

# Neuroeconomics

## Decision Making and the Brain

Edited by

*Paul W. Glimcher, PhD*  
Center for Neuroeconomics  
New York University  
New York, NY, USA

*Colin F. Camerer, PhD*  
Division of Humanities and Social Sciences  
California Institute of Technology  
Pasadena, CA, USA

*Ernst Fehr, PhD*  
Institute for Empirical Research in Economics  
University of Zürich  
Zürich, Switzerland

*Russell A. Poldrack, PhD*  
Department of Psychology  
University of California Los Angeles  
Los Angeles, CA, USA



AMSTERDAM • BOSTON • HEIDELBERG • LONDON  
NEW YORK • OXFORD • PARIS • SAN DIEGO  
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO  
Academic Press is an imprint of Elsevier



# Mechanisms for Stochastic Decision Making in the Primate Frontal Cortex: Single-neuron Recording and Circuit Modeling

*Daeyeol Lee and Xiao-Jing Wang*

## OUTLINE

|   |     |   |     |
|---|-----|---|-----|
| <b>Introduction</b>   | 481 | <i>What is Spiking Network Modeling?</i>                        | 491 |
| <b>Game-theoretic Studies of Decision Making in Non-human Primates</b>            | 483 | <i>A Recurrent Circuit Mechanism of Decision Making</i>         | 492 |
| <i>Game Theory and Learning in Social Decision Making</i>                         | 483 | <i>Neural substrate of a Decision Threshold</i>                 | 494 |
| <i>Choice Behavior During a Matching Pennies Game</i>                             | 484 | <b>Reward-dependent Plasticity and Adaptive Choice Behavior</b> | 495 |
| <i>Choice Behavior During a Rock–Paper–Scissors Game</i>                          | 485 | <i>Computation of Returns by Synapses</i>                       | 495 |
| <b>Neurophysiological Studies of Decision Making in Competitive Games</b>         | 487 | <i>Matching Law: Melioration Versus Optimization</i>            | 496 |
| <i>Role of the Dorsolateral Prefrontal Cortex in Decision Making</i>              | 487 | <i>Random Choice Behavior in Matching Pennies Game</i>          | 497 |
| <i>Role of the Anterior Cingulate Cortex in Decision Making</i>                   | 488 | <b>Conclusion</b>   | 499 |
| <b>Models of Decision Making</b>  | 490 | <b>Acknowledgments</b>  | 500 |
| <i>Drift-diffusion, Leaky Competing Accumulator, and Attractor Network Models</i> | 490 | <b>References</b>   | 500 |

## INTRODUCTION

Decision Making refers to the process of evaluating the desirabilities of alternative choices and selecting a particular option. Thus, many brain functions can be characterized as decision making. For instance,

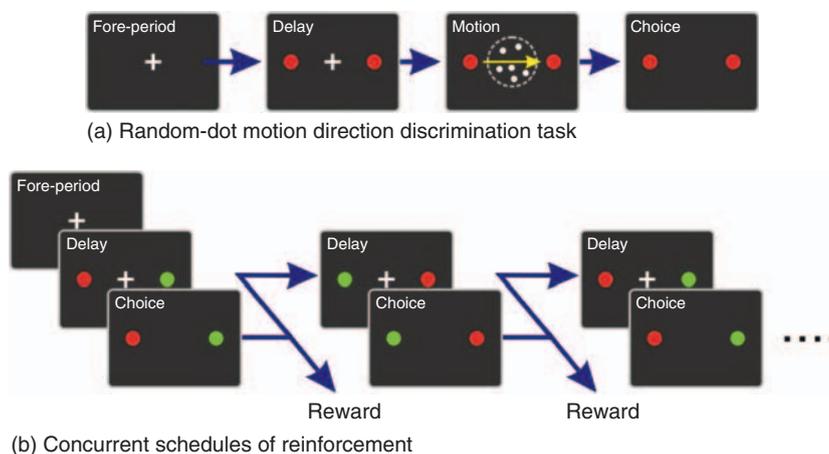
animals seldom behave solely based on their intrinsic metabolic needs, as sensory information about the environment must be taken into account in determining which action the animal should produce to obtain the most desirable outcomes. Sensory inputs are always noisy, and perceptual analysis of these inputs

reduces the uncertainty about the nature of sensory stimuli encountered in the animal's environment in order to categorize stimuli and select the most likely interpretation of the world. Once the relevant stimuli in the animal's environment are properly interpreted, it is then necessary to evaluate the desirability of the outcome expected from each of the alternative behavioral plans. Finally, even when the behavioral goal is set, a particular spatio-temporal sequence of muscle activation must be chosen to carry out the desired action.

Theoretical analyses of steady-state choice behavior are often formulated based on the principles of optimality and equilibrium. For example, game theory seeks to describe optimal decision-making strategies for multiple decision makers trying to maximize their own self-interests through a variety of complex social interactions (von Neumann and Morgenstern, 1944; Nash, 1950; see also Chapter 5 for an introduction to game theory). However, such approaches are limited when the decision makers do not possess all the information about the environment, or when the environment changes frequently. In such cases, the decision makers may need to improve their decision-making strategies incrementally by trial and error. This is often referred to as *melioration* (Herrnstein *et al.*, 1997). Similarly, models based on reinforcement learning (Sutton and Barto, 1998) have been developed to account for how various choice behaviors change dynamically.

Neuroeconomics applies a variety of techniques to investigate the biological processes responsible for decision making. Among them, physiological techniques that probe the activity of individual neurons at millisecond temporal resolution are generally applicable only to animals, due to their invasive nature.

In particular, non-human primates can be trained to perform relatively complex decision-making tasks, and their choice behaviors display many features similar to human choice behaviors. Since the brains of these two species display many structural and functional similarities, the results of single-neuron recording experiments conducted in non-human primates can provide valuable insights into the basic neural circuit mechanisms of decision making in humans. In recent years, electrophysiological studies in behaving monkeys have begun to uncover single-neuron activity that is correlated with specific aspects of perceptual decision making and reward-dependent choice behavior. For instance, Shadlen and his colleagues used a perceptual decision-making task based on random dot motion direction discrimination (Figure 31.1a), and found that activity of individual neurons in the primate posterior parietal cortex reflects the process of accumulating evidence (Roitman and Shadlen, 2002). Others (Sugrue *et al.*, 2004; Lau and Glimcher, 2005) examined the choice behavior of monkeys during a decision-making task based on concurrent schedules of reinforcement (Figure 31.1b), and found that their choice behavior largely conformed to the predictions of the matching law (Herrnstein *et al.*, 1997). Furthermore, the activity of neurons in the posterior parietal cortex encoded the rate of reward or utility expected from a particular choice (Dorris and Glimcher, 2004; Sugrue *et al.*, 2004). In contrast to perceptual discrimination tasks and concurrent schedules of reinforcement, competitive games involve interactions among multiple decision agents. It has been shown that monkeys are capable of producing stochastic choice behaviors that are nearly optimal for such competitive games (Dorris and Glimcher,



**FIGURE 31.1** Decision-Making tasks used in monkeys. (a) A random-dot motion discrimination task. When the animal fixates a central target, two peripheral choice targets are presented. Then, random-dot motion stimuli are presented, and the animal is required to shift its gaze towards one of the choice targets according to the direction of random-dot stimuli. (b) During a decision-making task based on concurrent reinforcement schedules, each target is baited with a particular probability (variable rate) or after a time interval sampled from a particular distribution (variable interval).

2004; Lee *et al.*, 2004, 2005). During such tasks, neurons in the dorsolateral prefrontal cortex and the anterior cingulate cortex exhibited firing activity that reflected history of past choice and rewards (Barraclough *et al.*, 2004; Seo and Lee, 2007; Seo *et al.*, 2007).

These experiments have spurred theoretical work that uses mathematical approaches to illuminate experimental observations. For instance, accumulator models have been widely applied to perceptual decision making (Smith and Ratcliff, 2004; Gold and Shadlen, 2007; see also Chapter 4). Reward-based choice behavior has been described by reinforcement learning models (Sutton and Barto, 1998; see also Chapter 22). We will briefly summarize these models in relationship to our neurophysiological recording experiments in non-human primates. Our focus, however, will be neural circuit modeling, which attempts to go one step further and explore how decision behavior and correlated neural activity can be explained by the underlying circuit mechanisms. For example, what are the neural circuit substrates for time integration of sensory evidence about alternative choices and for action selection? Is valuation of actions instantiated by neurons or synapses, and how does a neural circuit make dynamic decisions adaptively over time? What are the sources of stochasticity in the brain that underlie random choice behavior? We will introduce neural circuit models and illustrate their applications to a perceptual discrimination task (Wang, 2002; Machens *et al.*, 2005; Lo and Wang, 2006; Wong and Wang, 2006; Miller and Wang, 2006), a foraging task based on concurrent schedules of reinforcement (Soltani and Wang, 2006), and a matching pennies game task (Soltani *et al.*, 2006). These computational studies showed that a general neural circuit model can reasonably account for salient behavioral and electrophysiological data from diverse decision tasks, suggesting common building blocks of decision-making circuitry that may be duplicated throughout different stages of sensori-motor transformation in the primate brain.

## GAME-THEORETIC STUDIES OF DECISION MAKING IN NON-HUMAN PRIMATES

### Game Theory and Learning in Social Decision Making

When each of the alternative actions produces a particular outcome without any uncertainty, optimal decision making consists simply of choosing the action

that produces the most desirable outcome. When there is uncertainty about the outcomes expected from various actions, the animal's choice should be influenced by the likelihood of desirable outcomes expected from each action. A large number of economic theories, such as the expected utility theory (von Neumann and Morgenstern, 1944) and prospect theory (Kahneman and Tversky, 1979), have been proposed to account for such decision making under uncertainty or risk. In reality, however, the environment changes constantly, and this frequently alters the likelihood of various outcomes resulting from different actions. Consequently, optimality is rarely achieved, and typically subjects can only approximate optimal decision strategies by learning through experience (Sutton and Barto, 1998).

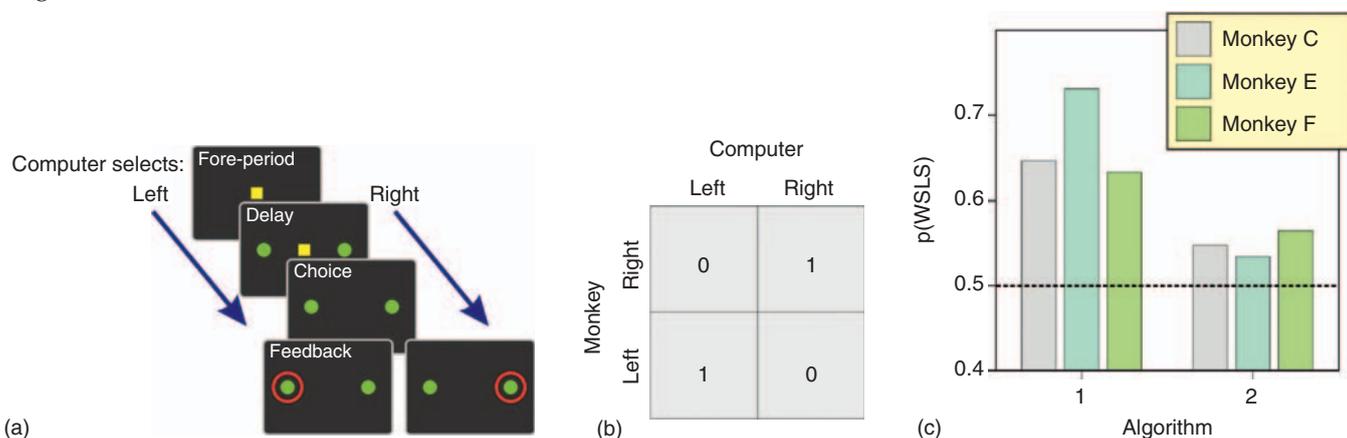
The complexity and hence difficulty of such learning would depend on the nature of dynamic changes in the animal's environment, which can occur for a number of reasons. Some are cyclical, such as seasonal changes, and others are predictable changes resulting from the animal's own actions, such as the depletion of food and other resources. Animals living in social groups face even more difficult challenges, because competitive interactions with other animals tend to make it quite difficult to predict the outcomes resulting from their own actions. Nevertheless, decision making in such social settings provides a unique opportunity to test various theories about the behavioral dynamics and underlying mechanisms of decision making.

One way to tackle mathematically the problems of decision making in a social context is formulated by game theory (von Neumann and Morgenstern, 1944; see also Chapter 5 of this volume). In game theory, a game is specified by a particular number of decision makers or players, a list of alternative choices available to each player, and the payoff matrix that assigns a particular outcome or payoff to each player according to the combination of actions chosen by all players. In other words, the payoff given to a player does not depend simply on that player's own action, but on the actions of all players in the game. In addition, a strategy for a given player is defined as a probability distribution over a set of available choices. A pure strategy refers to choosing a particular action exclusively, whereas a mixed strategy refers to a case in which multiple actions are chosen with positive probabilities. One of the predictions from game theory is that a set of players trying to maximize their self-interests would converge onto a set of strategies known as Nash Equilibrium, which is defined as a set of strategies for all players that cannot be changed by any individual player to increase his payoff. A game is called a mixed-strategy game when its equilibrium

strategy is mixed. In addition, a game is referred to as zero-sum when the sum of the payoffs for all players is always zero, so that someone's gain necessarily means someone else's loss. In the following sections, we describe the results from behavioral experiments to illustrate that non-human primates can approximate the mixed equilibrium strategies through iterative learning algorithms in competitive zero-sum games with two and three alternative actions. These correspond to the familiar games of matching pennies and Rock–Paper–Scissors, respectively.

### Choice Behavior During a Matching Pennies Game

To test whether and how monkeys approximated optimal decision-making strategies in simple competitive games through experience, previous studies have examined the choice behavior of monkeys in a computer-simulated binary zero-sum game, commonly referred to as matching pennies (Lee *et al.*, 2004; Figure 31.2a). In this game, each of two players chooses from two alternative options, and one of the players wins if his choices match and loses otherwise. During the experiment, a monkey was required to begin each trial by fixating a small yellow square presented in the center of a computer screen. 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 its target according to a pre-specified algorithm described below, and the animal was rewarded only when it chose the same target as the computer (Figure 31.2b).



**FIGURE 31.2** Monkey's choice behavior during the matching pennies game. (a) During this task, the animal made a saccadic eye movement towards one of the two peripheral targets to indicate its choice and was rewarded only when it chose the same target as the computer opponent. (b) The payoff matrix for the matching pennies game. (c) The probability that the animal would choose its target according to the so-called win–stay–lose–switch strategy during the matching pennies game against the computer opponent programmed with two different algorithms (1 and 2).

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. Any other strategy could be exploited by the opponent. In the monkey experiment, the strategy of the computer opponent was systematically manipulated to determine how the animal's choice behavior would be affected by the strategy of its opponent. Initially, the computer opponent chose the two targets with the same probabilities, regardless of the animal's choices. This was referred to as algorithm 0, and corresponded to the Nash Equilibrium strategy pursued unilaterally without regard to the opponent's behavior. In this case, the animal's expected payoff would be fixed regardless of how it chose its target. Therefore, it was not surprising that all three monkeys tested with algorithm 0 displayed a strong bias to choose one of the two targets more frequently. Overall, the monkeys C, E, and F chose the right-hand target in 70.0%, 90.2%, and 33.2% of the trials, respectively. In the next stage of the experiment, the computer opponent applied a set of statistical tests to determine whether the animal's choice was randomly divided between the two targets. If not, the computer decreased its probability of choosing a particular target as the animal chose the same target more frequently. This 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 in using the information about its previous rewards to determine its future choices. When tested with algorithm 1, the animals chose the two targets more or less equally frequently. Overall, during algorithm 1, the monkeys C, E, and F chose the right-hand target

in 48.9%, 51.1%, and 49.0% of the trials, respectively. In addition, the animal's successive choices were relatively independent, and as a result, the animal's overall reward rate was close to the optimal value of 0.5 (Lee *et al.*, 2004). Interestingly, the animals were more likely to choose the same target in the next trial if it was rewarded in a given trial (win–stay) and switch to the other target otherwise (lose–switch). Such 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 this so-called win–stay–lose–switch (WSLS) strategy substantially more than in 50% of the trials (Figure 31.2c).

In reinforcement learning models (Sutton and Barto, 1998; Camerer, 2003), the animal's choice is modeled by a set of value functions that are adjusted according to the outcome of the animal's choice. To test whether the animal's choice during the matching pennies game was accounted for by a reinforcement learning model, the value function at trial  $t$  for a given target  $x$ ,  $V_t(x)$ , was updated after each trial according to the following (Lee *et al.*, 2004):

$$V_{t+1}(x) = \alpha V_t(x) + \Delta_t(x) \quad (31.1)$$

where  $x = L$  and  $R$  for the right-hand and left-hand targets, respectively,  $\alpha$  is a decay factor, and  $\Delta_t(x)$  reflects a change in the value function that depends on the outcome of a choice. It was assumed that if the animal chose the target  $x$ , the value function for  $x$  was adjusted according to whether the animal was rewarded or not. Namely, when the animal selected the target  $x$  in trial  $t$ ,  $\Delta_t(x) = \Delta_{\text{rew}}$  if the animal was rewarded, and  $\Delta_t(x) = \Delta_{\text{unrew}}$  otherwise. For the target not chosen by the animal,  $\Delta_t(x) = 0$ . The probability that the animal would select the right-hand target was then determined by the following softmax rule:

$$p_t(R) = \exp V_t(R) / \{\exp V_t(R) + \exp V_t(L)\} \quad (31.2)$$

The model parameters,  $\alpha$ ,  $\Delta_{\text{rew}}$  and  $\Delta_{\text{unrew}}$  were estimated using a maximum likelihood procedure (Pawitan, 2001). This reinforcement learning model was compared with an alternative model based on the WSLS strategy. This WSLS model assumed that the animal chooses its target in each trial according to the WSLS strategy with some probability,  $p_{\text{WSLS}}$ . For example, if the animal was rewarded for choosing a particular target in the previous trial, this model assumes that the animal would choose the same target in the current trial with the probability of  $p_{\text{WSLS}}$ . If the animal was not rewarded in the previous trial, it

would switch to the other target with the same probability. Whether the animal's choice behavior was better accounted for by the reinforcement learning model or the WSLS model was determined by the Bayesian Information Criterion (BIC), defined as

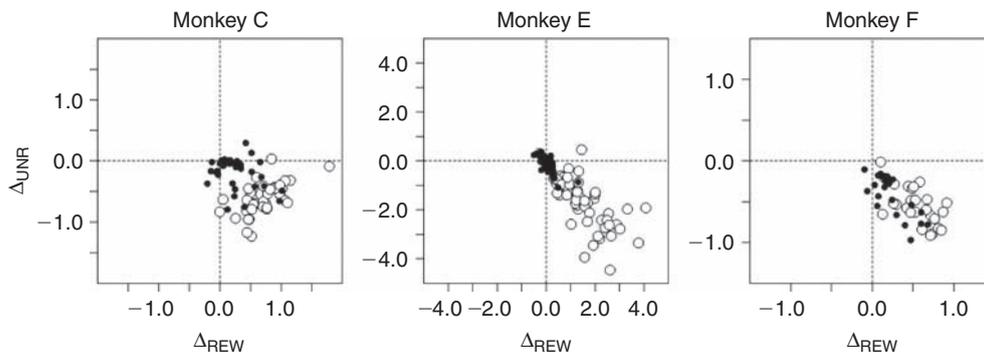
$$\text{BIC} = -\log 2 \log L + k \log N \quad (31.3)$$

where  $L$  denotes the likelihood of the model,  $k$  the number of model parameters ( $k = 1$  and  $3$  for the WSLS model and reinforcement learning model, respectively), and  $N$  the number of trials (Hastie *et al.*, 2001). For each session, the model that minimized the BIC was chosen as the best model. The results of model-fitting showed that for the reinforcement learning model applied to the choice data obtained with algorithm 1, the value functions for a given target tended to increase (decrease) when the animal was (not) rewarded for choosing the same target (Figure 31.3). In addition, the reinforcement learning model performed better than the WSLS model in 56.0% of the sessions (70/125 sessions). Therefore, for algorithm 1, the performance of these two models was comparable, suggesting that the animal's choice was largely determined by the WSLS strategy.

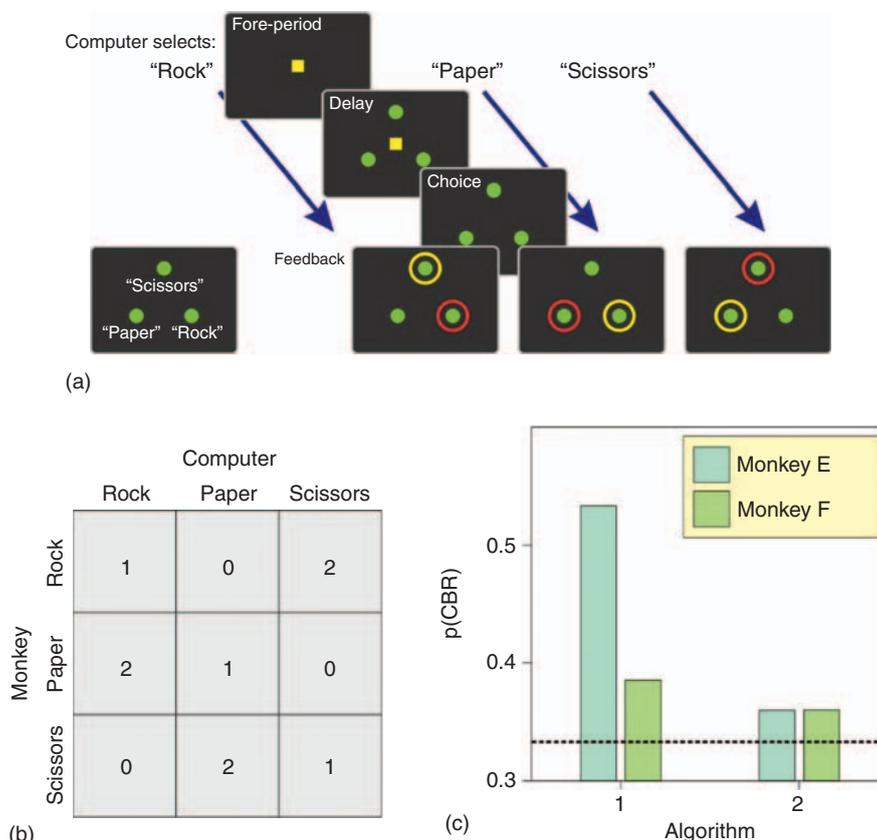
The same animals were then further tested against a third computer opponent which tested not only whether the animal's choice sequences were random, but also whether the animal's choice was independent of its previous choice outcomes (reward vs no reward). This was referred to as algorithm 2 (Lee *et al.*, 2004). Compared to the choice behavior during algorithm 1, the choice behavior resulting from algorithm 2 was dramatically different in that now the tendency to use the WSLS strategy was significantly reduced in all animals (Figure 31.2c). When the choice behavior was analyzed with the reinforcement learning model, it was found that the trial-to-trial changes in the value functions were much smaller compared to the results obtained in algorithm 1 (Figure 31.3). In addition, the reinforcement learning algorithm now accounted for the choice behavior better than the WSLS model in 63.9% of the sessions (62/97 sessions). Therefore, the animals reduced their reliance on the WSLS strategy when it was no longer profitable.

### Choice Behavior During a Rock–Paper–Scissors Game

Two monkeys were tested in a computerized Rock–Paper–Scissors game (Figure 31.4a). During this game, the Nash Equilibrium is to choose randomly each of the three options with the probability of  $1/3$ . Any other strategy can be exploited by the opponent.



**FIGURE 31.3** Parameters in the reinforcement learning model fit to the choice behavior of three different monkeys tested in the matching pennies games. The large empty circles and small filled circles correspond to the parameters obtained for individual sessions tested with algorithms 1 and 2, respectively.



**FIGURE 31.4** Monkey’s choice behavior during the rock–paper–scissors game. (a) During this task, the animal chooses among three alternative targets, and receives different payoffs according to the rules of the rock–paper–scissors game. (b) The payoff matrix for the rock–paper–scissors game. (c) The probability that the animal would choose its target in the rock–paper–scissors game corresponding to the so-called Cournot best response against the opponent using two different algorithms.

Similar to the matching pennies game, each trial began when the animal fixated a small yellow square at the center of the computer screen. Then, three identical green disks were presented, with their spatial positions arbitrarily designated as “Rock,” “Paper,” and “Scissors.” As in the matching pennies task, the computer chose its target according to one of three

different algorithms, and the animal was rewarded according to the payoff matrix of the Rock–Paper–Scissors game (Figure 31.4b). Namely, the animal was rewarded by a drop of fruit juice when it chose the same target as the computer opponent, and by two drops of juice when its choice beat the computer’s (e.g., when the animal and the computer chose

Rock and Scissors, respectively). The animal was not rewarded when its choice was beaten by the computer's choice.

Similar to the results obtained from the matching pennies task, each animal displayed an idiosyncratic bias to choose a particular target, while the computer opponent unilaterally complied with the Nash Equilibrium by choosing the three targets with the same probabilities regardless of the animal's choices (Lee *et al.*, 2005; algorithm 0). Furthermore, as in the matching pennies task, when the computer opponent predicted the animal's choice and behaved competitively using only the information about the animal's choice sequences (algorithm 1), the animal started choosing the three targets with more or less equal probabilities. Nevertheless, there was a significant bias to choose the target that would have beaten, and therefore was the best response to, the computer's choice in the previous trial. For example, if the animal chose Rock in a given trial, it tended to choose Scissors in the next trial. This is referred to as the *Cournot best response* (Lee *et al.*, 2005), which becomes the WLS strategy when there are only two alternative choices. In the final stage of the experiment (algorithm 2), the computer opponent utilized information about both the animal's choice sequences and their outcomes. Therefore, the computer could now detect whether there was any bias for the animal to rely on such strategies as the Cournot best response. During algorithm 2, the probability of using the Cournot best response was significantly reduced in both animals. Nevertheless, both animals chose their targets corresponding to the Cournot best response significantly more frequently than expected by chance (36.0% for both animals; Figure 31.4c). These results were largely consistent with the predictions of a reinforcement learning algorithm (Lee *et al.*, 2005).

## NEUROPHYSIOLOGICAL STUDIES OF DECISION MAKING IN COMPETITIVE GAMES

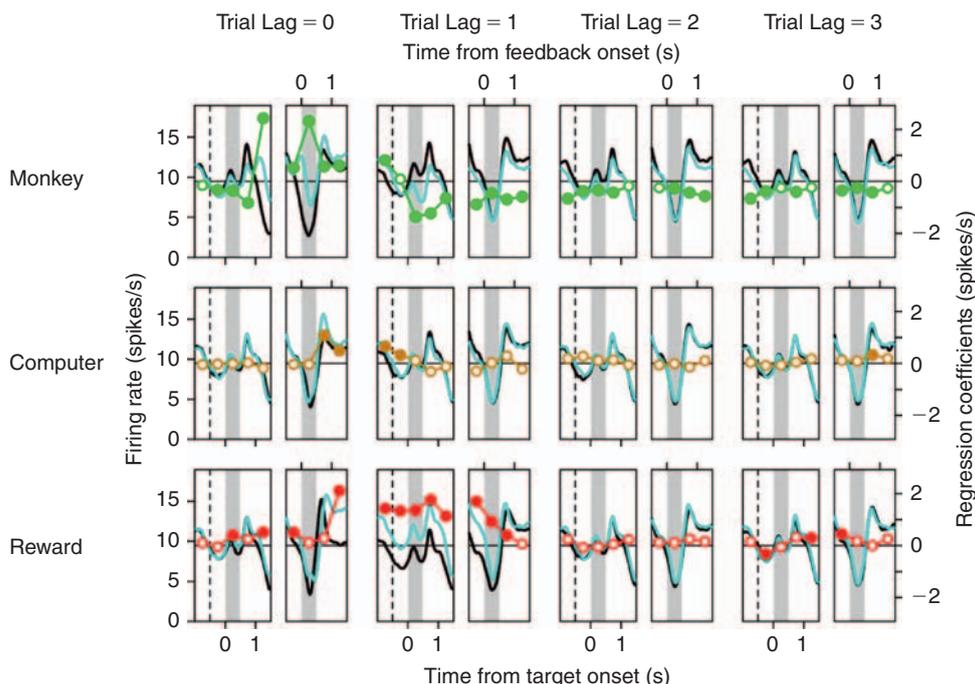
The results described in the previous section suggest that monkeys might approximate equilibrium strategies in competitive games using reinforcement learning algorithms. Recently, a relatively large number of studies have identified neural signals related to the key components of reinforcement learning in multiple brain areas. For example, signals related to the discrepancy between the predicted and actual rewards, commonly referred to as reward-prediction error, have been found in the anterior cingulate cortex (Matsumoto

*et al.*, 2007; Seo and Lee, 2007) as well as the midbrain dopamine neurons (Schultz, 1998). In addition, signals resembling value functions have been identified in various cortical areas and the basal ganglia (see the other chapters in Part 5 of this volume). Thus, many of these cortical and subcortical areas might also be involved in updating the animal's decision-making strategy during competitive games. The studies described below tested this in the dorsolateral prefrontal cortex and the anterior cingulate cortex (Barraclough *et al.*, 2004; Seo and Lee, 2007; Seo *et al.*, 2007).

### Role of the Dorsolateral Prefrontal Cortex in Decision Making

The primate dorsolateral prefrontal cortex (DLPFC) is anatomically connected with the multiple cortical areas involved in the processing of sensory, motor, and affective signals (Barbas and Pandya, 1987; Vogt and Pandya, 1987; Barbas, 1995; Carmichael and Price, 1996; Fuster, 1997; Luppino *et al.*, 2003). Consistent with such diverse sources of its inputs, neurons in the DLPFC display modulations in their activity that reflect various aspects of sensory stimuli, motor responses, and mnemonic processes (Chafee and Goldman-Rakic, 1998; Constantinidis *et al.*, 2001; Amemori and Sawaguchi, 2006). In addition, many neurons in the DLPFC change their activity according to the nature of the reward expected from a particular state of the environment (Watanabe, 1996; Leon and Shadlen, 1999; Kobayashi *et al.*, 2002). These findings suggest that the DLPFC might play an important role in decision making by encoding a particular state of the environment and the desirability of the outcome expected from such a state (Lee *et al.*, 2007).

Whether the neurons in the DLPFC encode signals necessary to optimize the animal's decision-making strategy was tested during the matching pennies task. The results from these experiments showed that the DLPFC neurons encode a multitude of signals that can contribute to different stages of decision making (Barraclough *et al.*, 2004; Seo *et al.*, 2007). First, some neurons in the DLPFC modulated their activity according to the value functions of the two alternative targets, and therefore might be involved in computing or encoding the decision variable necessary to determine the animal's choice. Second, activity of many neurons in the DLPFC encoded not only the upcoming choice of the animal, but also the animal's choices in the previous trials (Figure 31.5, top; Figure 31.6, top). In real life, the consequence of a particular action might be revealed after some temporal delay and often after several intervening actions. This leads to the problem



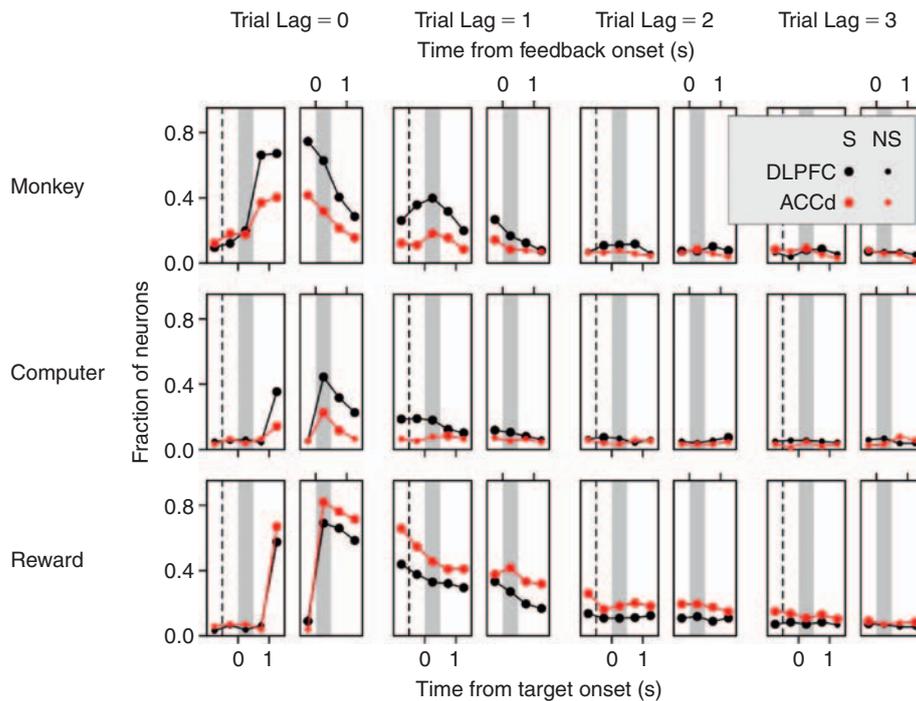
**FIGURE 31.5** An example neuron recorded in the dorsolateral prefrontal cortex of a monkey playing the matching pennies game against a computer opponent. Each panel shows the average activity of the neuron aligned to the onset of the targets or the feedback stimulus, separately according to the animal's choice (top), the choice of the computer opponent (middle), or the reward (bottom) in the current (Trial lag = 0) or previous trials (Trial lag = 1 to 3). Black and blue lines indicate whether the animal (top) or the computer (middle) selected the leftward and rightward target, or whether the animal was rewarded (black) or not (blue). Filled and empty circles correspond to the regression coefficients quantifying the effect of each variable in different time bins and also indicate whether each coefficient was statistically significant (filled;  $t$ -test,  $P < 0.05$ ) or not (empty). Dotted vertical lines indicate the time when the animal fixated the central target. The gray background in the left panels corresponds to the cue period during which the animal was required to maintain its fixation on the central target, whereas the gray background in the right panels corresponds to the feedback period.

of temporal credit assignment, since it may be important to identify correctly the action that was responsible for a particular outcome. Memory signals related to the animal's previous actions, such as those found in the DLPFC, are referred to as eligibility trace, and may play an important role in solving such problems. Third, some neurons in the DLPFC modulated their activity according to the previous choices of the computer opponent (Figure 31.5, middle; Figure 31.6, middle). During the matching pennies game, the average reward expected from choosing a particular target is directly related to the probability that the same target will be chosen by the computer opponent. Therefore, the signals related to the computer's previous choices provide the information necessary to estimate the likelihood of reward from a particular target. Finally, activity of the DLPFC neurons during the feedback period was often influenced by the outcomes of the animal's previous choices, namely, whether the animal was rewarded or not in the previous trials (Figure 31.5, bottom; Figure 31.6, bottom). This suggests that the activity in the DLPFC might be influenced by the animal's reward history and therefore by the context

in which a particular reward is delivered or omitted. Overall, these results suggest that the DLPFC might be closely involved in updating the animal's decision-making strategies as well as determining the animal's choice in a given trial.

### Role of the Anterior Cingulate Cortex in Decision Making

The primate anterior cingulate cortex is closely linked to the lateral prefrontal cortex and the orbitofrontal cortex (Vogt and Pandya, 1987; Barbas and Pandya, 1989; Carmichael and Price, 1996; Luppino *et al.*, 2003). Similar to the lateral prefrontal cortex, the anterior cingulate cortex is also connected to cortical areas associated with motor functions, such as the supplementary motor area and supplementary eye-field (Wang *et al.*, 2001; Luppino *et al.*, 2003). Compared to the lateral prefrontal cortex, however, the anterior cingulate cortex is more densely connected with the subcortical structures involved in emotion and reward processing, such as the amygdala (Barbas



**FIGURE 31.6** Fraction of neurons in the dorsolateral prefrontal cortex (black) and the dorsal anterior cingulate cortex (red) that significantly modulated their activity at a particular time bin according to the animal's choice (top), the choice of the computer opponent (middle) or the reward received by the animal (bottom) in the current (Trial lag = 0) or previous (Trial lag = 1 to 3) trials. Large symbols indicate that the number of neurons with significant effects was larger than expected by chance (binomial test,  $P < 0.05$ ).

and De Olmos, 1990; Ghashghaei *et al.*, 2007) and the ventral tegmental area (Williams and Goldman-Rakic, 1993; Lewis *et al.*, 2001). Accordingly, neurons in the anterior cingulate cortex often modulate their activity according to the expected and actual reward (Niki and Watanabe, 1979; Amiez *et al.*, 2005, 2006), but neural signals reflecting the information about specific actions are less frequently observed in the anterior cingulate cortex than in the lateral prefrontal cortex (Hoshi *et al.*, 2005; Matsumoto *et al.*, 2007).

Consistent with these findings, the activity of neurons in the anterior cingulate cortex was commonly modulated by the outcomes of the animal's choices during the matching pennies game. For example, many ACC neurons increased or decreased their activity when the animal received a feedback signal indicating that it would be rewarded in a given trial. Such activity related to the choice outcome or reward was often maintained across multiple trials, and, as a result, the reward-related activity in the ACC was often influenced by the animal's reward history (Seo and Lee, 2007). In addition, the time-course of such activity related to the outcome of the animal's previous choice varied substantially across different neurons. In some neurons, activity was influenced similarly by the animal's rewards in the current and previous trials, suggesting that the activity of such

neurons might signal the local rate of reward. Such signals might be used as a reference point against which the value of the reward in a current trial is evaluated. In contrast, the activity of other neurons was modulated antagonistically by the rewards in the current and previous trials. For example, a particular neuron might increase its activity when the animal is rewarded in the current trial, but its activity might be decreased if the animal was rewarded in the previous trial. Such neurons would then show their maximal or minimal activity when the animal was rewarded after losing a reward in the previous trial. Analogous to the activity of dopamine neurons that encode reward-prediction error (Schultz, 1998), some neurons in the ACC might therefore encode the deviation of the reward received by the animal in the current trial relative to the reward expected from the outcomes in the previous trial. To test this further, the value functions of the two alternative targets as well as the reward-prediction error were estimated on a trial-by-trial basis using a standard reinforcement learning model, and these estimates were used to test whether the activity of each ACC neuron was correlated with the sum or difference of the value functions or with the reward-prediction error. The results showed that the activity of many ACC neurons was correlated with the sum of the value function, but relatively infrequently with

the difference of value functions. Since reinforcement learning algorithms choose a particular action based on the difference in value functions, these results suggest that the ACC makes only a minor contribution to the process of action selection. In contrast, the activity of many neurons in the ACC was strongly correlated with the reward-prediction error. Presumably, outcome-related activity in the ACC that is modulated by the animal's reward history might influence the process of updating the value function of the alternative targets and therefore the animal's subsequent actions. Precisely how this is accomplished, however, is currently not known.

## MODELS OF DECISION MAKING

Experiments using different types of decision behaviors have led to two broad classes of models for decision making. On the one hand, so-called sequential-sampling models describe information processing that unfolds in time and determine performance accuracy and reaction times in perceptual and memory tasks. On the other hand, game-theoretical models and reinforcement learning models account for dynamic choice behavior which is based on utility maximization and interplay with the environment or other decision agents. These models are important for quantitatively describing behavioral data and assessing theoretical ideas about the cognitive processes of decision making. To truly understand the biological basis of decision behavior, however, it is critical to construct realistic neural circuit models that allow us to explore how key computations of decision making are instantiated by cellular machineries and collective dynamics of neural networks in the brain. Here we summarize models of varying degrees of complexity, with a focus on recently advanced neural models of decision circuits. We show that both perceptual decisions and value-based choice behaviors can be described by a recurrent neural circuit model that is endowed with reward-dependent synaptic plasticity.

### Drift-diffusion, Leaky Competing Accumulator, and Attractor Network Models

Sequential-sampling models are based on the intuitive idea that a decision is reached when stochastic accumulation of information about alternative choices reaches a particular threshold. For two-alternative forced choice tasks, a specific implementation is

called the drift-diffusion model (Ratcliff, 1978; Smith and Ratcliff, 2004). In this model, an activity variable  $X$  represents the difference between the respective amounts of accumulated information about the two alternatives, say  $X_1$  and  $X_2$ ,  $X = X_1 - X_2$ . The dynamics of  $X$  is given by the drift diffusion equation,

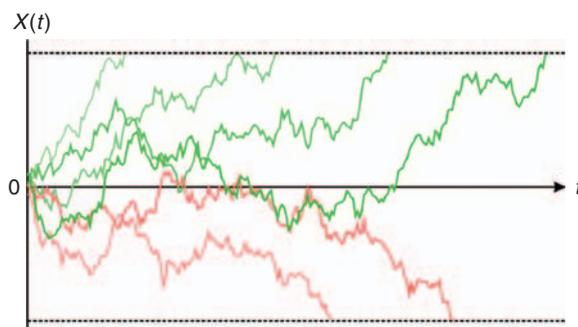
$$dX/dt = \mu + w(t) \quad (31.4)$$

where  $\mu$  is the drift rate and  $w(t)$  is a white noise of zero mean and standard deviation  $\sigma$ . The drift rate  $\mu$  represents the bias in favor of one of the two choices (and is zero if there is no net bias). For instance, in a random-dot motion direction discrimination task,  $\mu$  is proportional to the strength of motion signal. This system is a perfect integrator of the input:

$$X(t) = \mu t + \int_0^t w(t') dt' \quad (31.5)$$

The integration process is terminated and the decision time is read out whenever  $X(t)$  reaches a positive threshold  $\theta$  (choice 1) or a negative threshold  $-\theta$  (choice 2). If the drift rate  $\mu$  is positive, then choice 1 is correct, while choice 2 is an error (Figure 31.7). Therefore, this type of models is commonly referred to as "ramping-to-threshold" model, with the average ramping slope given by  $\mu$ .

The drift-diffusion model has been widely applied to fit behavioral data from perceptual and memory experiments (Ratcliff, 1978; Smith and Ratcliff, 2004). This model is the continuous-time equivalent of the discrete-time Sequential Probability Ratio Test (SPRT), which is the optimal procedure for making binary choices under uncertainty, in the sense that it



**FIGURE 31.7** Drift diffusion model for time integration in reaction time tasks. The variable  $X(t)$  represents the difference in the activity at time  $t$  between two neural pools A and B.  $X(t)$  undergoes a biased random walk, and decision A (or B) is reached when  $X$  reaches a prescribed threshold  $\theta$  (or  $-\theta$ ). The time it takes to reach the threshold is read out as the response time. Six sample trials (four correct trials, green; two error trials, red) are shown.

minimizes the mean decision time among all tests for a given lower bound of error rate (Wald and Wolfowitz 1948; Bogacz *et al.*, 2006).

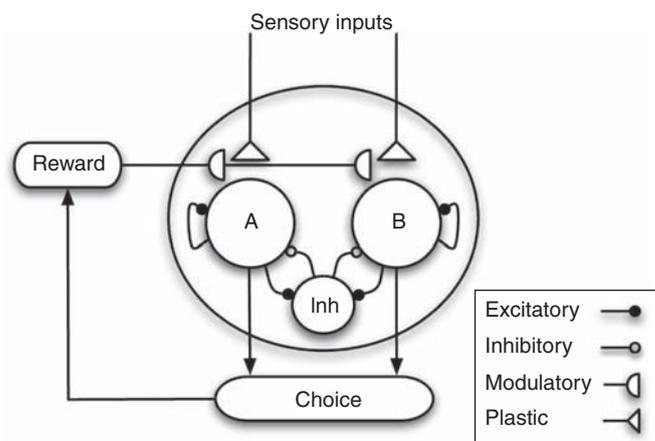
Can a ramping-to-threshold mechanism be instantiated by neural circuits? One key issue here is to determine the biological basis of time integration. The drift-diffusion model is a perfect integrator, whereas neurons and synapses are leaky, with short time constants of tens of milliseconds (Kandel *et al.*, 2000). Usher and McClelland (2001) extended the drift-diffusion model by incorporating a leak so that the integrator becomes forgetful with a decay time constant. Moreover, there is a competition between the two dynamical variables  $X_1$  and  $X_2$  through mutual inhibition. This leaky competitive accumulator model proved to fit many behavioral data equally as well as the drift-diffusion model, provided that the integration time is sufficiently long.

We have proposed that a long integration time can be realized biophysically in a decision neural network through recurrent excitation (Wang, 2002). Reverberating excitation represents a salient characteristic of cortical local circuits (Douglas and Martin, 2004). When this positive feedback is sufficiently strong, recurrent excitation in interplay with synaptic inhibition can create multiple stable states (“attractors”). Therefore, this class of models is referred to as “attractor network models.” Such models have been initially proposed for working memory – namely, the brain’s ability to actively hold information online in the absence of direct sensory stimulus. If a cortical network is endowed with a spontaneously active resting state and multiple memory states (for maintaining different items of information), transient inputs (such as a sensory cue) can switch the system from the resting state to one of the memory states, which is then self-sustained in the form of persistent neural activity. This idea has been explored using biophysically realistic neural circuit models (Wang, 2001; Amit and Mongillo, 2003). Recently, it has been realized that the same models can also implement decision-making computations (Wang, 2002; Machens *et al.*, 2005; Miller and Wang, 2006; Wong and Wang, 2006). Interestingly, physiological studies in behaving non-human primates often reported neural activity correlated with decision making in cortical areas, such as the prefrontal cortex or the parietal cortex, that also exhibit mnemonic persistent activity during working memory. Hence, recurrent network (attractor) models can be viewed as a promising neuronal instantiation of the ramping-to-threshold model, and they serve as a framework to elucidate a common circuit mechanism for decision making and working memory in the brain.

## What is Spiking Network Modeling?

Physiological experiments in behaving animals are critical to uncover neural signals correlated with specific aspects of decision making. Biophysically-based neural modeling can delineate circuit mechanisms that give rise to the observed neural signals, and identify key computational principles at the conceptual level. For certain questions about decision making, discussed below, it is important to capture neural firing of action potentials or spikes (electrical signals often described mathematically as point processes) through which neurons transmit information and communicate with each other. To this end, single cells should be described by a spiking neuron model, rather than a firing-rate model (as in the Usher-McClelland model). A popular choice is either the leaky integrate-and-fire model or the Hodgkin-Huxley model. Such a model is calibrated by physiological measurements, such as the membrane time constant and the input–output function (the spike firing rate as a function of the synaptic input current), which can be different for excitatory pyramidal cells and inhibitory interneurons. Furthermore, it is worth emphasizing that, in a biophysically-based model, synapses must be modeled accurately. Unlike connectionist models in which coupling between neurons is typically an instantaneous function of firing activity, synapses have their own rise-time and decay time constant, and exhibit summation properties. Synaptic dynamics turn out to be a crucial factor in determining the integration time of a neural circuit dedicated to decision making, as well as controlling the stability of a strongly recurrent network. Finally, “building blocks” (single cells and synapses) are used to construct a network endowed with a biologically plausible architecture. A commonly assumed circuit organization is the so called “Mexican-hat connectivity” – local excitation between neurons of similar selectivity combined with a more global inhibition (see Figure 31.8 for a schematic depiction of a simple two-neural pool network). Dynamic balance between synaptic excitation and inhibition is another feature of cortical microcircuits that has been increasingly recognized experimentally and incorporated in cortical network models.

Although this type of spiking network model may seem complicated compared with simple mathematical models such as the drift-diffusion model, it is actually minimal for quantitatively accounting for both behavioral data (determined by collective network dynamics) and spike train data of single neurons obtained from physiological experiments, and for exploring the synaptic circuit mechanisms required to explain these observations.



**FIGURE 31.8** A simple model for decision making with two alternatives. There are two pyramidal cell groups, each of which is selective to one of the two choice options, A or B. Within each pyramidal neural group there are strong recurrent excitatory connections that can sustain persistent activity triggered by a transient preferred stimulus. The two neural groups compete through feedback inhibition from interneurons. Depending on the task design, one of the choices may be rewarded with some probability, whereas the other may not, in any given trial. The outcome signal (reward or not) is assumed to modulate Hebbian plasticity of input synapses  $c_A$  and  $c_B$ . Since the network's decision dynamics depends on  $c_A$  and  $c_B$ , altered synaptic strengths lead to adaptive choice behavior across trials.

## A Recurrent Circuit Mechanism of Decision Making

A simple model for decisions with two alternative choices is schematically illustrated in Figure 31.8 (Wang, 2002). Two neural pools are selective for choice options (A or B), each consisting of a number of spiking neurons. The network connectivity exhibits salient features of a cortical circuit model: (1) direct excitatory connections are strong between pyramidal cells with similar selectivity (within each neural pool) and weak between those with entirely different selectivity (between the two neural pools); (2) there is a balance between synaptic excitation and inhibition; and (3) inhibitory interneurons induce competition between the two neural pools. As a result, interactions between the two neural pools are effectively inhibitory (Wong and Wang, 2006), similar to the Usher-McClelland model (Bogacz *et al.*, 2006). However, in contrast to the linear leaky competitor model, non-linear attractor dynamics that depend on reverberating excitation are critical to winner-takes-all competition.

Evidence for the two alternatives is provided by inputs  $I_A$  and  $I_B$  to the two neural pools. For instance, choice should be obvious if  $I_A$  is large and  $I_B$  is zero, whereas the weights of evidence for the two alternatives are equal if  $I_A = I_B$ . This is shown in Figure 31.9, where the input magnitudes are expressed in terms

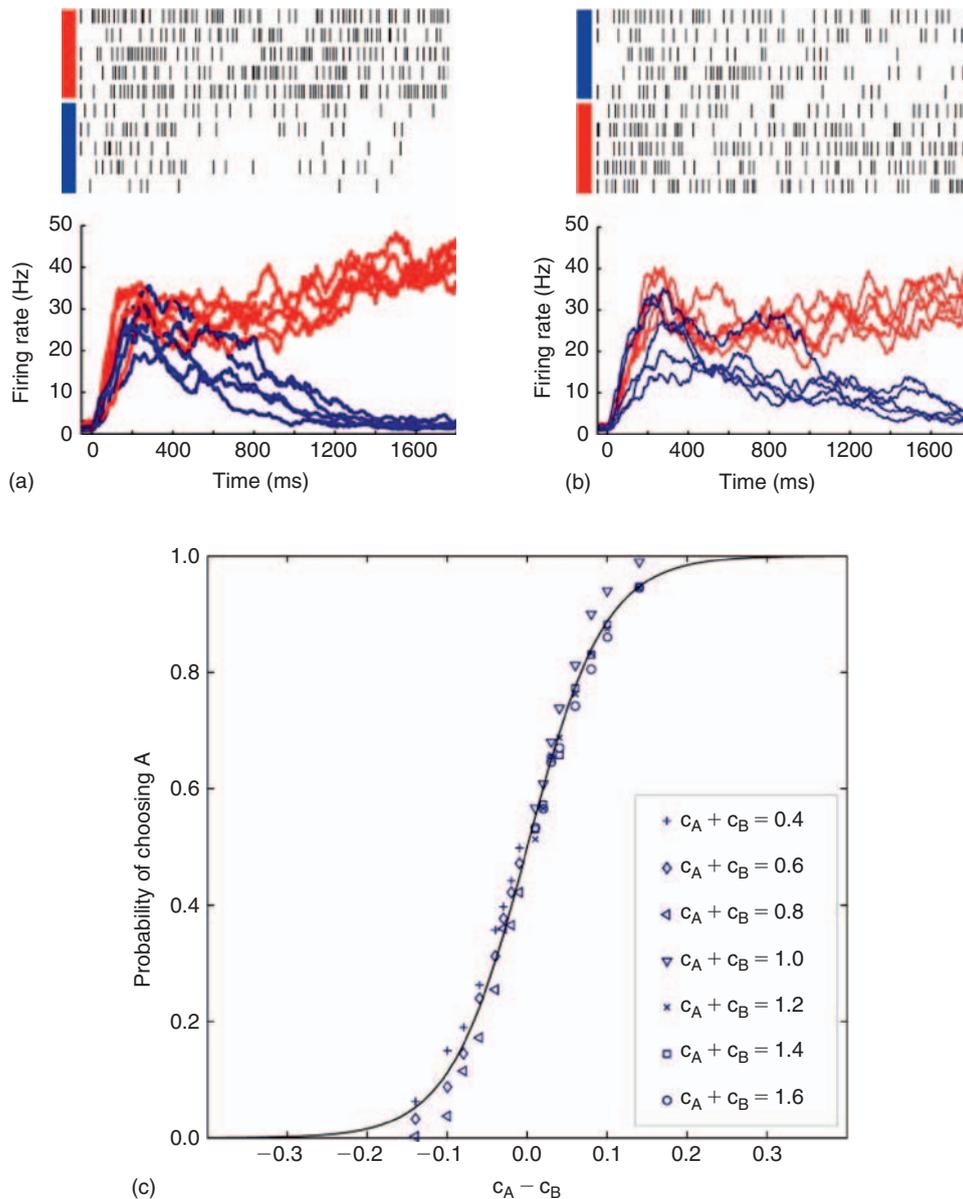
of the synaptic strengths  $c_A$  and  $c_B$  of the input pathway. Figures 31.9a and 31.9b display the time-course of network activity when  $c_A$  is larger than  $c_B$  in several simulation trials. In Figure 31.9a, the population firing rates  $r_A$  and  $r_B$  of the two neural pools first increase together, then start to diverge:  $r_A$  ramps up for hundreds of milliseconds, whereas  $r_B$  gradually decays away. Therefore, the decision outcome of the network is A, which is the correct choice. Figure 31.9b shows several error trials, in which the neural pool B rather than A wins the competition. Across many trials, the probability of choosing A (the correct response) is larger than 0.5 when evidence in favor of A is stronger than that for B ( $c_A$  is larger than  $c_B$ ). Figure 31.9c shows that this probability is a sigmoid function of the difference  $c_A - c_B$ . Note that even when the inputs are equal ( $c_A = c_B$ ), stochastic neural spiking dynamics still gives rise to winner-takes-all competition in each trial, with the choice probability at chance level (50%). Interestingly, this choice probability function is relatively insensitive to the absolute strengths  $c_A$  and  $c_B$ ; therefore the network's decision performance can be described by a softmax function

$$P_A(c_A - c_B) = 1/(1 + \exp(-(c_A - c_B)/\sigma)) \quad (31.6)$$

where  $\sigma$  expresses the amount of stochasticity due to irregular spike firing in the network and also depends on other model properties, such as the firing rates of input neurons. Importantly, a softmax decision criterion is widely assumed in more abstract models of choice behavior; indeed, it is the same equation used in the reinforcement learning model for fitting monkey's behavioral data (equation (31.2)). Our neural circuit modeling lends support to this general assumption, and allows insights into its underlying stochastic recurrent neural dynamics.

Can this model account for reaction times as well as accuracy in a decision task? This network exhibits quasi-linear ramping activity for many hundreds of milliseconds (Figure 31.9a, 31.9b). This ramping activity is realized in spite of a short neuronal membrane time constant ( $\sim 10$ ms) by slow reverberation mediated by the NMDA receptors (with a time constant of 50–100 ms) at the excitatory recurrent synapses. As in the drift-diffusion model, a fixed firing threshold  $\theta$  can be set such that a decision is made whenever  $r_A$  or  $r_B$  reaches  $\theta$  first, and the time elapsed since the stimulus onset can be read out as the decision time. This model also predicts that reaction times are longer in error trials than in correct trials, which is commonly observed in decision tasks.

Our model has been applied to a monkey experiment using a visual motion direction discrimination



**FIGURE 31.9** Neuronal activity and decision performance of the model shown in Figure 31.8. (a, b) The population firing rate of neurons is shown separately for trials in which the network’s choice is the neurons’ preferred (red) or non-preferred (blue) target. Raster plots show spike trains for two selected neurons in population A (top) and B (bottom) in the same trials. Neural groups A and B win the competition in trials shown in A and B, respectively. Activity is aligned at the onset of the visual targets. A few hundred milliseconds after the input onset, the average firing rates in the two populations start to diverge. Spiking activity is higher when the chosen target is preferred for the neuron (compare red and blue lines), and when its input is larger (compare red traces in the left and right panels). Moreover, firing activity is higher when the chosen target is non-preferred for the neuron that receives a larger input (compare blue traces in the left and right panels). In these simulations, the synaptic strengths are  $c_A = 0.33$  and  $c_B = 0.27$ . (c) Choice behavior of the decision-making network for different sets of synaptic strengths. The probability of choosing target A is a function of the difference between the two synaptic strengths and it can be fitted by a sigmoid function. Different symbols represent different overall synaptic strengths.

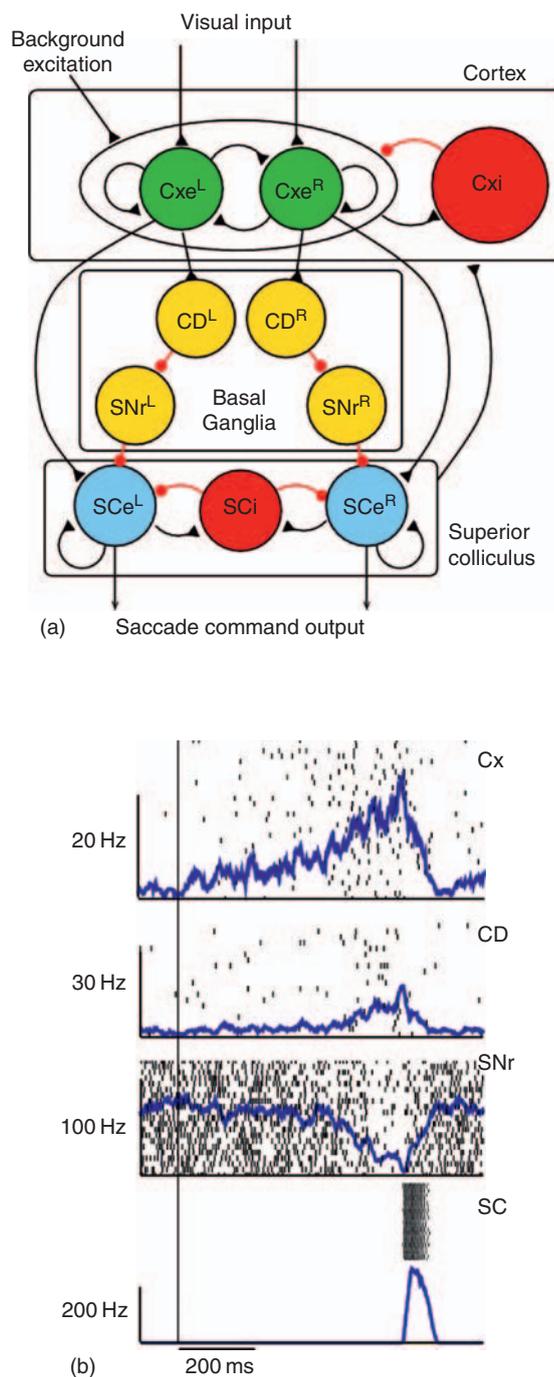
task (Figure 31.1a). In this experiment, a subject was shown a display of moving random dots, a fraction of which moved coherently in one of two possible directions (say A = left, B = right), and the remaining dots moved in random directions. The task difficulty was varied from trial to trial by the motion coherence (0 to 100%). In monkeys performing this

task, single neurons in the lateral intraparietal (LIP) cortex exhibited slow ramping activity that correlated with the perceptual decision about the direction (leftward or rightward) of the motion stimulus (Gold and Shadlen, 2007). At lower motion coherence the subject’s reaction time was longer, and the ramping of the LIP neuronal firing rate was slower but reached the

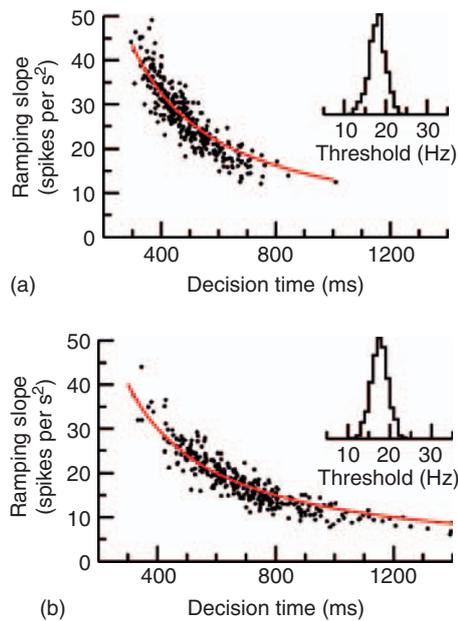
same firing activity level at the time when the behavioral response was produced, regardless of the motion coherence (Roitman and Shadlen, 2002). Thus, LIP neurons display a ramping-to-threshold process at the cellular level. Our neural circuit model has been used to simulate this monkey experiment, with the motion coherence given by the relative input strength  $(I_A - I_B)/(I_A + I_B)$  (between  $-100\%$  and  $+100\%$ ). This model reproduces the monkey's performance and reaction times, as well as salient physiological data of LIP neurons (Wang, 2002).

### Neural Substrate of a Decision Threshold

Other studies have also revealed ramping-to-threshold neural activity at the single cell level that is correlated with action selection and preparation (Schall, 2001). How can a decision threshold be instantiated by neurons, rather than prescribed in an *ad hoc* manner? A natural hypothesis is that when decision neurons integrate inputs and reach a particular firing rate level, this event triggers an all-or-none response in downstream neurons and leads to the generation of a behavioral output. This idea was tested for oculomotor decision tasks in which the motor response is a rapid saccadic eye movement. In an extended, two-stage circuit model (Lo and Wang, 2006), decision neurons in the cortex (as described above) project to movement neurons in the superior colliculus (SC), an important command center for saccades (Figure 31.10a). This model also includes a direct pathway in the basal ganglia, with an input layer (caudate, CD) and an output layer (substantia nigra reticulata, SNr). As a neural pool in the cortex ramps up in time, so do the synaptic inputs to the corresponding pool of SC movement neurons. When this input exceeds a well defined threshold level, an all-or-none burst of spikes is triggered in the movement cells, signaling a particular (A or B) motor output. In this scenario, a decision threshold (as a bound of firing rate of decision neurons) is instantiated by a hard threshold of synaptic input for downstream motor neurons. Figure 31.10b shows a sample trial of such a model simulation for the visual motion direction discrimination experiment. The rate of ramping activity fluctuates from trial to trial, as a result of stochastic firing dynamics in the cortex, and is inversely related to the decision time (as defined by the time when a burst is triggered in the SC) on a trial-by-trial basis (Figure 31.11). Moreover, when the task is more difficult (with a lower motion coherence) ramping activity is slower, leading to longer reaction times [(b) compared to (a) in Figure 31.11)]. However, the threshold of cortical



**FIGURE 31.10** Multiple-circuit model of a decision threshold in the oculomotor system. (a) Schematic model architecture. Neural pools in the cortical network integrate sensory information, as well as compete against each other. They project to both superior colliculus, and caudate nucleus in basal ganglia. Excitatory and inhibitory connections are indicated as black and red lines, respectively. (b) Reaction time simulation using the oculomotor network model of spiking neurons. A saccadic response is triggered by a burst of spikes in SC movement neurons, when the ramping activity in one of the selective neural pools in the cortex exceeds a threshold, so that SC movement neurons receive sufficient cortical excitation, and are released from inhibition by the basal ganglia as a result of cortical drive to the caudate.



**FIGURE 31.11** The slope of ramping activity of cortical decision neurons is inversely correlated with the reaction time in single trials of simulated random-dots motion direction discrimination task. When the motion strength is weaker (3.2% in (b)) compared to 12.8% in (a)), the ramping slopes are smaller and the reaction times are longer. However, the variability of the cortical activity threshold remains the same (inserts), and therefore the variability in the reaction time is primarily caused by the variability of the ramping slope due to neural firing fluctuations across trials.

firing activity that is read out by the downstream motion system has the same narrow distribution (inserts in Figure 31.11), regardless of the ramping speed or reaction times. Therefore, this model realizes a robust threshold detection mechanism, and the variability of reaction times is mostly attributed to the irregular ramping of neural activity itself rather than a stochastic decision bound.

How can a decision threshold be adaptively tuned in this circuit? For instance, in a speed–accuracy trade-off, too low a threshold leads to quicker responses but more errors, whereas too high a threshold improves the accuracy but prolongs response times. Neither of these yields maximal rewards. Since in our model the decision threshold is defined as the minimum cortical firing needed to induce a burst response in the downstream SC neurons, one would expect that this threshold could be adjusted by plastic changes in the cortico-collicular pathway: the same level of cortical input to the superior colliculus could be achieved with less firing of cortical neurons, if the synapses of the cortico-collicular projection were stronger. Interestingly, this is not the case when the system is gated by the basal ganglia. This is because neurons in SNr normally fire tonically at a high rate (Figure 31.10), and provide a sustained inhibition to SC movement neurons

(Hikosaka *et al.*, 2000). This inhibition must be released, as ramping activity in the cortex activates CD neurons, which in turn suppress the activity in the SNr, in order for SC neurons to produce a burst output. This highly non-linear disinhibition mechanism implies that the decision threshold is much more readily adjustable by tuning the synaptic strength of the cortico-striatal pathway (Lo and Wang, 2006). Indeed, such an adaptive tuning of the decision threshold is expected to depend on reward signals (Reynolds *et al.*, 2001), and cortico-striatal synapses represent a major target of innervation by dopamine neurons which play a critical role in reinforcement signaling (Reynolds and Wickens, 2002). Our work suggests that dopamine-dependent plasticity of cortico-striatal synapses is a likely neural locus for adaptive tuning of the decision threshold in the brain.

## REWARD-DEPENDENT PLASTICITY AND ADAPTIVE CHOICE BEHAVIOR

### Computation of Returns by Synapses

To describe adaptive decision making, we have incorporated reward-dependent learning in our decision-making model. Consider a local neural network shown in Figure 31.8. Recall that the network’s behavior is described by a softmax decision criterion – that is, the probability of choosing A versus B is a function of the difference in the synaptic strengths  $c_A$  and  $c_B$  for the inputs to the two competing neural pools (Figure 31.9c). Supposing that synaptic connections are plastic, then synaptic modifications will alter the network’s future decision behavior, which in turn will lead to further changes in the synapses (Figure 31.8). Specifically, we used binary synapses that undergo a stochastic Hebbian learning rule, namely that synaptic plasticity depends on coactivation of presynaptic and postsynaptic neurons and takes place stochastically (Fusi, 2002). Specifically, synapses between two neurons are assumed to have two states (Down and Up). Synaptic potentiation (depression) corresponds to a probabilistic transition from the Down to Up (Up to Down) state that depends on the activity of both the pre- and post-synaptic neurons (Hebb, 1949; Fusi, 2002). In addition, it is assumed that synapses for inputs to decision neurons are potentiated only if the choice is rewarded, and depressed otherwise (Soltani and Wang, 2006; Soltani *et al.*, 2006; Fusi *et al.*, 2007). This is inspired by the suggestion that the presence or absence of dopamine signal can reverse

the sign of modification at certain classes of synapses in the brain (Reynolds and Wickens, 2002). Our working hypothesis is that input synapses onto a decision circuit are updated according to such a reward-dependent Hebbian learning rule (see also Seung, 2003). As a result of synaptic modifications, the input strengths for the competing neural groups of the decision network vary from trial to trial, leading to adaptive dynamics of choice behavior.

This synaptic learning rule is a biologically plausible instantiation of reinforcement learning, and the model endowed with such synaptic plasticity is a general one rather than one designed for a particular task. We have tested the model with two types of choice behavior. First, we used the model to investigate a foraging task, in which a subject makes successive choices adaptively in a stochastic environment (Sugrue *et al.*, 2004; see also Chapter 30 of this volume). Second, the model was applied to simulate the behavior in a game-theoretic task that involves dynamic interplay between decision agents (Barraclough *et al.*, 2004), and which was summarized earlier in this chapter. In these tasks, whether a subject's choice yields reward or not depends either on the stochastic environment or on the competing agent. In either case, the model simulates a decision maker whose choice outcomes lead to synaptic plasticity that in turn influences future choices, thereby learning to forage adaptively or play a dynamic game with an opponent. We found that during learning, the synaptic strengths  $c_A$  and  $c_B$  compute returns  $R_A$  and  $R_B$  (the amount of reward per choice) rather than income (the amount of reward per unit time) for the alternative options A and B (Soltani and Wang, 2006). Moreover, because synapses are potentiated or weakened stochastically over time, they are forgetful and behave like a leaky integrator of past choice outcomes. In our model, synaptic integration of past rewards has a time constant of a few trials, and therefore the decision behavior is influenced only by rewards harvested locally in time, in agreement with behavioral (Lee *et al.*, 2004; Sugrue *et al.*, 2004; Lau and Glimcher 2005) and neurophysiological (Figure 31.6; Seo and Lee, 2007; Seo *et al.*, 2007) observations.

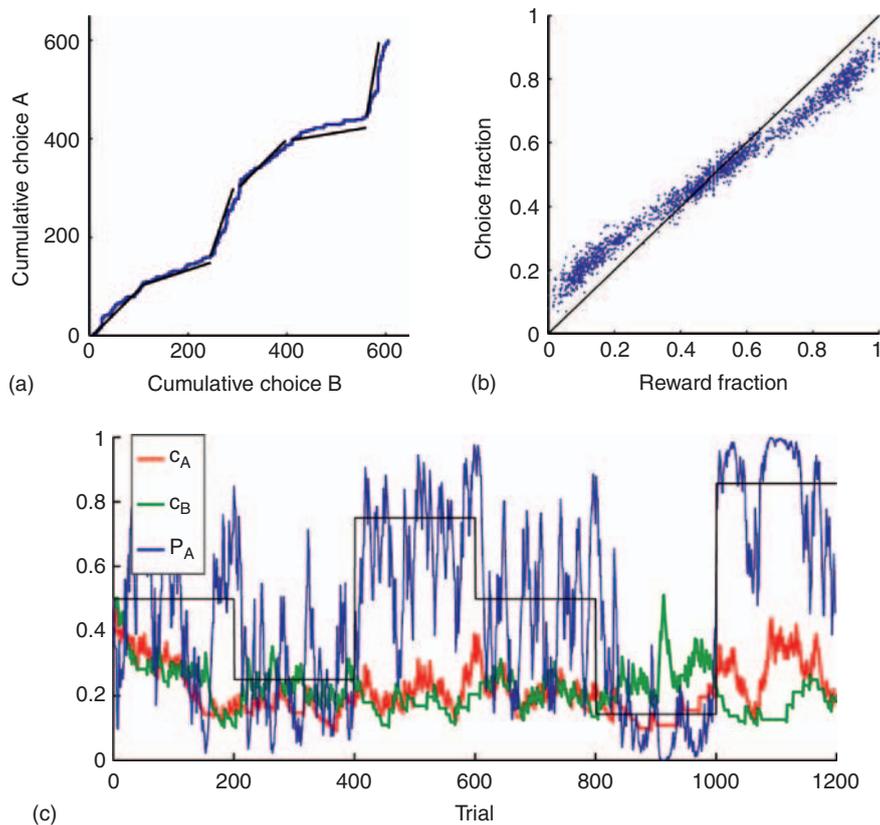
### Matching Law: Melioration Versus Optimization

In foraging tasks commonly used in laboratories, rewards are delivered to two response options stochastically at baiting rates  $\lambda_A$  and  $\lambda_B$ , respectively, according to a particular concurrent reinforcement schedule (Figure 31.1b; Sugrue *et al.*, 2004; Lau and Glimcher, 2005). Behavioral studies using this task have led to Herrnstein's matching law, which states that a subject

allocates his choices in a proportion which matches the relative reinforcement obtained from these choices (Herrnstein *et al.*, 1997). Moreover, the spiking activity of neurons in the lateral intraparietal cortex (LIP) is modulated by a representation of value that the authors defined as fractional income (Sugrue *et al.*, 2004; see also Chapter 30 for details). To explore a cortical circuit mechanism of matching behavior, we have used our neural circuit model of decision making, endowed with reward-dependent synaptic plasticity (Figure 31.8). As shown in Figure 31.12a–b, the model reproduces the matching behavior observed in the monkey experiment. As the reward rate  $\lambda_A/\lambda_B$  varies from one block of trials to the next, the choice behavior of the model changes quickly, so that the probability of choosing A versus B matches approximately  $\lambda_A/\lambda_B$ . We showed analytically that the synaptic strengths ( $c_A$  and  $c_B$ ) are proportional to the returns (reward per choice) of the two targets, namely  $c_A \propto R_A$  and  $c_B \propto R_B$ . Furthermore, the model qualitatively reproduces neural activity observed in LIP that is modulated by the values of the response options (Figure 31.13).

Figure 31.12c shows the probability of choosing option A ( $P_A$ ) along with the input synaptic strengths ( $c_A$  and  $c_B$ ) across six blocks of trials. The process of synaptic plasticity is stochastic, and there is considerable variability within each block of 200 trials. However, on average (indicated by the blue line for  $P_A$ ), the choice probability ratio matches that of rates at which rewards are delivered to the two targets, and this matching behavior is learned through plastic synapses. For instance, if in a block of trials the reward probability  $\lambda_A$  is larger than  $\lambda_B$ , then  $c_A$  is more likely to be potentiated than  $c_B$  through the successive decisions of the network across trials because the return from choosing A is higher, leading to a larger  $P_A$ . The converse occurs in a block of trials where  $\lambda_A$  is smaller than  $\lambda_B$ .

Note that synaptic modifications take place on a trial-by-trial basis, locally in time. There is no prescription in the model for global optimization. The model's performance is close to the matching behavior, which is achieved dynamically through a melioration process – i.e., the model chooses the alternative with a higher return, so that the interplay between decision behavior and synaptic plasticity iteratively improves the total income (reward per unit time) to the maximum possible, given the constraints of the stochastic neuronal and synaptic dynamics (Soltani and Wang, 2006). Moreover, the model also reproduces the observation that, in the monkey experiment, matching is not perfect, and the relative probability of choosing the more rewarding option is slightly smaller than the relative reward rate (“under-matching”) (Figure 31.12b). A model analysis showed that under-matching is a



**FIGURE 31.12** An attractor network model shows matching behavior in a dynamic environment. (a) For one session of the simulated matching experiment, the cumulative choice probability for target A is plotted against the cumulative choice probability for target B. Black straight lines show the baiting probability ratio in each block. The slope of the cumulative plot is equal to the choice ratio, and is approximately equal to the baiting probability ratio. In this session the following baiting probability ratios are used in sequence [1:1, 1:3, 3:1, 1:1, 1:6, 6:1]. (b) Each point shows the block-wise choice fraction as a function of the block-wise reward fraction for a block of trials on which the baiting probabilities are held constant. Most of the points fall close to the diagonal line (perfect matching), but the choice fraction is slightly lower than the reward fraction when the latter is larger than 1/2 (under-matching). (c) The synaptic strengths,  $c_A$  (red) and  $c_B$  (green), and the choice probability (blue) plotted as a function of time. The thin black line indicates the baiting probability ratio in each block. In each block the synaptic strengths fluctuate according to the value of returns from the two choices (not shown).

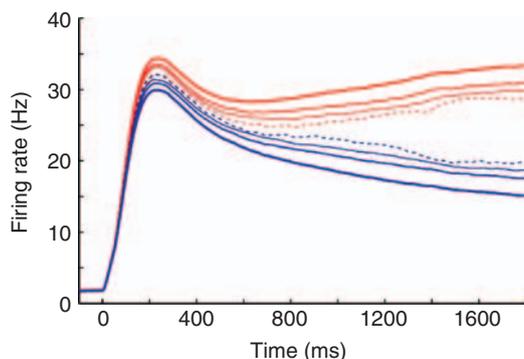
natural consequence of stochasticity in neural activity (Soltani and Wang, 2006).

In summary, our model suggests that although activity of LIP neurons depends on values of response options (Figure 31.13), valuation may occur elsewhere, at the synaptic level and in the form of returns. For the sake of simplicity we have considered a local network model, but we are agnostic regarding the actual site of synaptic plasticity that is critically involved with valuation. Candidate loci include the cortico-striatal connections in the basal ganglia (Lo and Wang, 2006), or synaptic pathways within the orbitofrontal cortex (see Chapter 29).

### Random Choice Behavior in Matching Pennies Game

Our model can be extended to decision making in competitive games between multiple agents, which is

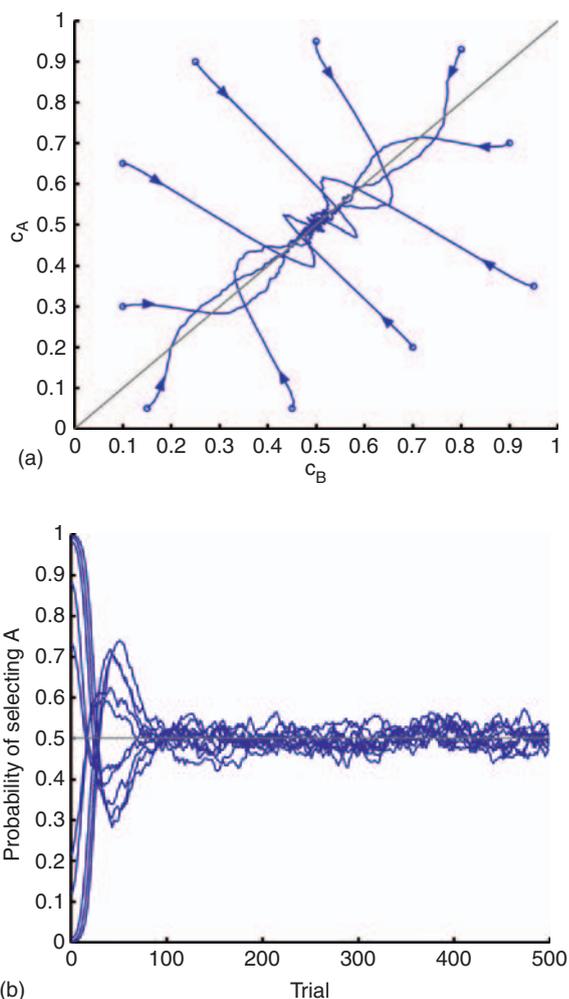
the main topic of the earlier sections of this chapter. The idea is that several such models, each simulating a “decision maker,” can interact according to a payoff matrix. We have used our model to simulate the monkey experiment of Barraclough *et al.* (2004), in which monkeys play matching pennies with a computer opponent that uses three different algorithms (0, 1, and 2, see description above). The model reproduces many salient behavioral observations (Soltani *et al.*, 2006). If the opponent is not interactive (using Algorithm 0), the model decision behavior is idiosyncratic and might, for instance, choose one of the targets exclusively. When the opponent uses algorithm 1, the model exhibits prominent win–stay–lose–switch (WSLS) behavior, as observed in monkeys. Finally, when the opponent uses algorithm 2 and is fully interactive according to the rules of matching pennies, the model behavior becomes quasi-random. This is shown in Figure 31.14, with several different sets of initial



**FIGURE 31.13** Graded activity of model neurons as a function of the input synaptic strengths which encode the values (returns) of choice options in a matching task. The activity of decision neurons shows a graded pattern, if single-trial firing rates are sorted and averaged according to the network's choice and the difference between synaptic strengths. Activity is aligned by the onset of two targets, and is shown separately for the choices corresponding to the neurons' preferred (red) or non-preferred (blue) target. In addition, trials are subdivided into four groups according to the difference between the strength of synapses to the two competing neural populations [ $c_A - c_B = -0.16$  to  $-0.05$  (dashed),  $-0.05$  to  $0$  (thin),  $0$  to  $0.05$  (medium),  $0.05$  to  $0.16$  (thick)].

values for the synaptic variables  $c_A$  and  $c_B$  (Figure 31.14, (a)). Different  $c_A$  and  $c_B$  values yield a different initial probability  $P_A$  of choosing response A versus B (Figure 31.14, (b)). Competitive interaction with the opponent, however, quickly equalizes the synaptic variables (Figure 31.14, (a)), and the choice probability becomes very close to 0.5 (Figure 31.14, (b)), regardless of the initial state of the system. For instance, if initially the system chooses target A more frequently because  $c_A$  is larger than  $c_B$ , it would be exploited by the opponent, and the unrewarded outcomes from choosing A would induce depression of  $c_A$  of the synapses to the neural pool A, so that the difference  $c_A - c_B$  decreases over time, and the system gradually chooses B more frequently.

Interestingly, our model, with a reinforcement learning rule that changes only synapses onto the neurons selective for the chosen option, does not capture all the details of the monkey's behavior. In particular, it shows a probability of WSLS,  $P(\text{WSLS})$ , below a limited value (about 0.65), whereas  $P(\text{WSLS})$  can be nearly 1 in monkeys with algorithm 1. We found that  $P(\text{WSLS}) \sim 1$  can be realized in our model with a different learning rule, according to which synapses onto both neural populations (selective for the chosen and unchosen targets) are modified in each trial. This is akin to a "belief-dependent learning rule" (Camerer, 2003; Lee *et al.*, 2005). It is also in line with the conclusion that both past reinforcers and past choices influence future decision behavior (Lau and Glimcher, 2005).



**FIGURE 31.14** Model simulation of the dynamic choice behavior during the matching pennies game. (a) The synaptic strengths  $c_A$  and  $c_B$  plotted against each other show the time evolution of adaptive learning. (b) The corresponding choice probability for A, which is a softmax function of  $c_A - c_B$ . Regardless of the initial state of the neural circuit (different  $c_A$  and  $c_B$  values), the network quickly evolves towards the Nash Equilibrium state of random mixed strategy, with  $c_A \approx c_B$  (the diagonal line in (a)), and the choice probability becomes chance level (0.5 in (b)).

Although our model can reproduce monkey's behavior obtained with different algorithms, different model parameters are required for each algorithm. How can these model parameters be tuned adaptively, as the opponent's algorithm is changed? To address this question, we incorporated a meta-learning rule proposed by Schweighofer and Doya (2003) that maximizes long-term rewards. We found that the enhanced model captures the very slow changes of the monkey's behavior, as the opponent's algorithm changes from session to session (Soltani *et al.*, 2006).

A general insight of this work is that a decision circuit produces random choice behavior, not necessarily

because the system has a prescribed “random number generator,” but because the trial-to-trial choice dynamics force the decision agent to play randomly. This is well demonstrated in our model, because the same model produces either stereotypical responses or random responses, depending on the behavior of its opponent. The model decision maker does not have the goal to play randomly, but simply tries to play its best, given the environment and other decision agents involved in the game. This conclusion is consistent with previous behavioral studies and models, emphasizing the critical importance of feedbacks in the production of quasi-random behavior (Rapoport and Budescu, 1992; Camerer, 2003). Moreover, our model suggests that irregular neural firing that gives rise to sigmoid decision criterion, and the stochastic nature of synaptic learning, contribute to the generation of random choice behavior, which can be desirable and even optimal in interactive decision tasks. Thus, this model sheds insights into neural processes in the brain underlying the randomness observed at the psychological level (Glimcher, 2005). Therefore, neurobiologically based neural modeling helps to bridge the gap between cognitive behavior and its underlying neural network mechanisms.

## CONCLUSION

Much of the research in behavioral economics focuses on how decision makers choose among various options when the information about the uncertain future prospects are provided explicitly. For example, in studies on decision making under risk, the decision makers are given specific information about the magnitudes and probabilities of possible payoffs from each choice. In real life, however, information about the magnitude, likelihood, and temporal delay of reward and punishment resulting from a particular choice often has to be estimated through experience. Furthermore, such reward contingencies often change over time, and this happens frequently when multiple agents interact. Especially valuable insight into the cognitive processes underlying such adaptive choice behaviors comes from studies on choice behaviors in games. Simple competitive zero-sum games, such as matching pennies and Rock–Paper–Scissors, also provide opportunities to study the underlying neural mechanisms for dynamic decision making, because non-human primates can be easily trained to play such games against a computer-simulated opponent.

Recently, neurophysiological studies have identified different types of signals encoded by individual

neurons in the frontal cortex and the posterior parietal cortex during such computer-simulated competitive games that might be used to determine the choice of the animal and update its strategy based on the previous choice outcomes. These studies have largely focused on the neural correlates of value functions and the signals necessary to update the value functions according to simple reinforcement learning algorithms. In contrast, the behavioral strategies of humans and animals might be also influenced by high-level cognitive factors, such as abstract rules and behavioral contexts, and the knowledge of the dynamic properties of the environment (Hampton *et al.*, 2006). The neural mechanisms responsible for updating the animal’s decision-making strategies based on such multiple sources of information are not well understood, and should be pursued in future research. Future studies also need to elucidate the neural processes involved in integrating different types of costs and benefits during decision making (Roesch *et al.*, 2006; Lee *et al.*, 2007; Rushworth *et al.*, 2007). It might be also possible to investigate the neural processes specialized for social interactions, since non-human primates can be also trained to play relatively simple experimental games in laboratory settings (Chen and Hauser, 2005; Jensen *et al.*, 2007).

The results from neurophysiological recording experiments need to be linked to mechanistic models about how information regarding the outcomes of previous choices is incorporated into a network of spiking neurons, allowing the animal to adjust its decision-making strategies adaptively. We have developed such a model of spiking neurons that is characterized by strongly recurrent or attractor dynamics and endowed with reward-dependent Hebbian synaptic plasticity. These studies have begun to provide important clues as to how adaptive stochastic decision making, such as matching behavior in a foraging task or approximate Nash Equilibrium in a competitive game, result from a dynamic interplay between a decision-making network and its environment. This model will need to be extended to investigate how a neural network or system of networks can optimally combine the information about various aspects of reward and punishment, such as their magnitude, probability, and immediacy. Also, the biophysical basis of reward-dependent plasticity in the brain remains to be elucidated. We expect that progress in this direction will ultimately account for the discrepancy between the choice behaviors of humans and animals, and the rational choice behaviors prescribed by normative economic theories.

## Acknowledgments

We thank Dominic Barraclough and Hyojung Seo for their contributions to the neurophysiological studies, Chung-Chuan Lo and Alireza Soltani for their contributions to the computational modeling and help with the figure and Alireza Soltani for his comments on the manuscript. This research was supported by NIH grants MH073246 (D.L. & X.W.) and NS044270 (D.L.).

## References

- Amemori, K. and Sawaguchi, T. (2006). Contrasting effects of reward expectation on sensory and motor memories in primate prefrontal neurons. *Cerebral Cortex* 16, 1002–1015.
- Amiez, C., Joseph, J.P., and Procyk, E. (2005). Anterior cingulate error-related activity is modulated by predicted reward. *Eur. J. Neurosci.* 21, 3447–3452.
- Amiez, C., Joseph, J.P., and Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. *Cerebral Cortex* 16, 1040–1055.
- Amit, D.J. and Mongillo, G. (2003). Selective delay activity in the cortex: phenomena and interpretation. *Cerebral Cortex* 13, 1139–1150.
- Barbas, H. (1995). Anatomical basis of cognitive-emotional interactions in the primate prefrontal cortex. *Neurosci. Biobehav. Rev.* 19, 499–510.
- Barbas, H. and De Olmos, J. (1990). Projections from the amygdala to basventral and mediodorsal prefrontal regions in the rhesus monkey. *J. Comp. Neurol.* 300, 549–571.
- Barbas, H. and Pandya, D.N. (1987). Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. *J. Comp. Neurol.* 256, 211–228.
- Barbas, H. and Pandya, D.N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 286, 353–375.
- Barraclough, D.J., Conroy, M.L., and Lee, D. (2004). Prefrontal cortex and decision-making in a mixed-strategy game. *Nat. Neurosci.* 7, 404–410.
- Bogacz, R., Brown, E., Moehlis, J. et al. (2006). The physics of optimal decision-making: a formal analysis of models of performance in two-alternative forced choice tasks. *Psychol. Rev.* 113, 700–765.
- Camerer, C.F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, NJ: Princeton University Press.
- Carmichael, S.T. and Price, J.L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *J. Comp. Neurol.* 371, 179–207.
- Chafee, M.V. and Goldman-Rakic, P.S. (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J. Neurophysiol.* 79, 2919–2940.
- Chen, M.K. and Hauser, M. (2005). Modeling reciprocity and cooperation in primates: evidence for a punishing strategy. *J. Theoretical Biol.* 235, 5–12.
- Constantinidis, C., Franowicz, M.N., and Goldman-Rakic, P.S. (2001). The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nat. Neurosci.* 4, 311–316.
- Dorris, M.C. and Glimcher, P.W. (2004). Activity in the posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44, 365–378.
- Douglas, R.J. and Martin, K.A.C. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451.
- Fusi, S. (2002). Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. *Biol. Cybern.* 87, 459–470.
- Fusi, S., Asaad, W., Miller, E.K., and Wang, X.-J. (2007). A neural model of flexible sensori-motor mapping: learning and forgetting on multiple timescales. *Neuron* 54, 319–333.
- Fuster, J.M. (1997). *The Prefrontal Cortex: Anatomy, Physiology and Neuropsychology of the Frontal Lobe*. Philadelphia, PA: Lippincott-Raven Publishers.
- Ghashghaie, H.T., Hilgetag, C.C., and Barbas, H. (2007). Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *NeuroImage* 34, 905–923.
- Glimcher, P.W. (2005). Indeterminacy in brain and behavior. *Annu. Rev. Psych.* 56, 25–56.
- Gold, J.I. and Shadlen, M.N. (2007). The neural basis of decision-making. *Annu. Rev. Neurosci.* 30, 535–574.
- Hampton, A.N., Bossaerts, P., and 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, 8360–8367.
- Hastie, H., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.
- Hebb, D.O. (1949). *Organization of Behavior*. New York, NY: Wiley.
- Herrnstein, R.J., Rachlin, H., and Laibson, D.I. (1997). *The Matching Law: Papers in Psychology and Economics*. Cambridge, MA: Harvard University Press.
- Hikosaka, O., Takikawa, Y., and Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiol. Rev.* 80, 953–978.
- Hoshi, E., Sawamura, H., and Tanji, J. (2005). Neurons in the rostral cingulate motor area monitor multiple phases of visuomotor behavior with modest parametric selectivity. *J. Neurophysiol.* 94, 640–656.
- Jensen, K., Call, J., and Tomasello, M. (2007). Chimpanzees are rational maximizers in an ultimatum game. *Science* 318, 107–109.
- Kahneman, D. and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–292.
- Kandel, E.R., Schwartz, J.H., and Jessel, T.M. (2000). *Principles of Neural Science*. New York, NY: McGraw-Hill.
- Kobayashi, S., Lauwereyns, J., Koizumi, M. et al. (2002). Influence of reward expectation on visuomotor processing macaque lateral prefrontal cortex. *J. Neurophysiol.* 87, 1488–1498.
- Lau, B. and Glimcher, P.W. (2005). Dynamic response-by-response models of matching behavior in rhesus monkeys. *J. Exp. Anal. Behav.* 84, 555–579.
- Lee, D., Conroy, M.L., McGreevy, B.P., and Barraclough, D.J. (2004). Reinforcement learning and decision-making in monkeys during a competitive game. *Cogn. Brain Res.* 22, 45–58.
- Lee, D., McGreevy, B.P., and Barraclough, D.J. (2005). Learning and decision-making in monkeys during a rock-paper-scissors game. *Cogn. Brain Res.* 25, 416–430.
- Lee, D., Rushworth, M.F.S., Walton, M.E. et al. (2007). Functional specialization of the primate frontal cortex during decision-making. *J. Neurosci.* 27, 8170–8173.
- Leon, M.I. and Shadlen, M.N. (1999). Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24, 415–425.
- Lewis, D.A., Melchitzky, D.S., Sesack, S.R. et al. (2001). Dopamine transporter immunoreactivity in monkey cerebral cortex: regional, laminar, and ultrastructural localization. *J. Comp. Neurol.* 432, 119–136.
- Lo, C.C. and Wang, X.-J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat. Neurosci.* 9, 956–963.

- Luppino, G., Rozzi, S., Calzavara, R., and Matelli, M. (2003). Prefrontal and agranular cingulate projections to the dorsal premotor areas F2 and F7 in the macaque monkey. *Eur. J. Neurosci.* 17, 559–578.
- Machens, C.K., Romo, R., and Brody, C.D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* 18, 1121–1124.
- Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.* 10, 647–656.
- Miller, P. and Wang, X.-J. (2006). Inhibitory control by an integral feedback signal in prefrontal cortex: a model of discrimination between sequential stimuli. *Proc. Natl Acad. Sci. USA*, 103, 201–206.
- Nash, J.F., Jr. (1950). Equilibrium points in  $n$ -person games. *Proc. Natl Acad. Sci. USA* 36, 48–49.
- Niki, H. and Watanabe, M. (1979). Prefrontal and cingulate unit-activity during timing behavior in the monkey. *Brain Res.* 171, 213–224.
- Pawitan, Y. (2001). In *All Likelihood: Statistical Modelling and Inference using Likelihood*. Oxford: Oxford University Press.
- Rapoport, A. and Budescu, D.V. (1992). Generation of random series in two-person strictly competitive games. *J. Exp. Psych. Gen.* 121, 352–363.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev.* 85, 59–108.
- Reynolds, J.N. and Wickens, J.R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks* 15, 507–521.
- Reynolds, J.N., Hyland, B.I., and Wickens, J.R. (2001). A cellular mechanism of reward-related learning. *Nature* 413, 67–70.
- Roesch, M.R., Taylor, A.R., and Schoenbaum, G. (2006). Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. *Neuron* 51, 509–520.
- Roitman, J.D. and Shadlen, M.N. (2002). Responses of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* 22, 9475–9489.
- Rushworth, M.F.S., Behrens, T.E.J., Rodebeck, P.H., and Walton, M.E. (2007). Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. *Trends Cogn. Sci.* 11, 168–176.
- Schall, J.D. (2001). Neural basis of deciding, choosing and acting. *Nat. Rev. Neurosci.* 2, 33–42.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27.
- Schweighofer, N. and Doya, K. (2003). Meta-learning in reinforcement learning. *Neural Networks* 16, 5–9.
- Seo, H. and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* 27, 8366–8377.
- Seo, H., Barraclough, D.J., and Lee, D. (2007). Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cerebral Cortex* 17, i110–117.
- Seung, H.S. (2003). Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron* 40, 1063–1073.
- Smith, P.L. and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci.* 27, 161–168.
- Soltani, A. and Wang, X.-J. (2006). A biophysically based neural model of matching law behavior: melioration by stochastic synapses. *J. Neurosci.* 26, 3731–3744.
- Soltani, A., Lee, D., and Wang, X.-J. (2006). Neural mechanism for stochastic behavior during a competitive game. *Neural Networks* 19, 1075–1090.
- Sugrue, L.P., Corrado, G.S., and Newsome, W.T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science* 304, 1782–1787.
- Sutton, R.S. and Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
- Usher, M. and McClelland, J.L. (2001). On the time course of perceptual choice: the leaky competing accumulator model. *Psychol. Rev.* 108, 550–592.
- Vogt, B.A. and Pandya, D.N. (1987). Cingulate cortex of the rhesus monkey: II. Cortical afferents. *J. Comp. Neurol.* 262, 271–289.
- von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.
- Wald, A. and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Stats.* 19, 326–337.
- Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* 24, 455–463.
- Wang, X.-J. (2002). Probabilistic decision-making by slow reverberation in cortical circuits. *Neuron* 36, 955–968.
- Wang, Y., Shima, K., Sawamura, H., and Tanji, J. (2001). Spatial distribution of cingulate cells projecting to the primary, supplementary, and pre-supplementary motor areas: a retrograde multiple labeling study in the macaque monkey. *Neurosci. Res.* 39, 39–49.
- Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. *Nature* 382, 629–632.
- Williams, S.M. and Goldman-Rakic, P.S. (1993). Characterization of the dopaminergic innervation of the primate frontal cortex using a dopamine-specific antibody. *Cerebral Cortex* 3, 199–222.
- Wong, K.-F. and Wang, X.-J. (2006). A recurrent network mechanism for time integration in perceptual decisions. *J. Neurosci.* 26, 1314–1328.