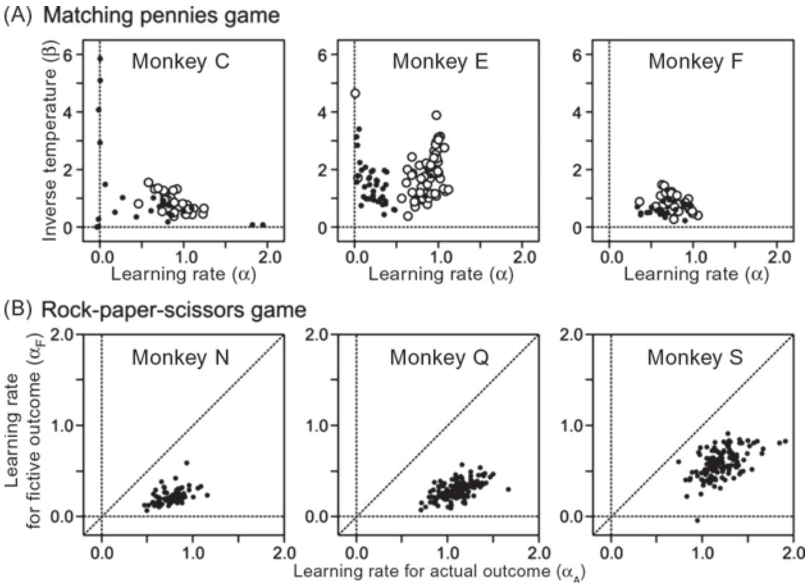

FIGURE 26.4 Behavioral performance of monkeys during the matching pennies (A) and rock-paper-scissors task (B). (A) Inverse temperatures (β) and learning rates (α) estimated from individual sessions in which the computer opponent selected its target using the algorithms 1 (empty circles) and 2 (dots) during the matching pennies task are shown for three different animals. (B) Learning rates for actual (α_A) and fictive (α_H) outcomes that were estimated using the hybrid learning model to fit the behaviors during the rock-paper-scissors task are shown for three different animals.

fictive outcomes that alternative actions would have produced, a feature of the EWA algorithm described above. Recently, it was found that monkeys can also adjust their strategies according to the fictive outcomes from unchosen actions during a rock-paper-scissors game (Abe and Lee, 2011; Lee *et al.*, 2005). Nevertheless, consistent with the results from studies in humans, the choices of monkeys were more strongly influenced by the actual outcomes from the actions chosen by the animals, than by the fictive outcomes from unchosen actions.

To demonstrate this, Abe and Lee (2011) trained monkeys to play the rock-paper-scissors game (Figure 26.1D). The monkeys first fixated a small central target at the beginning of each trial (Figure 26.2C). After a brief delay, three green disks were presented as choice targets, and the animal was free to shift its gaze towards one of these targets when the central target was extinguished. Each of these three targets was designated as rock, paper, or scissors, and whether the animal would be rewarded as a result of this choice, and the amount of juice reward given to the animal, were determined by the payoff matrix of a biased rock-paper-scissors game in competition with a computer opponent. For example, if the animal chose the "rock" target and the computer the "paper" target, then the animal did not receive any rewarded. When the result was a tie, the animal was reward with a single drop of juice. When the animal won by choosing rock, paper, and scissors, it received two, three, and four drops of juice, respectively. The payoff for the winning trial was varied so that the behavioral and neurophysiological effects of fictive outcomes could be examined quantitatively (Figure 26.1D). During this experiment, the animals were not required to memorize the rules of the rock-paper-scissors game, since the payoffs from all three choices determined by the choice of the computer opponent were visually indicated by the colors of the feedback stimuli (Figure 26.2C).

To test whether and how the animal's choices during this rock-paper-scissors game were influenced by fictive outcomes, the behavioral data obtained during this experiment were analyzed with several different learning models (Abe and Lee, 2011). This included a model-free reinforcement learning model, similar to the one described above, as well as a belief-learning model. In the belief-learning model, the players update their beliefs about the strategies of other players after each trial, and make their choices expecting to produce the best outcomes given such beliefs. This model was applied to the animal's choices during the rock-paper-scissors game by updating the value functions for all three choices according to the outcomes determined by the choice of the computer opponent. For example, when the computer selected the "rock" target, the outcome for the animal choosing rock, paper, scissors would be 1, 2, and 0, respectively, and these values were used as actual or fictive rewards to update their value functions for rock, paper, and scissors. Finally, in a hybrid-learning model, the value functions for chosen and unchosen actions were updated using two separate learning rates. Namely,

$$\begin{aligned} Q_{t+1}(x) &= Q_t(x) + \alpha_A\{actual\_reward_t - Q_t(x)\}, \\ &\quad \text{if x was chosen, and} \\ Q_{t+1}(x) &= Q_t(x) + \alpha_H\{fictive\_reward_t(x) - Q_t(x)\}, \\ &\quad \text{if x was not chosen} \end{aligned} \tag{26.3}$$

where *fictive_reward*$_t(x)$ denotes the fictive reward that could have been obtained from choosing x in trial t. In addition, α_A and α_F denote the learning rate for the actual and fictive outcomes, respectively. The results from these analyses showed that the hybrid learning model accounted for the monkey's choices during the rock-paper-scissors task better than the model-free reinforcement learning model and the belief learning model (Abe and Lee, 2011). In addition, the learning rates for the fictive outcome were always smaller than those for the actual outcomes (Figure 26.4B), indicating that actual outcomes exerted more powerful influence on the animal's subsequent choices.

CORTICAL MECHANISMS OF REINFORCEMENT LEARNING DURING ITERATIVE GAMES

Neural Activity Related to Values and Choices

One of the first areas hypothesized to be important in representing the value of visual targets in a manner that could be used to select strategic actions was the lateral intraparietal area (area LIP). Area LIP was selected for study because it is situated at the end of visual processing stream and its outputs impact regions of the brain involved in planning and executing upcoming saccades (Bisley and Goldberg, 2003; Grefkes and Fink, 2005; Pare and Wurtz, 2001). Previous work had demonstrated that activity in this region may encode the saliency of visual targets in a manner that can be used to allocate attentional resources and/or to select between upcoming saccade goals (Andersen, 1995; Goldberg *et al*., 2006). A pioneering study conducted by Platt and Glimcher (1999) demonstrated that important economic variables such as the probability and magnitude of reward impact the firing rates of LIP neurons and, in doing so, provided an alternative decision theoretic framework for studying the role of brain regions in simple sensory-to-motor transformations.

Given that area LIP lies at a nexus between sensory and motor processing and is influenced by economic variables, Dorris and Glimcher (2004) hypothesized that it could play an important role in representing the *subjective value* of choice targets, a neural correlate of economic objects like expected utility, during competitive games. In their experiment, monkeys competed against a computer opponent during the mixed-strategy inspection game (Figure 26.2B). From the monkey's perspective the target opposite the response field yielded a certain small amount of juice each time it was selected. The target in response field was "risky" in that it could pay double the certain amount or nothing. The payoff matrix was experimentally manipulated across blocks of trials so that the mixed-strategy Nash equilibrium solution for the monkey ranged from choosing the target in the center of the neuron's response field from 10–90% of the time. This equilibrium was manipulated by manipulating the computer opponent's "cost of inspection" (Figure 26.2B, variable *I*). Effectively, if *I* is low, the equilibrium shifts so the risky option is chosen infrequently, whereas if *I* is high, the equilibrium shifts so the risky option is chosen frequently. Importantly, the computer opponent's probability of inspecting remains 50% at equilibrium independent of the value of I, a core feature of game theory. Experimentally, Dorris and Glimcher (2004) found that both humans and monkeys approached the predicted equilibrium frequencies when playing this computer opponent although they tended to choose the risky option slightly too often when the cost of inspection was low. They reasoned that if LIP encoded the probability of movement, its activation would vary across blocks of trials as those movement probabilities changed. If, however, LIP encoded the subjective value (or expected utilities) of the targets, its activation should remain relatively constant as game theory suggests that this value remains constant at mixed strategy equilibrium, independent of movement probabilities. This latter interpretation is a critical feature extension of the Nash equilibrium concept presented in Chapter 2, and follows from the fact that the theoretical conclusion drawn by Nash is that subjective value (or expected utility) should be, on average, equal between the options mixed during mixed-strategy equilibrium play. LIP activity was, indeed, shaped by the subjective value of choice stimuli; firing rates varied along with changing value under forced-choice conditions (Dorris and Glimcher, 2004; Platt and Glimcher, 1999) and remained constant throughout the behavioral equilibria established during mixed-strategy conditions (Dorris and Glimcher, 2004; Figure 26.5A).

Although the Nash equilibrium concept rests on the idea that there can be no incentive to change one's overall strategy once at behavioral equilibrium (Nash, 1950), it does not specify what this means at a trial-by-trial level. The precise signals obtained from recording single neurons allow us to examine whether LIP is correlated to subjective value trial by trial as a function of the choice the subject actually made. To estimate subjective value on a trial-by-trial basis, Dorris and Glimcher (2004) optimized a simple model-free reinforcement learning algorithm, similar to ones described above and in Chapter 15. They fit the model to the monkey's pattern of behavioral choices using maximum likelihood methods in order to try and predict dynamically the monkey's pattern of choice from

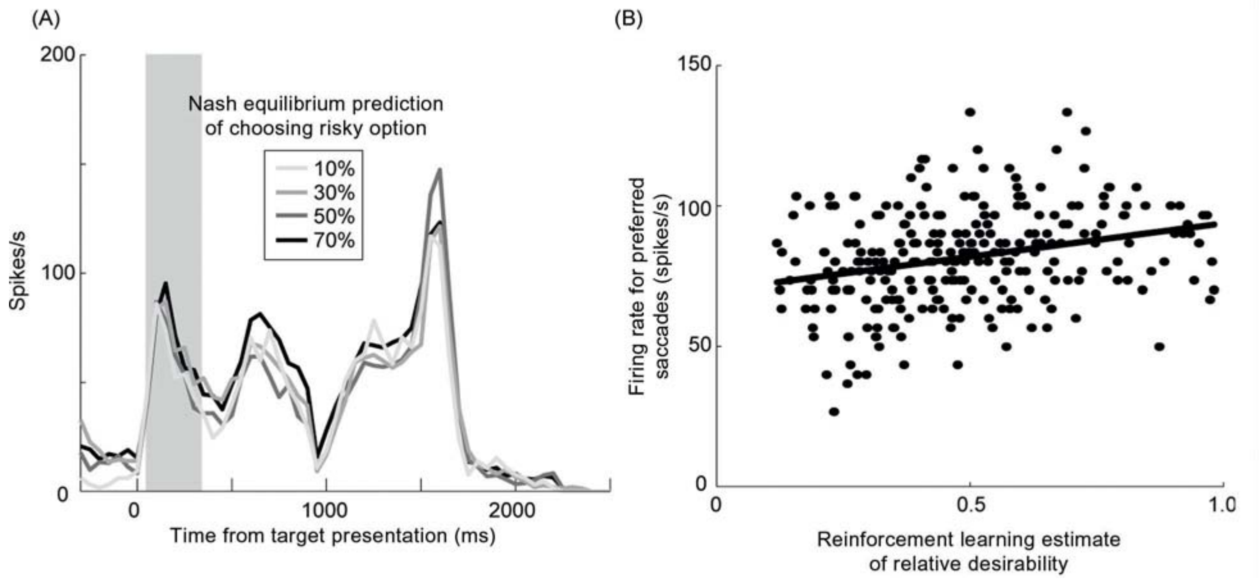

FIGURE 26.5 Encoding the subjective value of visual targets in area LIP. (A) Activity of a neuron during mixed-strategy inspection game task. Despite changes in the probability of preferred responses, LIP activity remained relatively constant which is consistent with an overall equivalency in subjective value at mixed-strategy equilibria. (B) Trial by trial variability in activity during the visual epoch was significantly correlated to a behavioral estimate of subjective value. *Adapted from* Dorris and Glimcher, (2004).

trial-to-trial. Briefly, they hypothesized subjective value of each option was incremented, if reward was received for choosing the risky option, or decremented, if reward was withheld for choosing the risky option. The only free parameter was the *learning rate* at which value was updated based on this reward information. The iterative nature of this reinforcement learning algorithm resulted in an estimate of subjective value derived from all of the subject's previous choices with the most recent choices being weighted most heavily. The authors found that trial-by-trial fluctuations in LIP activity co-varied with this trial-by-trial behavioral estimate of subjective value (Dorris and Glimcher, 2004; Figure 26.5B).

In addition to area LIP, activity that reflects both target value and saccade choices has also been identified in many other brain areas, including the prefrontal cortex. Activity of neurons in each of these regions is related to the value functions for specific actions or their transformations. Lee and colleagues demonstrated this in a series of studies in which activity was recorded from individual neurons in the dorsolateral prefrontal cortex (dlPFC; Barraclough *et al.*, 2004), the dorsal anterior cingulate cortex (ACC; Seo and Lee, 2007), and LIP (Seo *et al.*, 2009). The results from these studies showed that immediately before the animal chose its target (during the delay period, Figure 26.2A), neurons in all of these areas encode not only the animal's upcoming choice, but also the sum of the value functions for two different actions and their difference. This was demonstrated by using the following regression model to examine neuronal firing rates:

$$\begin{aligned} S_t = b_0 + b_1\,C_t + b_2\{Q_t(\text{right}) - Q_t(\text{left})\} \\ + b_3\{Q_t(\text{right}) + Q_t(\text{left})\} \end{aligned} \qquad (26.4)$$

where S_t denotes the spike rate of a given neuron during the delay period in trial t, C_t the animal's choice ($C_t = 1$ if the animal chose the rightward target and 0 otherwise), and the value function for target x in trial t, $Q_t(x)$, were estimated from the model-free reinforcement learning descried above. The difference in the value functions used in this model might be used by the animal to determine its choice, whereas their sum might be related to the *state value function* (Belova *et al.*, 2008; Cai *et al.*, 2011; Lee *et al.*, 2012; Seo and Lee 2008). The state value function corresponds to the average of action value functions weighted by the probability of taking each action, and therefore indicates the overall goodness of options faced by the animal at any given time. During the matching pennies game, for example, both actions are chosen with roughly equal probabilities, so the average of the value functions is a good estimate of the state value function. The results of this analysis revealed that signals related to the sum of the value functions are widespread in the brain at the level of single neurons (Lee and Seo, 2011; Seo and Lee, 2007, 2008; Seo *et al.*, 2009). In addition, a significant proportion of the neurons in the dlPFC and LIP, but not in the ACC, also modulated their activity according to the difference in the value functions (Seo and Lee, 2007, 2008). These results suggest that the cortical network consisting of the prefrontal and parietal areas might be important for value-based action selection during iterative competitive games (Lee *et al.*, 2012). It also seems likely that the value-related signals observed in these brain areas during matching pennies game are likely to contribute to reinforcement learning in non-social context as well, in which the subject's choices are well described by model-free reinforcement learning algorithms (Sugrue *et al.*, 2004).

Neural Activity Related to Choice and Reward Histories

The results described in the previous section suggest that the neurons in multiple cortical areas, such as the dlPFC and LIP, might play an important role in integrating the signals related to the animal's previous choices and their outcomes to update the value functions. To test this directly, Lee and colleagues applied the following regression model that includes the previous choices of the animal and computer opponent as well as the animal's choice outcomes:

$$\mathrm{S_t} = B[\, 1\ \mathrm{u_t}\ \mathrm{u}_{t-1}\ \mathrm{u_{t-2}}\ \mathrm{u_{t-3}}]' \tag{26.5}$$

where u_t is a row-vector consisting of three dummy variables corresponding to the animal's choice (0 and 1 for the leftward and rightward choices, respectively), the computer's choice (coded in the same way as the animal's choice), and the reward (1 for rewarded trials and 0 otherwise) in trial t, and B is a vector of 13 regression coefficients. Thus, the regression coefficients in this model quantify how strongly the activity of a given neuron is modulated by the current and past choices of the animal and their outcomes as well as the choices of the computer opponent. This analysis was performed separately for the spike rates measured with a series of non-overlapping 0.5-s bins defined relative to the time of target onset or feedback onset. The results showed that many neurons in the dlPFC and LIP encoded signals related to the animal's choice and its outcome as well as the computer's choice not only in the current trial, but also those in the last several trials (Seo and Lee, 2007, 2008; Seo *et al.*, 2009; Figure 26.6). The activity related to the previous choices of the computer opponent might of course be related to the value functions for alternative choices, since during the matching pennies game animals are rewarded for choosing the same target as the computer.

The signals related to the animal's previous choices might function as temporary memory signals encoding the animal's choice history. In reinforcement learning

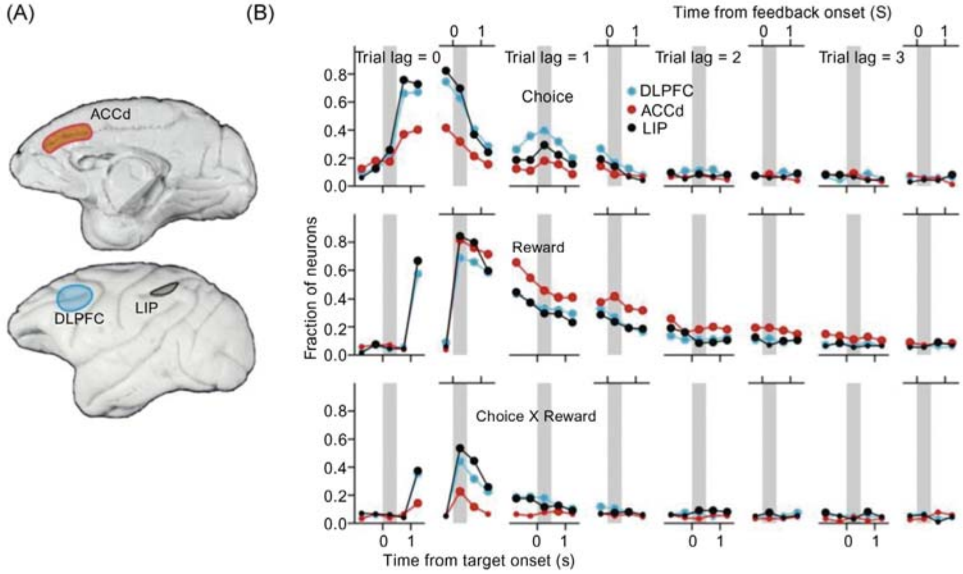

FIGURE 26.6 Cortical areas (A) and summary of neural activity (B) examined during the matching pennies task. (A) Locations of the dorsal anterior cingulate cortex (ACCd), dorsolateral prefrontal cortex (DLPFC), and lateral intraparietal area (LIP). (B) The time course of signals related to the animal's choice (top), reward (middle), and their conjunctions (bottom). Each row shows the proportion of neurons in each cortical area that significantly modulated their activity according to the animal's choice, reward, or their conjunctions (or computer's choice) in the current (trial lag = 0) and three previous (trial lag = 1, 2, 3) trials. A large symbol indicates that the effects were found in significantly more neurons than expected by chance.

theory, how delayed rewards are attributed to previous actions is referred to as the problem of *temporal credit assignment*, and the memory signals related to the animal's previous choices, often referred to as *eligibility trace*, can be used to resolve this problem, an issue discussed in Chapter 15. Thus, the neural activity related to the animal's previous choices that were found in both the dlPFC and LIP might correspond to the eligibility traces hypothesized in temporal-difference learning models.

Activity related to the animal's reward history was also found in the prefrontal cortex and posterior parietal cortex. Signals related to the reward history were particularly strong in the ACC, consistent with the idea that the medial prefrontal cortex, including the ACC, plays an important role in monitoring the outcomes of different actions. The activity related to the animal's reward history might also play an important role in computing the average rate of reward and how a particular reward deviates from the reward expected from the animal's reward history, the reward prediction error (Seo and Lee, 2007). Interestingly, the signals related to the animal's choice and reward histories found in these different cortical areas were heterogeneous and their time constants were well described by a power function, suggesting that the time constants for signals related to previous choices and outcomes might be relatively long in a small number of neurons (Bernacchia *et al*., 2011). This raises the possibility that neurons in these different cortical areas might provide a reservoir of time constants that can be selected flexibly according to the optimal time scale specific for a particular behavioral task (Beherens *et al*., 2007; see also Chapter 23).

Neural Activity Related to Fictive Outcomes

The analyses of behavioral data from the rock-paper-scissors experiment described above have shown that not only the actual outcomes of the actions chosen by the animal, but also fictive outcomes from alternative unchosen actions, influence the animal's subsequent choices. To determine whether the prefrontal cortex is involved in incorporating both actual and fictive outcomes into different value functions, the activity of individual neurons in the dlPFC and orbitofrontal cortex (OFC) was recorded in monkeys playing the rock-paper-scissors game (Abe and Lee, 2011). Consistent with findings from previous studies, the results from this study showed that neurons in both dlPFC and OFC often encode the actual outcomes from the animal's choices. The activity related to actual outcome was often seen during the feedback period in which the information about the actual outcome from the chosen target and the fictive outcomes from unchosen target were revealed to the animal (Figure 26.2C). For example, the OFC neuron illustrated in Figure 26.7A increased its activity with the magnitude of reward obtained by the animal during the feedback period of winning trials. Neurons in both dlPFC and OFC also encoded the outcomes from specific actions. For example, some neurons changed their activity according to the outcomes from choosing rock, while others modulated their activity according to the outcomes from choosing paper. This tendency was stronger in the dlPFC than in the OFC, suggesting that the dlPFC might play a more important role in updating the action value functions (Abe and Lee, 2011).

More importantly, neurons encoding fictive outcomes were also found in both dlPFC and OFC. The OFC neuron illustrated in Figure 26.7B changed its activity only slightly during the feedback period of the winning trials, but increased its activity systematically according to the magnitude of fictive reward that the animal could have earned in tie or loss trials by choosing one of the remaining targets. For some neurons in both dlPFC and OFC, the activity related to the fictive reward from the unchosen winning target changed according to the position of the winning target, and this tendency was stronger in the dlPFC than in the OFC. These results suggest that the dlPFC and OFC might play an important role in encoding not only the actual outcomes from chosen actions, but also fictive outcomes from unchosen actions, and might use those signals to update the value functions for both chosen and unchosen actions as prescribed in model-based reinforcement learning.

RESPONSE SELECTION BY THE FRONTAL EYE FIELDS AND SUPERIOR COLLICULUS

The frontoparietal areas outlined above (i.e., dlPFC, ACC, OFC, LIP) appear to represent important statistics related to social decision making ranging from the previous history of choices and their outcomes, to the evaluation of choices and their outcomes, to valuation functions and even to knowledge about "what could have been." However, it is unlikely that any of these regions ultimately selects or executes the choice response. This is evidenced by the difficulty in triggering saccades with micro-stimulation in these areas, the poor correlations of activity with saccadic reaction times and the relatively mild effects on saccade generation that result from lesion of these areas. The midbrain superior colliculus (SC) and, one of its main sources of cortical inputs, the frontal eye fields (FEF), are, by contrast, intimately involved in selecting

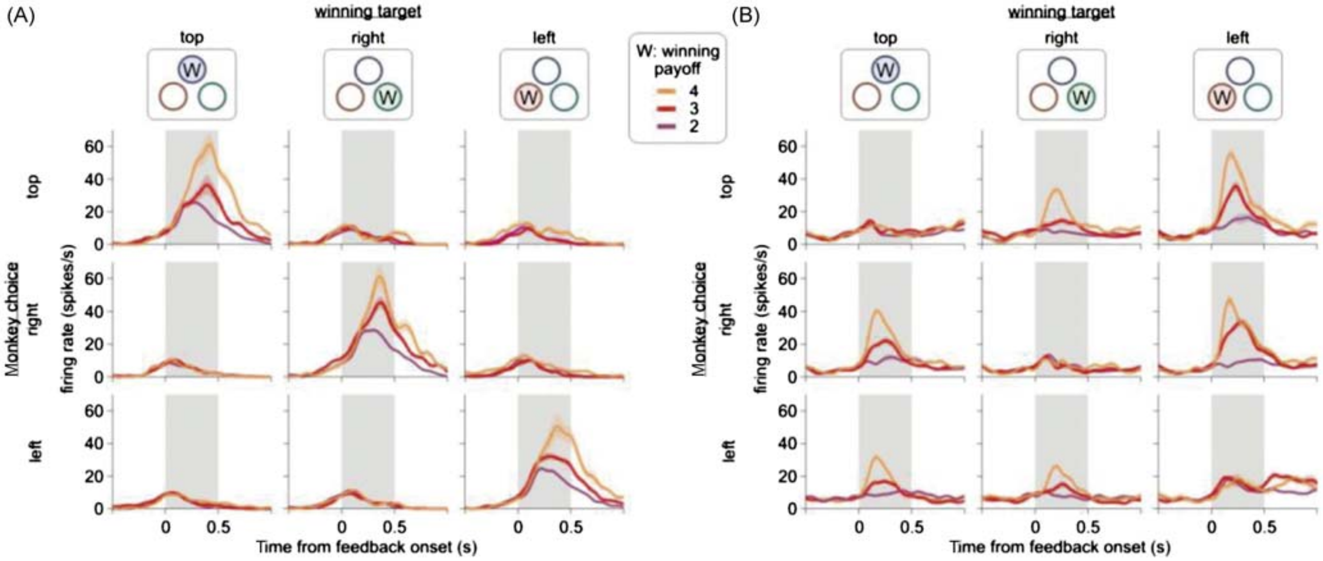

FIGURE 26.7 Example neurons recorded in the orbitofrontal cortex of monkeys that encoded actual (A) and fictive (B) outcomes during the rock-paper-scissors task. The activity of each neuron was plotted according to the animal's choice (rows), computer's choice (columns), and the payoff from the winning target (different colors). The gray background corresponds to the 0.5-s feedback period.

saccades and generating saccadic commands. Saccades are evoked with electrical micro-stimulation in these areas at low currents, activity patterns are predictive of both when and where a saccade will occur, both project to the brainstem circuits that directly control muscle forces and saccades can no longer be generated if these two structures are ablated (Dorris *et al.*, 1997; Glimcher and Sparks, 1992; Grantyn *et al.*, 2004; Robinson, 1972; Schiller *et al.*, 1980). In this section, we discuss how activity within the FEF and SC evolves to select one saccade response over another during the mixed-strategy game, matching pennies (Abunafessa and Dorris, 2011; Thevarajah *et al.*, 2009).

This matching pennies experiment borrowed the most sophisticated computer opponent from Lee and colleagues (2004), the level 2 algorithm outlined above (Figure 26.2A), with two important exceptions. First, during each experimental session the locations of the choice targets were tailored to the response field of the neuron under study. Recall that each neuron is most active for initiating saccades with a particular vector (for example a 10° saccade to the right). Once this vector was experimentally identified, one choice target was presented at that location (*inside* the response field) and the other choice target was presented at the mirror-image location relative to fixation (*opposite* the response field or 10° to left in this example). Second, a temporal warning period was introduced between the removal of the fixation point and the presentation of the choice targets. Therefore, the monkeys learned during a trial both where and when the targets would be presented. That, coupled with the requirement that a saccade choice be completed very rapidly after target presentation, encouraged choice selection during the temporal warning period. Behavioral choices were allocated to each target in equal proportions in a relatively unpredictable pattern replicating the behavioral patterns that Lee and colleagues (2004; Figure 26.4) had previously observed. Examination of SC neuronal activity revealed that one saccade becomes increasingly selected over the other as the time of target presentation approaches (Figure 26.8B). Interestingly, this neuronal selection process closely mirrors the process seen in perceptual decision making when neuronal activity accrues as a function of the quality of sensory evidence (e.g., Horwitz *et al.*, 2004). This suggests that similar principles that underlie well characterized accumulator models (see Chapters 3 and 19) apply to both perceptual and social forms of decision making. In other words, the degree to which neuronal activations segregate over time provides insight into the time course of response selection preceding strategic actions. Indeed, if the length of the warning period is changed the rate of neuronal selectivity scales accordingly (Thevarajah *et al.*, 2009).

To understand how this neuronal selection process becomes biased in favor of one of the choice targets at the level of the SC requires measuring neuronal activity from regions of the saccadic circuit that provide inputs to the SC. Abunafessa and Dorris (2011) recorded activity from the frontal eye fields (FEF) while monkeys played the matching pennies task. The

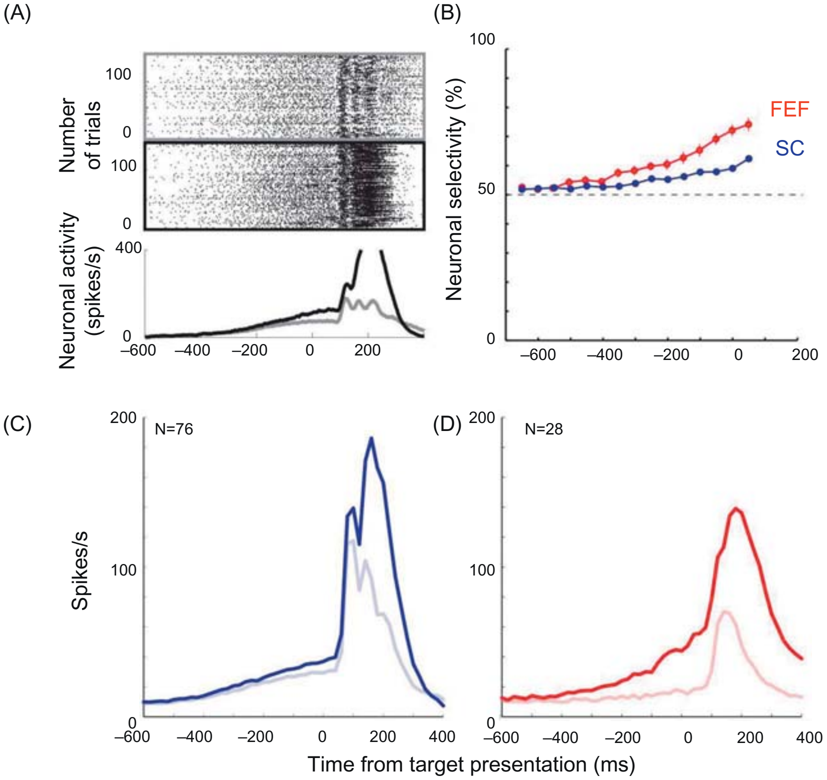

FIGURE 26.8 Neuronal selectivity of SC and FEF of upcoming mixed-strategy saccades during matching pennies task. (A) Rasters (top panels) and post-synaptic activation functions (bottom panel) are sorted based on saccades directed into (black) and opposite (gray) the neuron's response field. (B) Evolution of neuronal prediction over time. Receiver Operating Characteristic analysis for SC (N = 78 neurons; blue) and FEF (N = 28 neurons; red). Circles represent neuronal predictions based on successive 50 ms time bins throughout the warning period. (C–D) Population neuronal responses during the matching pennies task for SC (C) and FEF (D), respectively. Dark lines represent neuronal activity when the target in the neuron's response field was chosen and light lines represent neuronal activity when the target opposite the neuron's response fields was chosen.

FEF are strongly inter-connected with large portions of the frontal and parietal lobes and provide strong inputs to the SC. In addition, the FEF are particularly active during voluntary, goal-directed saccades, thus making them a likely candidate to be involved in choosing saccades during the mixed-strategy matching pennies task (Schall, 2002). Importantly, Abunafessa and Dorris recorded from the same monkeys in the same task as the SC studies, therefore, any differences in neuronal processing between the FEF and SC are unlikely to result from any differences in behavioral strategies. These authors found that neuronal selectivity occurred earlier and reached a higher overall level in the FEF than the SC, during the time leading up to the presentation of the choice targets (Figure 26.8). One might expect that neuronal selectivity would be stronger in the SC because it is closer to the ultimate motor output and integrates information across multiple frontoparietal areas described above. A possible explanation is that neuronal selectivity in the FEF reflects the ongoing decision process but, because the threshold level which neuronal activity must surpass to trigger a saccade is presumably located in the SC, this decision information is either not passed on to the SC as the decision evolves or the SC is partially inhibited prior to the presentation of the choice targets to prevent early crossing of the threshold and premature saccades.

This pre-target activity in the SC is modulated by the history of previous choices and their outcomes in a manner similar to that observed in higher cortical structures (Thevarajah *et al.*, 2010). A win-stay bias is particularly evident, that is, if a monkey chooses a saccade and it is rewarded during the matching pennies task, then on the subsequent trial, the pre-target activity in the SC representing that rewarded saccade grows at a faster rate. Faster accumulation of activity at a particular locus on the SC map is associated with a higher probability of choosing that action and faster reaction times. Interestingly, Thevarajah and colleagues (2010) found that trial-by-trial estimate of action value derived by applying the hybrid EWA model (Camerer and Ho, 1999) was correlated to trial-by-trial pre-target SC activity. This provided strong evidence that competition between neuronal populations within the brain's pre-motor structures is being regulated in a manner predicted by learning models to select strategic actions.

Direct perturbation of neural circuits has also been used in decision tasks to provide functional evidence regarding the contribution of a brain region to choice behavior (Carello and Krauzlis, 2004; Dorris *et al.*, 2007; Gold and Shadlen, 2000; Salzman *et al.*, 1990). Using a micro-stimulation paradigm adapted from Gold and Shadlen (2000, 2003; and discussed in Chapter 19), Thevarajah and colleagues (2009) tested

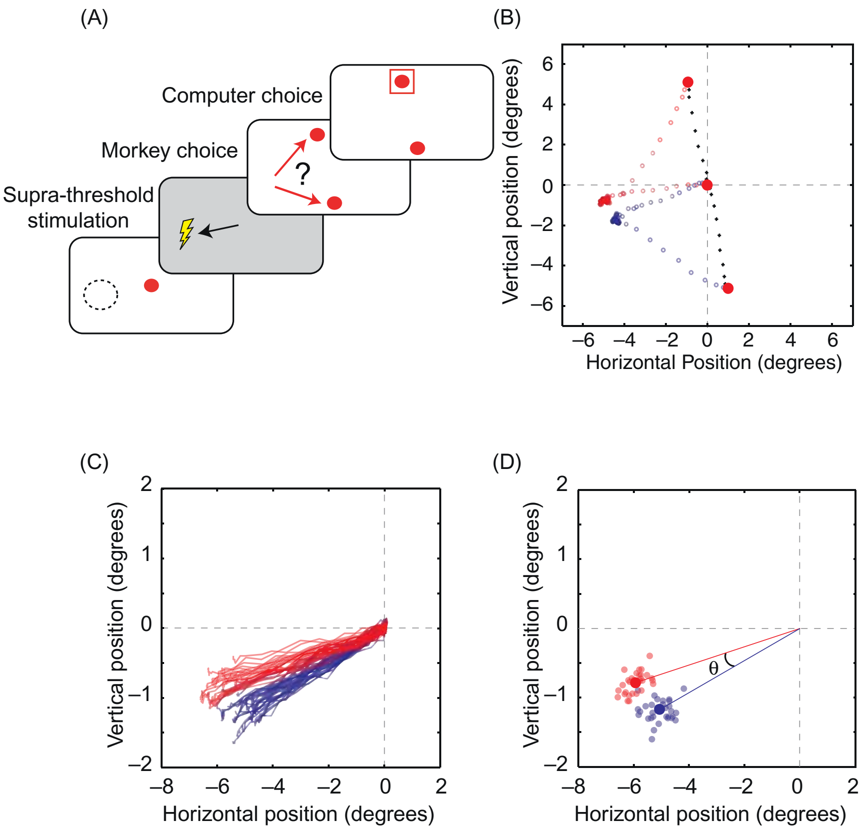

FIGURE 26.9 Using supra-threshold stimulation to test the role of the SC in preparing mixed-strategy saccades. (A) Behavioral task. Stimulation applied during the warning period triggered a saccade that was orthogonal to the targets. Afterwards, monkeys were free to choose either target. Dashed circle and associated lightning bolt indicate the vector of stimulation-evoked saccade. (B–D) Evoked saccades on stimulation trials were segregated based on the final target selection. Data is shown for those trials in which stimulation was applied 500 ms into the 600-ms warning period for a representative stimulation site. (B) On the majority of trials, stimulation was not applied and saccades were made directly to the targets (black crosses). Stimulation-evoked saccades were segregated into two categories: those in which the left (red) or right (blue) target was ultimately selected. (C) Stimulation saccades tended to deviate towards the target that was ultimately chosen. (D) Angular deviation (⊖) was calculated as the angle between the averages of the end-points between the two categories of stimulation-evoked saccades.

whether the predictive activity in the SC outlined above is functionally related to the process of response selection under game theoretic conditions (Figure 26.9A). To test this hypothesis, on a small proportion of matching pennies trials, the ongoing decision process was perturbed with a short burst of micro-stimulation (Figure 26.9B). This stimulated SC location elicited saccades orthogonal to the direction of the choice targets. Because saccade trajectories are determined by population activity across the topographically organized SC map (Lee *et al.*, 1998), stimulation-induced saccades deviate towards regions of pre-existing activity – effectively revealing what option the monkey was in the process of selecting. The authors found that these stimulation-induced saccades deviated towards the location the animal ultimately chose (Figure 26.9C). As the stimulation was applied closer to the time when the choice targets were presented the deviations became more pronounced. The pattern of stimulation-induced deviations over time tracked the time course of the neuronal selection process observed when recording from the SC of these monkeys (Figure 26.8B). Therefore, interrupting developing saccade plans at a range of times preceding the presentation of the choice targets opened a window into the time course of the gradual response selection process during mixed strategy decision making (Figure 26.9D).

Lastly, Thevarajah and colleagues (2009) applied sub-threshold stimulation to the SC in the time leading up to saccadic choices in the matching pennies task. This low level stimulation was enough to bias activity in the SC stimulation site but not enough to directly trigger saccades. The result was that the monkeys' strategies shifted from the predicted Nash equilibrium of equal responses to the two targets in favor of responses towards the site of stimulation. This provided direct, causal evidence that the SC is involved in the selection of mixed strategy saccades and, more generally, highlights how artificially perturbing activity within decision circuits can provide insight into the functional role that a particular brain region plays in the decision process.

CONCLUSIONS

This chapter has outlined the important advances that have been made in understanding the neural circuits subserving social decision making by combining state-of-the-art neurophysiological techniques in non-human primates with microeconomic tasks and statistical analyses. The invasive techniques used in non-human primates allow neural activity to be recorded at high spatial and temporal resolution and correlated to specific stages of game play, behavior or parameters of learning models. Furthermore, the functionality of localized patterns of neural activities on game play can be examined through artificial manipulation. These techniques have allowed researchers to begin to unravel the key fronto-parietal, basal ganglia and brainstem structures that are critical for social decision making. Particularly fruitful has been analyzing behavior and neural signals within the framework of learning models. This allows us to understand the mechanisms by which value representations are constructed according to the animal's previous choices and their outcomes and how choices are selected from these value representations on a trial-by-trial basis.

Once the important statistics of choices and outcomes during a particular game are calculated and various quantities of learning models are updated in associative frontoparietal cortices, the actual selection and execution of the choice must be made. The evidence suggests there is a competition between neuronal populations in premotor regions of the brain (the FEF and SC for saccades) that represent the available actions. Gradually, the activity in one population begins to dominate over the others and, once a threshold level of activation is reached, a movement, or choice, is triggered. It seems likely that a similar competition occurs for purely perceptual decisions where the race to action threshold is influenced by the quality of sensory information. For social decisions, the competition is shaped by economic factors such as the relative value of the targets, the history of past choices and their outcomes, and even fictive information representing the outcomes of what "could have been." Although more work has to be done, the higher order statistics and learning parameters coded in frontoparietal networks appear to shape the competition in spatially organized maps of potential actions such as those within the FEF and SC to bias the competition in favor of the option with the highest subjective value for the chooser. The fact that neuronal circuits are inherently noisy may actually be beneficial to social decision making; it could be a source of stochasticity ensuring that the most valuable action is only more likely – but not deterministically – to occur. Therefore, our brain circuits for social decision making might be designed to exploit the most valuable options during game play while injecting some stochasticity to prevent opponents from exploiting us.

Acknowledgments

We thank Hiroshi Abe, Dominic Barraclough, Hyojung Seo, Dhushan Thevarajah, Areh Mikulic, and Abduallah Abunafessa for their contribution to the research described in this article. We thank Alan Sanfey whose writing contributed to portions of this chapter that were carried over from the first edition of this book. The research of the authors was supported by the National Institute of Health (D. L.) and National Science and Engineering Research Council (M.C.D.).

References

Abe, H., Lee, D., 2011. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. Neuron. 70, 731–741.

Abunafessa, A., Dorris, M.C., 2011. Role of the frontal eye fields in choosing mixed-strategy saccades. Soc. Neurosci. Abstr.

Andersen, R.A., 1995. Encoding of intention and spatial location in the posterior parietal cortex. Cereb. Cortex. 5, 457–469.

Barraclough, D.J., Conroy, M.L., Lee, D., 2004. Prefrontal cortex and decision-making in a mixed-strategy game. Nat. Neurosci. 7, 404–410.

Behrens, T.E., Woolrich, M.W., Walton, M.E., Rushworth, M.F., 2007. Learning the value of information in an uncertain world. Nat. Neurosci. 10, 1214–1221.

Belova, M.A., Paton, J.J., Salzman, C.D., 2008. Moment-to-moment tracking of state value in the amygdala. J. Neurosci. 28, 10023–10030.

Bernacchia, A., Seo, H., Lee, D., Wang, X.J., 2011. A reservoir of time constants for memory traces in cortical neurons. Nat. Neurosci. 14, 366–372.

Bisley, J.W., Goldberg, M.E., 2003. Neuronal activity in the lateral intraparietal area and spatial attention. Science. 299, 81–86.

Byrne, R., Whiten, A., 1989. Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans. Oxford Univ Press.

Cai, X., Kim, S., Lee, D., 2011. Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. Neuron. 69, 170–182.

Camerer, C., Ho, T.H., 1999. Experience-weighted attraction learning in normal form games. Econometrica. 67, 827–874.

Camerer, C.F., 2003. Behavioral Game Theory. Princeton Univ Press, Princeton.

Carello, C.D., Krauzlis, R.J., 2004. Manipulation intent: evidence for a causal role of the superior colliculus in target selection. Neuron. 43, 575–583.

Corrado, G.S., Sugrue, L.P., Seung, H.S., Newsome, W.T., 2005. Linear-nonlinear-poisson models of primate choice dynamics. J. Exp. Anal. Behav. 84, 581–617.

Cui, H., Andersen, R.A., 2007. Posterior parietal cortex encodes autonomously selected motor plans. Neuron. 56, 552–559.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., Dolan, R.J., 2011. Model-based influences on humans' choices and striatal prediction errors. Neuron. 69, 1204–1215.

Dorris, M.C., Glimcher, P.W., 2004. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron. 44, 365–378.

Dorris, M.C., Olivier, E., Munoz, D.P., 2007. Competitive integration of visual and preparatory signals in the superior colliculus during saccadic programming. J. Neurosci. 27, 5053–5062.

Dorris, M.C., Pare, M., Munoz, D.P., 1997. Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. J. Neurosci. 17, 8566–8579.

Erev, I., Roth, A.E., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am. Econ. Rev. 88, 848–881.

Fudenberg, D., Levine, D.K., 1998. The Theory of Learning in Games. MIT Press, Cambridge.

Glimcher, P.W., 2003. The neurobiology of visual-saccadic decision-making. Annu. Rev. Neurosci. 26, 133–179.