

Cognitive Critique



GAME THEORY IN NEUROSCIENCE

HYOJUNG SEO

*Department of Neurobiology
Yale University School of Medicine, New Haven, CT*

EMAIL: hyojung.seo@yale.edu

TIMOTHY J. VICKERY

*Department of Psychology
Yale University, New Haven, CT*

EMAIL: tim.vickery@gmail.com

DAEYEOL LEE

*Department of Neurobiology
Department of Psychology
Yale University, New Haven, CT*

EMAIL: daeyeol.lee@yale.edu

Accepted December 31, 2011

KEYWORDS

game theory, reinforcement learning, reward, decision making, basal ganglia, prefrontal cortex

ABSTRACT

Decisions made during social interaction are complex due to the inherent uncertainty about their outcomes, which are jointly determined by the actions of the decision maker and others. Game theory, a mathematical analysis of such interdependent decision making, provides a computational framework to extract core components of complex social situations and to analyze decision making in terms of those quantifiable components. In particular, normative prescription of optimal strategies can be compared to the strategies actually used by humans and animals, thereby providing insights into the nature of observed deviations from prescribed strategies. Here, we review the recent advances in decision neuroscience based on game theoretic approaches, focusing on two major topics. First, a number of studies have uncovered behavioral and neural mechanisms of learning that mediate adaptive decision making during dynamic interactions among decision agents. We highlight multiple learning systems distributed in the cortical and subcortical networks supporting different types of learning during interactive games, such as model-free reinforcement learning and model-based belief learning. Second, numerous studies have investigated the role of social norms, such as fairness, reciprocity and cooperation, in decision making and their representations in the brain. We predict that in combination with sophisticated manipulation of socio-cognitive factors, game theoretic approaches will continue to provide useful tools to understand multifaceted aspects of complex social decision making, including their neural substrates.

INTRODUCTION

The brain is fundamentally the organ of decision making. It processes the incoming sensory information through several different modalities in order to identify the current state of the animal's environment, but this information is of no use unless it can be utilized to select an action that produces the most desirable outcome for the animal. This process of selecting the most appropriate action would be trivial, if the relationship between a given action and its outcome was fixed and did not change through evolution. In such cases, the problem might be solved most efficiently simply by hard-wiring the relationship between the animal's state and the action that produces the best outcome, as in various reflexes. For example, if a puff of air is applied to the eyes, the best response would be to

close the eye lids to prevent any damage to the cornea. In real life, however, the relationship between a particular action and its possible outcome in a given environment might change over time, and therefore the animal is often required to modify its decision making strategy, namely, the probability of taking a given action, given the animal's current environment and its knowledge of the dynamics of that environment.

The most challenging situation arises when the decision maker has to make a decision in a social setting. In this case, the outcome of an action is not only determined by the animal's environment and the action chosen by the animal, but also by the actions of other animals in the same environment. Social interactions are common for all animals, are manifested in various forms, ranging from mating to predation, and can be competitive or cooperative. Given the pervasive nature of social decision making, it is possible that the brain areas involved in decision making might have adapted to specific computational demands unique to social decision making. In fact, during the last decade, a large number of studies have investigated the brain mechanisms involved in social cognition, including those that play important roles during social decision making (Wolpert et al. 2003; Lee 2008; Rilling and Sanfey 2011). In many of these studies, researchers have focused on behavioral tasks that have been frequently analyzed using game theory. These game theoretic tasks have several advantages. First, rules of many games are relatively simple, although the behaviors elicited by them can be quite complex. As a result, these games can be described relatively easily to human subjects, and some animals, especially non-human primates, can be trained to perform virtual game theoretic tasks against computer opponents (Lee and Seo 2007; Seo and Lee 2008). Second, many studies have investigated the neural activity during repeated or iterative games, in which the subject plays the same game repeatedly, either against the same players or against different players across different trials. This is analogous to many real-life situations, and also provides an interesting opportunity to study the mechanisms of learning at work during social interactions (Fudenberg and Levine 1998). Finally, the use of game theoretic tasks in neurobiological studies also benefits from the rich literature on theoretical and behavioral studies. For many games, optimal strategies for rational players are known. In addition, a large number of behavioral studies have characterized how humans and animals deviate from such optimal strategies. In this review article, we first

provide a brief introduction to game theory. We then describe several different learning theories that have been applied to account for dynamic choice behaviors of humans and animals during iterative games. We also review the recent findings from neurophysiological and neuroimaging studies on social decision making that employ game theoretic tasks. We conclude with suggestions for important research questions that need to be addressed in future studies.

GAME THEORY

Game theory, introduced by von Neuman and Morgenstern (1944), is a branch of economics that analyzes social decision making mathematically. In this theory, a game is defined by a set of players, each choosing a particular action from a set of alternatives. An outcome is determined by the choices made by all players, and a particular outcome determines the payoff to each player. This is commonly summarized in a payoff matrix. For example, the payoff matrices for the matching pennies and prisoner's dilemma games are shown in Figure 1.

A. Matching Pennies Game

		Non-matcher	
		Head	Tail
		Head	(1, -1)
Matcher	Head	(-1, 1)	(1, -1)
	Tail	(-1, 1)	(1, -1)

B. Prisoner's Dilemma game

		Player II	
		Cooperate	Defect
		Cooperate	(3, 3)
Player I	Cooperate	(0, 5)	(5, 0)
	Defect	(1, 1)	

Figure 1. Payoff matrix for a matching pennies game (A) and prisoner's dilemma game (B). A pair of numbers within the parentheses indicate the payoffs from each combination of choices for the row and column players, respectively.

For each of these games, two players face the same choice. In the matching pennies game, both players choose between heads and

tails of a coin, whereas in the prisoner's dilemma game, they choose between cooperation and defection. In the matching pennies game, one of the players (matcher) wins if the two players choose the same option, and loses otherwise. The sum of the payoffs for the two players is always zero, namely, one player's gain is always the other player's loss. Hence, matching pennies is an example of a strictly competitive, zero-sum game. In contrast, the sum of payoffs to the two players is not fixed for the prisoner's dilemma game.

A strategy refers to the probability of taking each action, and it can be fixed or mixed. A fixed strategy refers to selecting a particular action with certainty, whereas a mixed strategy refers to selecting multiple strategies with positive probabilities, and therefore can be understood as a probability distribution over an action space. If the decision maker knows the strategies of all the other players, then he or she can compute the expected value of the payoff for each option. The optimal strategy in this case, referred to as the best response, would then be to choose the option with the maximum expected value of the payoff. For example, in the case of the matching pennies game shown in Figure 1A, if the matcher knows that the non-matcher chooses the heads and tails with probabilities of 0.2 and 0.8, respectively, then the expected value of the payoffs for choosing the heads and tails for the matcher would be $-0.6 (=0.2x1+0.8x(-1))$ and $0.6 (=0.2x(-1)+0.8x1)$, respectively. However, it is clear that this scenario is not rational for the non-matcher, since if the matcher adopts the best response strategy against this particular strategy of the non-matcher, the non-matcher would lose in the majority of the cases. This consideration leads to the concept of an equilibrium strategy. The so-called Nash equilibrium refers to a set of strategies defined for all players in a particular game such that no player can increase their payoff by deviating from their strategy unilaterally (Nash 1950). In other words, a Nash equilibrium consists of a set of strategies in which every player's strategy is the best response to the strategy of everyone else. For the matching pennies task, there is a unique Nash equilibrium, which is for each player to choose the heads and tails with equal probabilities. Therefore, the matching pennies game is an example of a mixed-strategy game, which refers to games in which the optimal strategy is mixed. There is also a unique Nash equilibrium for the prisoner's dilemma game, which is to defect. The outcome of mutual defection in the prisoner's dilemma game is worse for both players than that of mutual cooperation, leading to a dilemma.

Although game theory makes clear predictions about the choices of rational players in a variety of games, human subjects frequently deviate from these predictions (Camerer 2003). There are two possible explanations. One possibility is that human subjects are cognitively incapable of identifying the optimal strategy. This possibility is supported by the fact that for many simple games, the strategies of human subjects tend to approach the equilibrium strategies over time (Camerer 2003). Another possibility is that a key assumption in game theory about the self-interested, rational player is not entirely true (Fehr and Fischbacher 2003). In fact, as discussed below, humans and some animals are not entirely selfish and might behave in some cases to improve the well-being of other individuals.

LEARNING THEORIES FOR ITERATIVE GAMES

Human and animal behaviors are constantly shaped by their experience, and social decision making is not an exception to this rule. Therefore, in order to understand how humans and animals change their strategies during social interactions through experience, it is important to understand the nature of learning algorithms utilized by decision makers during iterative games. A powerful theoretical framework within which to study the process of learning and decision making is the Markov decision process (MDP), which is based on the assumption that the outcome of a decision is determined by the current state of the decision maker's environment and action (Sutton and Barto 1998). In this framework, commonly referred to as reinforcement learning, the probability of taking an action a_i is largely determined by its value function V_i . This probability, denoted as $p(a_i)$, increases with V_i , but the decision maker does not always choose the action with the maximum value function. Instead, actions with smaller value functions are chosen occasionally to explore the decision maker's environment. Commonly, $p(a_i)$ is given by the softmax transformation of the value functions

$$p(a_i) = \exp(\beta V_i) / \sum_j \exp(\beta V_j) \quad (1)$$

where β denotes the inverse temperature that controls the randomness in action selection and hence the degree of exploration.

The strategy or policy of a decision maker, denoted by $p(a_i)$, changes through experience, because the value functions are adjusted by the outcomes of past decisions. These learning algorithms can be divided roughly into two categories (Sutton and Barto 1998; Camerer 2003). First, in the so-called simple or model-free reinforcement learning algorithms, the value functions are adjusted only according to the discrepancy between the actual outcome and expected outcome. Therefore, after choosing a particular action a_i , only its value function is updated as follows:

$$V_i(t+1) = V_i(t) + \alpha \{ r_t - V_i(t) \} \quad (2)$$

where r_t denotes the reward or payoff from the chosen action and α corresponds to the learning rate. The value functions for the remaining actions are not updated. Second, in the so-called model-based reinforcement learning algorithms, the value functions can be adjusted more flexibly not only according to the outcomes of previous actions, but also on the basis of the decision maker's internal or forward model of his or her environment. For games, this corresponds to the player's model or belief about the strategies of other players. Thus, for model-based reinforcement learning or belief learning, it is the previous choices of other players that drive the process of learning by influencing the model or belief about the likely actions of other players. Previous studies on iterative games have largely found that the simple reinforcement learning accounts for choices of human subjects better than the model-based reinforcement learning or belief learning (Mookherjee and Sopher 1997; Erev and Roth 1998; Feltovich 2000). Similarly, choice behaviors of monkeys playing a virtual rock-paper-scissors game against a computer opponent were more consistent with the simple reinforcement learning algorithm than the belief learning algorithm (Lee et al. 2005).

Although the simple and model-based reinforcement learning algorithms have important differences, both can be understood as processes in which the value functions for different actions are adjusted through experience. Unlike the simple reinforcement learning model, the objective of the belief learning is to estimate the strategies of other players. The new observation about the choices of other players can then be translated into incremental changes in the value functions of various actions. For example, imagine that during an iterative prisoner's dilemma, player I has just observed that player II has cooperated (Figure 1B). This might strengthen player I's belief that player II is more likely to cooperate in the

future. Accordingly, player I's value functions for cooperation and defection might become closer to player I's payoffs for these two actions expected from player II's cooperation (3 and 5, respectively, for the payoff matrix shown in Figure 1B). As this example illustrates, the model-based reinforcement learning or belief learning can be implemented by updating the value functions for all actions of a particular player, including those of unchosen actions, according to the hypothetical outcome expected from the choices of other players. In other words, for belief learning, the value function for each action, a_i , can be updated as follows:

$$V_i(t+1) = V_i(t) + \alpha \{ h_t - V_i(t) \} \quad (3)$$

where h_t refers to the hypothetical outcome that would have resulted from action a_i given the choices of all the other players in the current trial.

The fact that both simple and model-based reinforcement learning during iterative games can be described by the same set of value functions has led to the insight that these two different learning models correspond to two extreme cases on a continuum. Accordingly, a more general type of learning algorithm that includes a pure, simple reinforcement learning algorithm and a pure belief learning algorithm has been proposed. This hybrid learning model was originally referred to as an experience-weighted-attraction (EWA) model (Camerer and Ho 1999). In this hybrid learning model, the value functions for all actions are updated after each round, but the value function for the chosen action and those for all the other actions can be updated at different learning rates. Behavioral studies in both humans and monkeys have found that this hybrid model can account for the behaviors observed during iterative games better than simple reinforcement learning or belief learning models (Camerer and Ho 1999; Lee et al. 2005; Abe and Lee 2011).

LEARNING OPTIMAL STRATEGIES DURING EXPERIMENTAL GAMES

Although analytical game theory prescribes optimal strategy to maximize expected utility, it does not predict how this equilibrium strategy can be reached when the games are actually played. This leads to a number of important empirical questions. For example,

do the actual strategies of human subjects conform to the normative prediction of game theory? What information is actually attended and utilized during the equilibration process? Are there algorithms that ensure that choices of players converge on the equilibrium strategies through iterative interactions?

A number of empirical studies have found that during simple two-person mixed-strategy games (e.g. zero-sum games) the choice strategy adopted by individual players frequently deviates from the optimal probabilistic mixture of alternative actions as prescribed by the game theory (Malcolm and Lieberman 1965; O'Neil 1987; Brown and Rosenthal 1990; Rapoport and Boebel 1992; Mookerjee and Sopher 1994, 1997; Ochs 1995). Aggregate choice probabilities averaged over time across subjects often exhibited a deviation from the prediction of optimal mixed strategies, even when all the players had complete information on the payoff structure of the game that they were playing. Even in cases where the large-scale choice probabilities approximated the equilibrium mixture, the actual sequences of successive choices produced by individual subjects were often incompatible with sequences predicted by random samples from a multinomial distribution. Systematic deviations from equilibrium strategies commonly found in these experimental studies often included serial dependence of individual player's choice on the history of past choices and subsequent payoffs of their own as well as of the opponents.

Using simulations based on extensive datasets from previous studies, Erev and Roth (1998) demonstrated that seemingly heterogeneous choice trajectories among individual players over repeated rounds could be effectively captured by a simple reinforcement learning model, if initial choice propensities of the model were set to reflect actual biases or beliefs peculiar to individual players. Interestingly, they also showed that simple reinforcement learning can provide a reasonable account for the deviation from equilibrium strategy that has been consistently found in iterative ultimatum games as well.

Although reinforcement learning successfully simulated choice strategies during competitive games, how people form beliefs based on other players' strategies and dynamically adjust them through repeated interactions has been experimentally investigated using non-competitive games in which there are multiple equilibrium strategies and therefore, cooperation plays a more important role to maximize the payoff than in strictly competitive games (Colman

1999; Camerer 2003). Early models of belief learning proposed that people chose the best response given the last choice (i.e., Cournot dynamics) or by averaging all the past choices (i.e., fictitious play) of the other players (Camerer 2003). However, more recent studies have shown that subjects actually took a strategy intermediate between these two extreme forms of belief learning, averaging the past choices of other players with greater emphasis on their more recent choices (Cheung and Friedman 1997). A coordination game has an advantage in experimentally investigating the dynamics of belief learning, since it makes it possible to estimate the player's belief on other players' strategies directly. For example, in order-statistic games, each of a group of players choose a number and the payoff to each player is determined by the deviation of the number chosen by that particular player from an order-statistic (e.g. median, minimum, mean) of all numbers chosen by the entire group of players. Using the EWA learning model, Camerer and Ho (1999) showed that the actual learning algorithm utilized by subjects during coordination games falls between strict reinforcement and belief learning.

Previous studies have also investigated whether animals are capable of learning and playing, using an optimal strategy during iterative interactions with another agent. For example, rhesus monkeys use approximately equilibrium strategies during competitive games, such as matching pennies games (Barraclough et al. 2004; Lee et al. 2004; Thevarajah et al. 2009) and inspection games (Dorris and Glimcher 2004). In some of these studies (Barraclough et al. 2004; Lee et al. 2004), rhesus monkeys played a two-choice, iterative matching pennies game against a computerized opponent, in which animals (matcher) were rewarded only when they made the same choice as the computer opponent. By systematically manipulating the strategy that was used by the computer opponent, these studies tested how rhesus monkeys dynamically adjusted their choice strategy in response to the changing strategy of their computer opponent. The unique equilibrium strategy for both players in this game is to choose the two options randomly with equal probabilities. When the computer opponent simulated an equilibrium-strategy player regardless of the animal's strategy (algorithm 0), choice patterns of animals were far from the equilibrium strategy, sometimes revealing strong and idiosyncratic bias toward one of the two alternative choices. However, this is not irrational, since once one player uses an equilibrium strategy, the expected payoffs

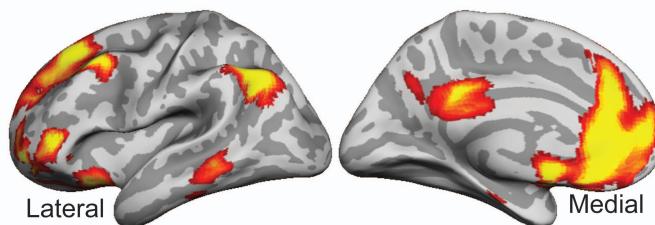
for the other player is equal for all strategies (Colman 1999). In the next stage of the experiment, the computer opponent started using a more exploitative strategy. In each trial, the computer opponent made a prediction about the animal's choice according to the animal's choice history, and chose its action with a higher expected payoff more frequently (algorithm 1). Such an exploitative strategy has often been found in human subjects (Colman 1999). Following the adoption of this exploitative strategy by the computer opponent, the monkey's choice behavior rapidly approached an equilibrium strategy in terms of aggregate probabilities. However, at the scale of individual choices, monkeys frequently used a win-stay/lose-switch strategy, revealing a strong dependence of choice strategy on the past history of their own choice and its resulting payoff. Accordingly, as in human subjects, choice patterns of animals were well captured by a simple reinforcement learning model (Lee et al. 2004; Seo and Lee 2008). Finally, this strong serial correlation between the animal's choice and the conjunction of its past choice and subsequent outcome was further reduced in response to the computer's strategy, which predicted the animal's next move by exploiting the past history of the monkey's choice and subsequent payoff (algorithm 2). Against this highly exploitative opponent, monkeys were capable of increasing overall randomness and decreasing the predictability in their choices by integrating the past payoffs over a longer time scale, rather than depending on a deterministic win-stay/lose-switch strategy (Seo and Lee 2008).

Using a competitive inspection game, Dorris and Glimcher (2004) also demonstrated that monkeys dynamically changed their choice strategies through iterative interaction against the computerized player. They systematically manipulated the optimal mixed-strategy of monkeys by changing the inspection cost and thus the payoff matrix of this experimental game. They showed that, against the computer with an exploitative strategy based on the animal's past choice history, aggregate probability of making a risky choice increased with inspection cost, approximately conforming to the equilibrium strategy prescribed by the game theory. An animal's choice probability was consistent with matching behavior which allocates choices according to the ratio of income obtained from two alternative actions (Herrnstein 1961), suggesting that monkeys reached an equilibrium strategy by dynamically adjusting their choice according to the change in the resulting payoff (Sugrue et al. 2004, 2005).

NEURAL MECHANISM FOR MODEL-FREE REINFORCEMENT LEARNING IN COMPETITIVE GAMES

Neural signals representing the utility or subjective desirability of anticipated and experienced decision outcomes are broadly distributed throughout cortical and sub-cortical networks (Schultz et al. 2000; Padoa-schioppa and Assad 2006; Wallis and Kennerley 2010). In particular, when the blood-oxygen-level-dependent (BOLD) signals related to the choice outcomes during the matching pennies game were analyzed using a multi-voxel pattern analysis (MVPA), reward-related signals were detected practically throughout the entire brain (Figure 2; Vickery et al. 2011).

A. GLM



B. MVPA

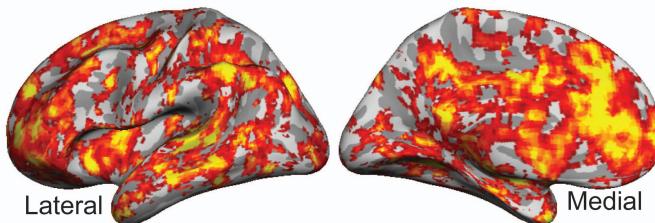


Figure 2. Voxels in the human brain (left hemisphere) showing significant reward-related modulations during a matching pennies task. A. Results obtained using the generalized linear model (GLM) designed to detect the mean changes in the local BOLD signals. B. Results from a multi-voxel pattern analysis designed to detect changes in the spatial pattern of BOLD signals (Vickery et al. 2011).

In addition, signals related to outcomes expected from specific actions or their value functions are often localized in the same brain regions involved in action preparation and execution, such as the dorsolateral prefrontal cortex, the mediofrontal and premotor

regions, and the lateral parietal cortex, as well as the basal ganglia (Platt and Glimcher 1999; Barraclough et al. 2004; Dorris and Glimcher 2004; Sugrue et al. 2004; Samejima et al. 2005; Lau and Glimcher 2008; Seo and Lee 2009; So and Stuphorn 2010; Cai et al. 2011; Pastor-Bernier and Cisek 2011). These regions are commonly connected to the final motor pathway, directly or indirectly. Therefore, signals related to action values can exert influence on action selection.

A number of studies have highlighted the function of the basal ganglia in computing the values of specific actions on the basis of past choices and their outcomes, although this is less well understood compared to the problem of value representation and action selection. Many models of the basal ganglia emphasize the role of dopamine as a teaching signal that drives the process of learning the values of different actions (Montague et al. 1996; O'Doherty 2004; Houk et al. 2005; Daw and Doya 2006). In particular, phasic activity of dopamine neurons resembles theoretical reward prediction error, which is a core computational signal that can update value function in temporal difference models (Schultz 2006). Therefore, dopamine-dependent modification of corticostriatal synaptic efficacy might contribute to updating value functions in the striatum (Reynolds et al. 2001; Houk et al. 2005; Shen et al. 2008; Gerfen and Surmeier 2011). This all-purpose machinery of reinforcement learning might also underlie the adaptive changes in decision making strategies during iterative games. For example, a network of neurons equipped with reward-dependent synaptic plasticity and attractor dynamics can simulate choice behaviors of monkeys observed during a computer-simulated matching pennies game (Soltani et al. 2006).

In addition to the signals related to the values of alternative actions and reward prediction errors, neurophysiological studies have also found that neurons distributed throughout the primate cortex encode multiple types of signals related to the animal's previous choice and reward history. These signals might reflect the neural building blocks used to update the value functions. For example, during the matching pennies task, neurons in the dorsolateral and medial prefrontal cortex, dorsal anterior cingulate cortex, as well as the lateral intraparietal cortex often encoded the information about the animal's choices and payoffs in previous trials (Figure 3; Barraclough et al. 2004; Seo and Lee 2007; Seo and Lee 2008; Seo and Lee 2009; Seo et al. 2009). The mnemonic signals encoding

recent choices can serve as eligibility traces, which can potentially be utilized to update value functions, when they are appropriately combined with the signals related to payoffs resulting from those past choices (Sutton and Barto 1998).

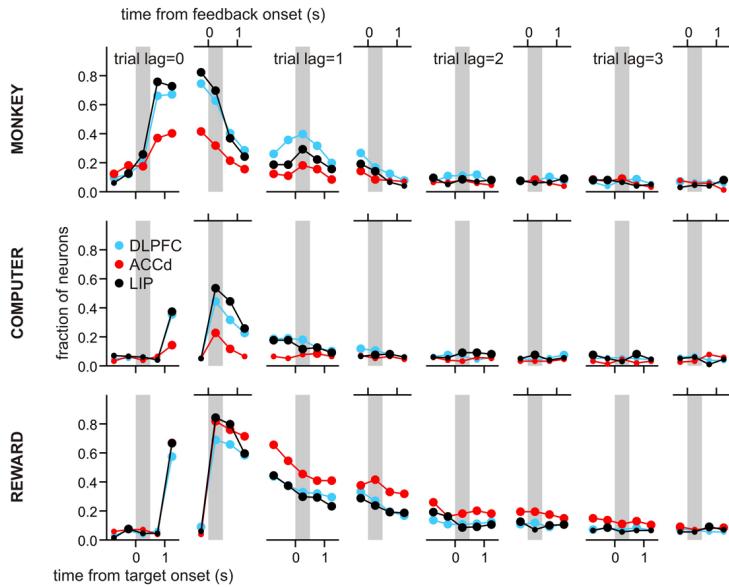


Figure 3. Time course of signals related to the animal’s choice (top), choice of the computer opponent (middle), and choice outcome (bottom) in three different cortical areas (DLPFC, dorsolateral prefrontal cortex; ACC, anterior cingulate cortex; LIP, lateral intraparietal cortex) during a matching pennies game. Figures in each row represent the fraction of neurons significantly modulating their activity according to choices or outcomes in the current (trial lag=0) or previous (trial lag=1 to 3) trials (Seo and Lee 2008). Note that the computer’s choice is equivalent to the conjunction of the animal’s choice and its outcome. The results for each trial lag are shown in two sub-panels showing the proportion of neurons in each cortical area modulating their activity significantly according to the corresponding factor relative to the time of target onset (left panels) or feedback onset (right panels). Large symbols indicate that the proportion of neurons was significantly higher than the chance level (binomial test, $p<0.05$). Gray background corresponds to the delay period (left panels) or feedback period (right panels).

Consistent with this possibility, neurons in the dorsolateral and medial prefrontal cortex often modulated their activity by the conjunction of past outcomes and the preceding choice that led to that particular outcome (Figure 3; Barraclough et al. 2004; Seo and Lee 2009). Consequently, the activity of these neurons was often correlated with the difference in the experienced values of alternative choices and thus signaled relative desirability of a specific action. Therefore, as predicted by the reinforcement learning theory, retrospective neural signals related to the animal's previous choices and outcomes might contribute to computing and updating the value functions and to action selection. In addition, the ability to identify a particular choice responsible for a given outcome is often critical for adaptive decision making, especially when action-outcome contingencies change frequently (Walton et al. 2010). Temporal neural traces related to the animal's previous choices and outcomes might be especially effective when the volatility of the animal's environment increases as a result of social interactions (Behrens et al. 2007, 2008, 2009). An important property of an equilibrium strategy in any game is that the expected utility should be equal for all actions at equilibrium. Indeed, Dorris and Glimcher (2004) found that during an inspection game, neuronal activity in the lateral intraparietal cortex (LIP) was relatively constant across different blocks of a trial, while the probability of monkeys choosing a risky option varied with different inspection costs. Therefore, these neurons might reflect the subjective desirabilities of two alternative actions during a steady state after monkeys reached an equilibrium strategy. However, neural activity was also correlated with the fluctuations in the payoffs expected from the recent choice history of the opponent on a trial-by-trial basis, suggesting that neural activity in the LIP might also represent the value functions of alternative actions on the basis of recent choice and outcome history (Dorris and Glimcher 2004; Sugrue et al. 2004, 2005). Parietal regions have also been implicated in decision making during mixed-strategy games for humans (Vickery and Jiang 2009), as well as in matching pennies games in which monkeys had to make a cross-modal choice between reaching vs. simply directing a gaze toward a visual target (Cui and Andersen 2007).

NEURAL MECHANISM FOR MODEL-BASED REINFORCEMENT LEARNING IN COMPETITIVE GAMES

In belief learning, agents form predictions on the next move of the other players based on their past choices, and this is equivalent to model-based learning in terms of reinforcement learning (Sutton and Barto 1998; Lee et al. 2005; Abe and Lee 2011). In model-based reinforcement learning, value functions of all actions can be potentially updated, even for unchosen actions, whenever the decision maker makes a new observation of the choices made by other players and, as a result, modifies his or her estimates about the strategies of other players. Therefore, the neural mechanisms for belief learning or model-based reinforcement learning during iterative games might include the representations of not only the actual outcomes from chosen actions but also hypothetical outcomes from unchosen actions.

Prediction errors associated with counterfactual inferences, such as regret and relief, can exert an influence on emotion as well as decision making even in non-social contexts. Such counterfactual error signals have been localized in a distributed prefrontal network that includes the orbitofrontal cortex (Mellers et al. 1999; Camille et al. 2004; Coricelli et al. 2005; Chandrasekhar et al. 2008; Fujiwara et al. 2009). For example, neurons in the primate anterior cingulate cortex (ACC) represent both experienced and fictive payoffs similarly, independently of actions associated with those outcomes (Hayden et al. 2009). In another study (Abe and Lee 2011), monkeys were trained to play a rock-paper-scissors game against a computerized opponent. Similar to human subjects, choice probability of the animal was influenced by both actual and fictive rewards with actual outcomes having greater influence than five rewards. Similar to the neural activity in the ACC, neurons in the dorsolateral prefrontal cortex (DLPFC) and orbitofrontal cortex (OFC) often modulated their activity according to both experienced and hypothetical outcomes (Figure 4). However, neurons in the DLPFC and OFC tended to encode the conjunction of specific actions and their hypothetical outcomes, namely, information about which action could have produced a particular hypothetical outcome. Furthermore, single neurons tended to represent the actual and hypothetical outcomes from a given action similarly, providing a possible neural substrate by which value functions of both chosen

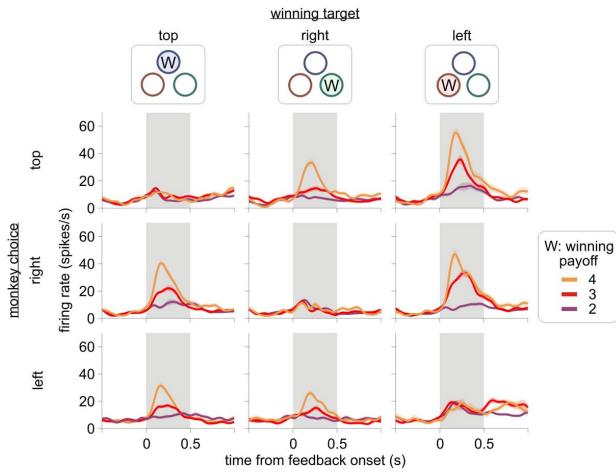


Figure 4. An example OFC neuron encoding the hypothetical outcome from the winning target during a rock-paper-scissors task. Each panel shows average spike density function estimated separately according to the position of the winning target (columns), the position of the target chosen by the animal (rows), and the winning payoff (colors). Thus, the results shown in the main diagonal are from the winning trials (Abe and Lee 2011).

and unchosen actions can be updated simultaneously. Similarly, several recent neuroimaging studies in human subjects have found that prediction errors based on both model-free and model-based reinforcement learning were represented in the ventromedial cortex as well as in the striatum (Lohrenz et al. 2007; Glasher et al. 2010; Daw et al. 2011; Simon and Daw 2011). Together, these findings suggest that common neural systems may underlie both model-free and model-based reinforcement learning.

NEURAL MECHANISMS FOR NON-COMPETITIVE GAMES AND ALTRUISM

Although the natural world is replete with examples of competitive interactions, humans and other animals also engage in numerous cooperative interactions, such as collaboration to maximize foraging outcomes (Fehr and Fischbacher 2003). In the parlance of game

theory, cooperative interactions often arise in non-zero sum games, because a net benefit can result for all participants. Cooperation is fascinating to behavioral scientists and neuroscientists because it forms the cornerstone of human society and requires complicated cognitive mechanisms to ensure that cooperation remains adaptive. Many such relationships depend upon concepts such as trust and reputation, which enable an organism to sacrifice a benefit in the short-term in order to reap future rewards. Intriguingly, social interactions characterized by game theory can reveal seemingly maladaptive or irrational behavior. For instance, altruism occurs when a participant purposefully sacrifices all or some of their own positive outcome to benefit another. Altruistic punishment occurs when an organism incurs a cost to itself in order to penalize another participant for bad behavior. Amorphous concepts such as fairness based on social norms also support cooperative behavior - violators can be detected and punished. This section reviews the use of games to investigate the mechanisms that enable a cooperative society, in particular those that are purported to reveal aspects of altruism, fairness, trust, and reputation. The three most prominent game theoretic tools used to investigate these concepts are the dictator, ultimatum, and prisoner's dilemma games.

ALTRUISM, GENEROSITY, AND THE DICTATOR GAME

Altruistic behavior is common amongst humans, and poses something of a puzzle in terms of evolution. For instance, why might a person donate a portion of their income to charitable causes, when no direct return on that investment is expected? Such behavior does not improve the fitness of the organism. Though it might seem a strong candidate for a purely human trait, altruism has been extensively studied by ethologists, who have purported to observe non-human animals behaving similarly. An example of such behavior is alarm calls amongst ground squirrels, which draw predators' attention to the caller (and increases their peril) but benefits the other ground squirrels (Sherman 1977). Another oft-cited example is sharing of food resources amongst vampire bats (Wilkinson 1984).

Using game theoretic principles and models of evolution, ethologists have explained altruistic behavior by means of two mechanisms. The first, kin selection, explains that altruistic behavior improves the chances of success for genetically related organisms (Hamilton 1964), and thus might have evolved to support the survival of genes responsible for cooperation. This might explain

the alarm calls of ground squirrels, which increase with the number of closely related animals nearby (Sherman 1977). The second mechanism is reciprocal altruism (Trivers 1971), which explains altruistic behavior on the basis of an expectation that such favors might be returned in the future. This mechanism, closely tied to notions of reputation and trust discussed below, might explain the example of food sharing amongst bats, which does not vary predictably based on kinship (Wilkinson 1984). Even if all elements of altruism can be explained in terms of kin selection and reciprocal altruism (or as epiphenomenal consequences of the evolution of mechanisms supporting those activities), altruistic behavior is interesting because of the complex cognitive traits that must have evolved to support its execution, such as kin recognition, cheater detection, and reputation.

Studies of *pure* altruism - that is, altruism that cannot be explained in terms of either kin selection or reciprocation - are frequently attempted in humans by means of the dictator game (DG). The dictator game is very simple: one agent is granted an endowment, and he or she makes a decision about the amount to donate to a second agent. Such a game played only once has one optimal strategy as prescribed by game theory: the dictator should keep the maximum amount possible, and grant the minimum possible amount to the other participant. Any deviation from this strategy could be interpreted as unselfish, altruistic behavior. In the lab and the field, human participants playing as the dictator do not conform to the selfish predictions of a basic game theoretic analysis, by making transfers to the recipient in more than 50% of cases, and often making substantial donations of roughly 20% (Forsythe et al. 1994; reviewed in Camerer 2003). Donations in the dictator game may be related to concepts of *fairness* and justice. For instance, Fehr and Fischbacher (2004) found that outside observers perceive even splits in the dictator game to be the most *fair*.

The dictator game has been used as a tool to understand the neural underpinnings of altruism in fMRI studies, neuropsychological studies of patients, and neuropeptide studies. In a naturalistic modification of the basic dictator game task, Moll and colleagues (2006) endowed participants in an fMRI study with \$128 and then posed a series of yes-or-no questions, some of which could affect the amount of the endowment taken away at the end of the experiment. Some of these decisions were costly donations, in which the participant could decide whether or not to give part of their endow-

ment to a charitable organization selected by the experimenters. Participants endowed a minimum of \$21 and on average sacrificed 40% of their endowment over the course of the experiment. Costly decisions to donate money were associated with higher activity in the ventral tegmental area and striatum, structures that are strongly related to reward processing. In the same experiment, trials that resulted in a reward to the participant also produced activation in these regions. This result was cited as support for the *warm-glow* effect, which explains charitable donations based on the intrinsic reward experienced by giving (Andreoni 1990). Charitable giving also activated another region, the subgenual area, which was not associated with direct rewards. The authors concluded that charitable donations generate a rewarding experience by activating the same regions that are activated by direct rewards. Since the subgenual region has previously been associated with feelings of social attachment, the authors suggest that empathy with causes elicits similar processes as social affiliation. However, subgenual activation is difficult to interpret in light of limited understanding of this region's function and its proximity to ACC, which plays a complex role in many cognitive tasks. Nevertheless, it is a site that is intriguingly associated with a high density of oxytocin (OT) receptors (Barberis and Tribolet 1996; Barraza et al. 2011). OT may play a role in dictator game behavior, but evidence is mixed. While intranasal OT administration reliably influences perceptions of trust and behavior in games such as the ultimatum game, discussed below, it was not found to influence dictator game responses (Zak et al. 2007). On the other hand, OT administration does increase quantity, but not rates, of charitable giving under similar task structure (Barraza et al. 2011), and genetic variation in oxytocin receptors has been associated with dictator game behavior (Israel et al. 2009).

Krajbich et al. (2009) compared dictator game offers in patients with damage in the ventromedial prefrontal cortex (VMPFC) to normal participants and brain-damaged controls. VMPFC patients made offers that were, on average, much lower than offers from the other groups. This, together with the behavior of these patients in ultimatum and trust games, led the authors to propose a role of the VMPFC in the sense of guilt. Regardless of the interpretation, these findings suggest an important role for the VMPFC in determining DG offers.

It is worth pointing out two of the many complications in interpreting dictator game behavior in terms of *pure* altruism. First, factors external to the game may influence behavior in a manner that muddles the interpretation. No experiment occurs in a vacuum, and participants may not want to be seen by the experimenters as greedy, for instance. Second, participants in experiments may have complex expectations about the consequences for their reward. Will a second, surprise phase of the experiment involve reciprocation? Will generous or selfish acts truly have no consequences? Intriguingly, the donations of participants in the dictator game are influenced by a number of factors, including the range of possible actions. When participants can decide to either donate money or take money from the other agent, average donations are reduced and some participants take money from opponents when such an option is available (Bardsley 2008). Other studies have demonstrated that feelings of *ownership* over the endowment (by requiring participants to earn it prior to the dictator game) lead to reduced donations (Cherry et al. 2002). Therefore, dictator game behavior may be linked not only to altruistic behavior, but also to notions of fairness and the expectations of experimenters.

DICTATOR GAME, RETRIBUTION, AND PUNISHMENT

Beyond altruism, when combined with other experimental elements, the dictator game is a useful tool to investigate social concepts such as fairness and affiliation. For instance, Fehr and Fischbacher (2004) enabled a third-party participant to *punish* the dictator for low offers. Third-parties indeed punished unfair offers, more often when they were more unfair, even though punishment was costly. To our knowledge, this paradigm has not been used in neuroscientific studies of fairness perception and punishment, but it could prove to be a useful tool.

Moor et al. (2012) showed that the dictator game can measure retribution based on behavior in a completely different task. Children of various age groups played a computer game called Cyberball, in which a ball is passed amongst several participants. Computer players were programmed to either include (by tossing the ball to the participant frequently) or exclude (by avoiding tossing the ball to the participant). Following this game, the participant made dictator game decisions involving each of these players. Participants made lower offers to players who excluded them than to those who included them. In addition, punishment elicited activ-

ity in tempoparietal junction (TPJ), superior temporal sulcus (STS), and ventrolateral PFC. TPJ and STS are frequently engaged during reasoning about the mental states of others (Gallagher and Frith 2003). This study suggests that the dictator game may prove useful as a measure of resentment of unfair social behavior in totally distinct situations.

ULTIMATUM GAME AND FAIRNESS

Another game that informs our understanding of fairness is the ultimatum game. The ultimatum game resembles the dictator game, but with a twist. Players in this game are the Proposer and the Responder. The Proposer is granted an endowment and chooses how much to share with the Responder, as in the dictator game. The Responder then chooses to either accept or reject the offer. If the offer is accepted, then the participants take home their shares. If the Responder rejects, then neither the Proposer nor the Responder get a reward. Basic game theory predicts that the Proposer should select the minimum non-zero offer, and the Responder should accept any non-zero offer, but of course, humans do not behave this way. Many investigators since Güth et al. (1982) found that Proposers tend to make high offers, many 50% of their totals, and that low offers tend to be rejected by Responders at very high rates (Camerer 2003). This has been interpreted as reflecting notions of fairness: low offers are perceived as unfair, and responders are willing to incur a loss in order to punish unfair Proposers.

The ultimatum game has been used extensively in neuroscientific studies of fairness since Sanfey et al. (2003), which examined fMRI signals associated with fair and unfair offers against human and computer opponents. They found that unfair offers elicited greater activity in a broad range of regions, including the DLPFC, anterior insula, and ACC, and that unfair offers from humans led to greater activity in the insula than unfair offers from computers. Insula activity also varied parametrically depending on the degree of unfairness or asymmetry in offers. The authors used insula activation, in particular, to support the role of emotion in responses to unfair offers, while DLPFC activity was related to cognitive control processes. While the insula and DLPFC clearly play important roles in ultimatum game responses, as supported by subsequent studies, such inferences seem problematic in light of these regions' complex and multifaceted roles in a diverse set of cognitive processes. For instance, insula is responsive to a variety of different manipulations,

and may also play a general role in attention (e.g., Dosenbach et al. 2006).

DLPFC involvement in reception and responses to unfair offers is also supported by repetitive transcranial magnetic stimulation (rTMS) and electroencephalography (EEG) studies. rTMS to the right DLPFC reduces rejection rates of unfair offers (Knoch et al. 2006) and causes longer reaction times (RT) in rejection responses to unfair ultimatum game proposals, but does not affect RT to fair offers (van 't Wout et al. 2005). An EEG study conducted on participants playing the ultimatum game also found that baseline cortical activity in the right prefrontal cortex predicts punishment behavior (Knoch et al. 2010). Tabibnia et al. (2008) employed an ultimatum game paradigm while scanning participants with fMRI, and focused on activity in response to fair offers. They showed that fair offers in the ultimatum game promoted activity in reward-related regions (ventral striatum, amygdala, VMPFC and OFC), and that this activity was dissociable from the monetary gain differences between fair and unfair offers. A region of VMPFC was more active in response to accepted than unaccepted unfair offers, and this was interpreted as support for emotional regulation in response to unfair offers. The ultimatum game has also been used to examine the role of neurotransmitters (particularly serotonin and OT) in responses to fairness. For instance, when serotonin was temporarily lowered by means of acute tryptophan depletion in responders, a significant increase in rates of rejection was observed, suggesting that serotonin modulates perception of fairness (Crocket et al. 2008). This might be related to the finding that ventral PFC damage has a similar effect on rejection rates (Koenigs and Tranel 2007).

In sum, these studies suggest that DLPFC and insula play important roles in perceiving and responding to unfairness or inequity, and that perception of fairness modulates reward-related circuitry. Serotonin and OT are clearly important in systems that modulate social behavior and perceptions, and the involvement of DLPFC and insula may be related to these neurochemicals. Further research should seek to dissociate more cleanly the role of these systems in general cognitive processes from their involvement in perceptions and reactions to fairness.

THE TRUST GAME AND THE PRISONERS' DILEMMA: TRUST, RECIPROCITY, AND REPUTATION

Much cooperative behavior pays off only when partners eventually reciprocate, and many cases of altruism and generosity in animals and humans can be explained by reciprocal relationships (Trivers 1971). Relying on reciprocation to forego smaller but more certain rewards depends on trust, which can be reinforced based on one's own and others' experiences that are reflected in their reputations. Two games that are important tools to study these constructs are the trust game and the prisoner's dilemma game.

In the trust game (Berg et al. 1995), one player (the investor) is given an endowment. The investor decides how much endowment to give to the trustee player. The entrusted amount is then multiplied by a predetermined factor (>1), and the trustee decides how much to return to the investor. Although basic game theory predicts a lack of trust and no investment in this game (since reciprocation is not enforced), and also predicts no return of the investment by the trustee, human players tend to exhibit high levels of trust and cooperation. They tend to invest about half of their endowment, and usually receive roughly the same amount back from the trustee (reviewed by Camerer 2003).

In the first fMRI study of participants playing the trust game, McCabe and colleagues scanned participants while they played trust games against either human or computer opponents in an attempt to examine neural correlates of theory-of-mind. They reported that a group with frequent cooperation showed higher activity in the prefrontal cortex in response to decisions against human than computers, whereas a group of noncooperators did not show such a difference. They interpreted this as evidence that the prefrontal cortex reflects theory-of-mind processes. Several fMRI studies have examined trust games while participants in the game repeatedly interacted while simultaneously scanned, in order to study the development of neural signals related to trust. For instance, King-Casas et al. (2005) scanned participants with fMRI as they played iterated trust games. They found that dorsal striatum responses following trial outcomes predicted trust plays on subsequent trials, and that this response peaked earlier as interaction with a given opponent increased. The authors suggest that this shift reflects the development of a model of the partner's intentions. In a similar study, Krueger et

al. (2007) scanned participants during repeated trust games. They focused on the paracingulate cortex (PcC) and the septal area (SA), two regions that showed higher activity during trust decisions than a control choice. They divided the experiment into first and second stages (*building* and *maintenance* stages), and split the participant pairs into two groups, *defectors* and *non-defectors*, based on whether either participant in the pair defected. In the building phase, they found higher activity in the PcC of the non-defector group than of the defector group during the first move of the game, a finding that the authors related to the role of PcC in mentalizing or theory-of-mind. During the maintenance phase, the authors found higher activation in the SA for non-defectors than defectors, a finding the authors suggested may be linked to this area's role in trust-related neurochemicals (OT and vasopressin). Indeed, OT seems to play a role in behavior during the trust game. OT administration increases the amount of investment during the trust game, but has no effect on risk preference otherwise (Kosfeld et al. 2005).

In an experiment linking trust and reward learning signals (Delgado et al. 2005), participants played a trust game against three fictitious partners, after receiving extensive background information that induced *good*, *bad* or *neutral* perceptions about these partners. The participants then played repeated trust games against all three partners while scanned using fMRI. Intriguingly, although all three partners behaved identically, reputation modulated neural responses. When the authors examined the activity in the ventral caudate nucleus (localized by the contrast between positive vs. negative feedback), they found that this region did not differentiate positive and negative outcomes as strongly for *good* or *bad* partners as for *neutral* partners, suggesting that reputation can sometimes override learning signals.

The prisoner's dilemma game has a similar form compared to the trust game, but decisions are made simultaneously. This game is most commonly described by analogy to two criminals being interrogated separately. Players have two choices - *Defect* or *Cooperate*. If both players cooperate (*remain silent*), they receive a small punishment (e.g., a month in jail). If one player defects and the other cooperates, the defector received the minimum punishment (no jail time), while the cooperator receives the maximum jail time (e.g., one year). If both players defect, they receive an intermediate punishment (e.g., three months in jail). In both one-shot and iterated versions of this game, classical game theory predicts

defection (i.e., Nash equilibrium), although participants do better by cooperating over time (Axelrod and Hamilton 1981), and human participants indeed frequently cooperate (Wedekind and Milinski 1996). To investigate the neural basis of cooperation, Rilling et al. (2002) scanned participants as they played iterated prisoner's dilemma games against other humans or computers. Reciprocated cooperation with humans elicited activation in reward-related regions such as the caudate nucleus, nucleus accumbens, VMPFC, OFC, and rostral ACC. Only the OFC was activated during cooperative interactions with computers. Therefore, these activations were separate from responses to monetary rewards. A similar study utilizing a one-shot prisoner's dilemma game showed similar results in the VMPFC and ventral striatum (Rilling et al. 2004). These results suggest that cooperation may be intrinsically rewarding in some cases.

CONCLUSIONS

During the last decade, the popularity of game theoretic frameworks in investigating the neural mechanisms underlying complex social decision making increased substantially. As a result, dynamical interactions among decision makers as well as norms of social exchange became increasingly subjected to scientific and quantitative research. Despite these rapid advances in understanding the formal and behavioral aspects of social decision making, our knowledge of the neural processes involved in social interactions is still limited and a number of important questions remain to be answered.

Fairness and altruistic behaviors might activate common reward systems. However, the neural mechanisms of social comparison that underlie these more abstract concepts need to be understood better. Furthermore, the process of social comparison is clearly modulated by contexts, for example, whether such comparisons are made in a cooperative or competitive environment. Parametric changes in the payoff matrix can be a useful tool to induce more competitive or cooperative behaviors, and thus to investigate the neural mechanism responsible for flexibly modulating basic social comparison.

Interestingly, many studies have found that brain regions involved in diverse cognitive functions, such as attention and executive functions, are also involved in social decision making. Therefore, how social behaviors are constrained by, or interact with, other cognitive functions, and whether individual differences in

cognitive functions are correlated with their behavioral orientations in social interactions, will be a fruitful area of research. Finally, understanding the precise nature of genetic and neuromodulatory factors that can explain individual variances in social interactions will be important for the effective treatment of abnormal social behaviors (Kishida et al. 2010).

REFERENCES

- Abe H, Lee D (2011) Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70:731-741
- Andreoni J (1990) Impure altruism and donations to public goods: a theory of warm-glow giving. *Econ J* 100:464-477
- Axelrod R, Hamilton RD (1981) The evolution of cooperation. *Science* 211:1390-1396
- Barberis C, Tribollet E (1996) Vasopressin and oxytocin receptors in the central nervous system. *Crit Rev Neurobiol* 10:119-154
- Bardsley, N (2008) Dictator game giving: altruism or artefact? *Exp Econ* 11:122-133
- Barraclough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci* 7:404-410
- Barraza JA, McCullough ME, Ahmadi S, Zak PJ (2011) Oxytocin infusion increases charitable donations regardless of monetary resources. *Horm Behav* 60:148-151
- Behrens TE, Hunt LT, Rushworth MF (2009) The computation of social behavior. *Science* 324:1160-1164
- Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. *Nature* 456:245-249
- Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. *Nat Neurosci* 10:1214-1221
- Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity, and social history. *Games Econ Behav* 10:122-142

- Brown JN, Rosenthal RW (1990) Testing the minimax hypothesis: a re-examination of O'Neil's game experiment. *Econometrica* 58:1065-1081
- Cai X, Kim S, Lee D (2011) Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during inter-temporal choice. *Neuron* 69:170-182
- Camerer CF (2003) Behavioral game theory: experiments in strategic interaction. Princeton University Press, Princeton, NJ
- Camerer CF, Ho TH (1999) Experience-weighted attraction learning in normal form games. *Econometrica* 67:827-874
- Camille N, Coricelli G, Sallet J, Pradat-Diehl P, Duhamel J-R, Sirigu A (2004) The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304:1167-1170
- Chandrasekhar PVS, Capra CM, Moore S, Noussair C, Berns GS (2008) Neurobiological regret and rejoice functions for aversive outcomes. *Neuroimage* 39:1472-1484
- Cherry TL, Frykblom P, Shogren JF (2002) Hardnose the dictator. *Am Econ Rev* 92:1218-1221
- Cheung YW, Friedman D (1997) Individual learning in normal form games: some laboratory results. *Games Econ Behav* 19:46-76
- Colman AM (1999) Game theory and its applications in the social and biological sciences. Routledge, New York, NY
- Coricelli G, Critchley HD, Joffily M, O'Doherty JP, Sirigu A, Dolan RJ (2005) Regret and its avoidance: a neuroimaging study of choice behavior. *Nat Neurosci* 8:1255-1262
- Crockett MJ, Clark L, Tabibnia G, Lieberman MD, Robbins TW (2008) Serotonin modulates behavioral reactions to unfairness. *Science* 320:1739
- Cui H, Andersen RA (2007) Posterior parietal cortex encodes autonomously selected motor plans. *Neuron* 56:552-559
- Daw ND, Doya K (2006) The computational neurobiology of learning and reward. *Curr Opin Neurobiol* 16:199-204
- Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204-1215

- Delgado MR, Frank RH, Phelps EA (2005) Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci* 8:1611-1618
- Dorris MC, Glimcher PW (2004) Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44:365-378
- Dosenbach NU, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlagger BL, Petersen SE (2006) A core system for the implementation of task sets. *Neuron* 50:799-812
- Erev I, Roth AE (1998) Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am Econ Rev* 88:848-881
- Fehr E, Fischbacher U (2003) The nature of human altruism. *Nature* 425:785-791
- Fehr E, Fischbacher U (2004) Third-party punishment and social norms. *Evol Hum Behav* 25:63-87
- Feltovich N (2000) Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games. *Econometrica* 68:605-641
- Forsythe R., Horowitz JL, Savin NE, Sefton M (1994) Fairness in simple bargaining experiments. *Games Econ Behav* 6:347-369
- Fudenberg D, Levine DK (1998) The theory of learning in games. MIT Press, Cambridge, MA
- Fujiwara J, Tobler PN, Taira M, Iijima T, Tsutsui K-I (2009) A parameteric relief signal in human ventrolateral prefrontal cortex. *Neuroimage* 44:1163-1170
- Gallagher HL, Frith CD (2003) Functional imaging of *theory of mind*. *Trends Cog Sci* 7:77-83
- Gerfen CR, Surmeier DJ (2011) Modulation of striatal projection systems by dopamine. *Annu Rev Neurosci* 34:441-466
- Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66:585-595
- Güth W, Schmittberger R, Schwarze B (1982) An experimental analysis of ultimatum bargaining. *J Econ Behav Organ* 3:367-388

- Hamilton WD (1964) The genetic evolution of social behavior. *J Theoret Biol* 7:1-52
- Hayden BY, Pearson JM, Platt ML (2009) Fictive reward signals in the anterior cingulate cortex. *Science* 324:948-950
- Herrstein RJ (1997) The matching law: papers in psychology and economics. Harvard University Press, Cambridge, MA
- Houk JC, Davis JL, Beiser DG (1995) Models of information processing in the basal ganglia. MIT press, Cambridge, MA
- Israel S, Lerer E, Shalev I, Uzefovsky F, Riebold M, Laiba E, Bachner-Melman R, Maril A, Bornstein G, Knafo A, Ebstein RP (2009) The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the dictator game and the social value orientations task. *PLoS One* 4:e5535
- King-Casas B, Tomlin D, Anen C, Camerer CF, Quartz SR, Montague PR (2005) Getting to know you: Reputation and trust in a two-person economic exchange. *Science* 308:78-83
- Kishida KT, King-Casas B, Montague PR (2010) Neuroeconomic approaches to mental disorders. *Neuron* 67:543-554
- Knoch D, Pascual-Leone A, Meyer K, Treyer V, Fehr E (2006) Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314:829-832
- Knoch D, Gianotti LRR, Baumgartner T, Fehr E (2010) A neural marker of costly punishment behavior. *Psychol Sci* 21:337-342
- Koenigs M, Tranel D (2007) Irrational economic decision-making after ventromedial prefrontal damage: evidence from the ultimatum game. *J Neurosci* 27:951-956
- Kosfeld M, Heinrichs M, Zak PJ, Fischbacher U, Fehr E (2005) Oxytocin increases trust in humans. *Nature* 435:673-676
- Krajbich I, Adolphs R, Tranel D, Denburg NL, Camerer CF (2009) Economic games quantify diminished sense of guilt in patients with damage to the prefrontal cortex. *J Neurosci* 29:2188-2192
- Krueger F, McCabe K, Moll J, Kriegeskorte N, Zahn R, Strenziok M, Heinecke A, Grafman J (2007) Neural correlates of trust. *Proc Natl Acad Sci USA* 104:20084-20089
- Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* 58:451-63

- Lee D (2008) Game theory and neural basis of social decision making. *Nat Neurosci* 11:404-409
- Lee D, Conroy ML, McGreevy BP, Barracough DJ (2004) Reinforcement learning and decision making in monkeys during a competitive game. *Cogn Brain Res* 22:45-58
- Lee D, McGreevy BP, Barracough DJ (2005) Learning and decision making in monkeys during a rock-paper-scissors game. *Cogn Brain Res* 25:416-430
- Lee D, Seo H (2007) Mechanisms of reinforcement learning and decision making in the primate dorsolateral prefrontal cortex. *Ann NY Acad Sci* 1104:108-122
- Lohrenz T, McCabe K, Camerer CF, Montague PR (2007) Neural signature of fictive learning signals in a sequential investment task. *Proc Natl Acad Sci USA* 104:9493-9498
- Malcolm D, Lieberman B (1965) The behavior of responsive individuals playing a two-person, zero-sum game requiring the use of mixed strategies. *Psychon Sci* 2:373-374
- Mellers B, Schwartz A, Ritov I (1999) Emotion-based choice. *J Exp Psychol Gen* 128:332-345
- Moll J, Krueger F., Zahn R, Pardini M, de Oliveira-Souza R, Grafman J (2006) Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci USA* 103:15623-15628
- Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16:1936-1947
- Mookerjee D, Sopher B (1994) Learning behavior in an experimental matching pennies game. *Games Econ Behav* 7:62-91
- Mookherjee D, Sopher B (1997) Learning and decision costs in experimental constant games. *Games Econ Behav* 19:97-132
- Moor BG, Güroğlu B, Op de Macks ZA, Rombouts SARB, Van der Molen MW, Crone EA (2012) Social exclusion and punishment of excluders: Neural correlates and developmental trajectories. *NeuroImage* 59:708-717
- Nash JF (1950) Equilibrium points in n-person games. *Proc Natl Acad Sci USA* 36:48-49

- Ochs J (1995) Games with unique, mixed strategy equilibria: An experimental study. *Games Econ Behav* 10:202-217
- O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr Opin Neurobiol* 14:769-776
- O'Neil B (1987) Nonmetric test of the minimax theory of two-person zerosum games. *Proc Natl Acad Sci USA* 84:2106-2109
- Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441:223-226
- Pastor-Bernier A, Cisek P (2011) Neural correlates of biased competition in premotor cortex. *J Neurosci* 31:7083-7088
- Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400:233-238
- Rapoport A, Boebel RB (1992) Mixed strategies in strictly competitive games: A further test of the minimax hypothesis. *Games Econ Behav* 4:261-283
- Reynolds JNJ, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. *Nature* 413:67-70
- Rilling JK, Gutman DA, Zeh TR, Pagnoni G, Berns GS, Kilts CD (2002) A neural basis for social cooperation. *Neuron* 35:395-405
- Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004) Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *NeuroReport* 15:2539-2543
- Rilling JK, Sanfey AG (2011) The neuroscience of social decision-making. *Ann Rev Psychol* 62:23-48
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337-1340
- Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003) The neural basis of economic decision-making in the ultimatum game. *Science* 300:1755-1758
- Schultz W (2006) Behavioral theories and the neurophysiology of reward. *Ann Rev Psychol* 57:87-115
- Schultz W, Tremblay L, Hollerman JR (2000) Reward processing in primate orbitofrontal cortex and basal ganglia. *Cereb Cortex* 10:272-283

- Seo H, 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, Lee D (2008) Cortical mechanisms for reinforcement learning in competitive games. *Philos Trans R Soc Lond B* 363:3845-3857
- Seo H, Lee D (2009) Behavioral and neural changes after gains and losses of conditioned reinforcers. *J Neurosci* 29:3627-3641
- Seo H, Barraclough DJ, Lee D (2009) Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. *J Neurosci* 29:7278-7289
- Shen W, Flajolet M, Greengard P, Surmeier DJ (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321:848-851
- Sherman PW (1977) Nepotism and the evolution of alarm calls. *Science* 197:1246-1253
- Simon DA, Daw ND (2011) Neural correlates of forward planning in a spatial decision task in humans. *J Neurosci* 31:5526-5539
- So NY, Stuphorn V (2010) Supplementary eye field encodes option and action value for saccades with variable reward. *J Neurophysiol* 104:2634-2653
- Soltani A, Lee D, Wang X-J (2006) Neural mechanism for stochastic behaviour during a competitive game. *Neural Netw* 19:1075-1090
- Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782-1787
- Sugrue LP, Corrado GS, Newsome WT (2005) Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat Rev Neurosci* 6:363-375
- Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. MIT Press, Cambridge, MA
- Tabibnia G, Satpute AB, Lieberman MD (2008) The sunny side of fairness - preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol Sci* 19:339-347

- Thevarajah D, Mkiulic A, Dorrice MC (2009) Role of the superior colliculus in choosing mixed-strategy saccades. *J Neurosci* 29:1998-2008
- Trivers RL (1971) The evolution of reciprocal altruism. *Quart Rev Biol* 46:35-57
- van 't Wout M, Kahn RS, Sanfey AG, Aleman A (2005) Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport* 16:1849-1852
- Vickery TJ and Jiang YV (2009) Inferior parietal lobule supports decision making under uncertainty in humans. *Cereb Cortex* 19:916-925
- von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton University Press, Princeton, NJ
- Wallis JD, Kennerley SW (2010) Heterogeneous reward signals in prefrontal cortex. *Curr Opin Neurobiol* 20:191-198
- Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF (2010) Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* 65:927-939
- Wedekind C, Milinski M (1996) Human cooperation in the simultaneous and the alternating Prisoner's Dilemma: Pavlov versus Generous Tit-for-Tat. *Proc Natl Acad Sci USA* 93:2686-2689
- Wilkinson GS (1984) Reciprocal food sharing in the vampire bat. *Nature* 308:181-184
- Wolpert DM, Doya K, Kawato M (2003) A unifying computational framework for motor control and social interaction. *Philos Trans R Soc Lond B* 358:593-602
- Zak PJ, Stanton AA, Ahmadi S (2007) Oxytocin increases generosity in humans. *PLoS One* 2:e1128