



HHS Public Access

Author manuscript

Trends Neurosci. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

Trends Neurosci. 2016 January ; 39(1): 40–48. doi:10.1016/j.tins.2015.11.002.

Neural Basis of Strategic Decision Making

Daeyeol Lee^{1,2,3} and Hyojung Seo¹

¹Department of Neuroscience, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA

²Kavli Institute for Neuroscience, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510, USA

³Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520, USA

Abstract

Human choice behaviors during social interactions often deviate from the predictions of game theory. This might arise partly from the limitations in cognitive abilities necessary for recursive reasoning about the behaviors of others. In addition, during iterative social interactions, choices might change dynamically, as knowledge about the intentions of others and estimates for choice outcomes are incrementally updated via reinforcement learning. Some of the brain circuits utilized during social decision making might be general-purpose and contribute to isomorphic individual and social decision making. By contrast, regions in the medial prefrontal cortex and temporal parietal junction might be recruited for cognitive processes unique to social decision making.

Keywords

game theory; reinforcement learning; arbitration; prefrontal cortex

Social decision making

In theories of decision making, utilities and values play a central role. The concept of utility allows a concise summary of the decision maker's preferences among numerous options, since choices are assumed to maximize utility [1]. In some cases, this assumption is relaxed and replaced by a monotonic relationship between the probability of choosing a particular option and its value [2]. This implies that all the factors affecting the choice of a decision maker must be incorporated into the utility or value functions, although in practice, specific theories and models consider only a subset of such factor. For example, the expected utility theory [1] and prospect theory [3] describe how uncertain outcomes of decision making can be evaluated, whereas the theories of temporal discounting specify how the value of a decision outcome might be influenced by its delay [4,5].

Corresponding author: Lee, D. (daeyeol.lee@yale.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

In contrast to individual decision making, interactions in social settings introduce two additional elements [6–9]. First, choices might be influenced by the decision-maker's other-regarding preferences. For example, purely altruistic behaviors might result when decision makers acts to increase the well-being of others despite the negative consequences of such actions to themselves. In other cases, individual decision makers might choose to punish those selfishly violating social norms [10, 11]. Second, in social settings, the outcome and hence its utility or value of an individual's choice depends on the choices of others in the group. For example, during a rock-paper-scissors game, each option beats one of the other two options but loses to the remaining one. Therefore, the ability to predict the choices of others and their underlying cognitive processes, often referred to as the theory of mind, plays an important role during social interactions [12, 13]. In this review, we will focus on the psychological processes of strategic reasoning during social interactions and their underlying neural mechanisms.

Decision making in a social setting is often referred to as a game, and analyzed formally by game theory [1]. For example, when everyone in a group forms accurate beliefs about the behaviors of everyone else and acts according to the principle of utility maximization, a Nash equilibrium (see Glossary) is obtained, namely, a set of strategies from which individual decision makers cannot deviate unilaterally without decreasing their utilities [14]. For example, during the rock-paper-scissors game, Nash equilibrium is for everyone to choose all three options randomly and equally often. Any other strategy, such as favoring any particular option, can be exploited by opponents. It should be noted that it is not necessarily optimal to adopt a strategy of Nash equilibrium when not everyone does the same. The optimal strategy expected for a set of strategies that is not necessarily at the Nash equilibrium is referred to as a best-response strategy. Therefore, at Nash equilibrium, everyone's strategy is the best response to everyone else's.

Although the concept of Nash equilibrium makes a precise prediction about the strategies in any social setting in which all decision makers have complete information about the nature of the game and act selfishly, these assumptions are often violated [15–17]. Moreover, choices observed in experimental games are often better accounted for by models of bounded rationality [18,19], such as the cognitive hierarchy model, in which individual decision makers select the best response based on their subjective and often inaccurate beliefs about the strategies of others [20, 21]. Such models have been successfully applied to a broad range of experimental games, including the beauty-contest game and stag-hunt game [12,17, 22]. These results suggest that strategic decisions during social interactions might be largely based on limited iterative reasoning about the likely behaviors of others.

Learning during social interactions

Most social interactions tend to occur repeatedly. Therefore, recursive inferences during social decision making might be adjusted and augmented by learning as decision makers encounter the same or similar games repeatedly. Indeed, the reinforcement learning theory [2], developed based on the Markov decision process [23], is not limited to the choices of decision makers acting individually in a non-social environment, but might also account for the dynamics of choice behaviors during repeated social interactions [15, 24]. It is assumed

in reinforcement learning theory that the outcome or reward is determined entirely, albeit probabilistically, by the current state of the decision maker's environment and their chosen action. Estimates for the outcomes are referred to as value functions, which can be defined for specific states (state value functions) or specific actions chosen from each state (action value functions).

Reinforcement learning algorithms can be divided into two categories, depending on how value functions are updated through experience [2, 24–28]. In simple or model-free reinforcement learning algorithms, the outcome of each chosen action is compared to the prediction from the current value functions. The resulting difference is referred to as reward prediction error and is used to update the value function for the chosen action or the most recently visited state. Model-free reinforcement learning algorithms are robust, since they rely on only a small number of parameters, but relatively inflexible, since the value functions are updated incrementally only for the chosen action and specific states visited by the decision maker. Therefore, if there is a sudden change in the environment, many trials would be required before all the value functions are appropriately adjusted. In contrast, model-based reinforcement learning algorithms compute the value functions directly from the decision maker's knowledge of their environment. Often, such knowledge is embodied in the transition probabilities between different states of the environment and reward probabilities from different states, which can be used to simulate the experience of the decision maker. The fictive or hypothetical outcomes generated during this simulation are then used to update the value functions. Although computationally intensive, such simulations can be performed extensively, allowing the value functions to be updated more quickly and rendering model-based reinforcement learning algorithms more flexible than model-free algorithms.

Both types of reinforcement learning can be applied to iterative decision making in a social context. For example, Nash equilibrium strategies in many games require the decision makers to make their choices stochastically, rather than choosing one option exclusively. This is referred to as a mixed strategy. Previous studies have shown that for games with mixed-strategy equilibria, human and non-human primates tend to use model-free reinforcement learning algorithms to approximate equilibrium strategies [15, 29, 30]. However, even for relatively simple games, model-free reinforcement learning algorithms might be supplemented by model-based reinforcement learning algorithm [31–36], analogous to the findings from studies with non-social multi-stage decision making tasks [37–39].

In game theory, players might update their beliefs about the choices of other players through experience and then choose their strategies expected to produce the best outcomes given such beliefs [15, 40,41]. This is known as belief learning, and is equivalent to model-based reinforcement learning. For example, in models of fictitious play, the belief about the strategy of a given player is determined by the frequency of each strategy previously chosen by the same player [42]. Algorithmically, belief learning or fictitious-play models can be implemented in two different ways. First, after each play, the estimates for the frequency of other players playing each strategy might be updated based on their observed behaviors. This is analogous to updating the transition probability according to the state prediction

errors during a non-social multistage decision making task [43]. Once reliable internal models of other plays are constructed, then, they can be used to compute the value functions of alternative actions during subsequent interaction with the same players through mental simulations, namely, recursive strategic reasoning [12,17, 22]. Second, it is also possible that the value functions for multiple alternative actions might be updated immediately after each trial according to the hypothetical outcome signals inferred from the observed behaviors of other players [34, 35]. For example, if the opponent chooses a rock during a rock-paper-scissors game, then the value functions for paper and scissors might be increased and decreased, respectively, according to their positive and negative hypothetical payoffs. In this case, the process of mental simulation would not be required during subsequent trials. It might be possible to distinguish between these two possibilities by carefully examining the amount of time needed for decision making, since mental simulation is likely to be a time-consuming process. In addition, understanding the nature and time course of different computations involved in belief learning would be important for the studies on the underlying neural mechanisms.

Neural mechanisms for learning in social domain

Since the computational process of comparing and updating the value functions for alternative actions are similar for both social and non-social decision making, the neural mechanisms for strategic social decision making might substantially overlap with those involved in individual decision making [9]. The neural mechanisms for model-free reinforcement learning might be distributed in a large number of cortical areas and the basal ganglia [24], and might be shared for both social and non-social decision making. For example, during model-free reinforcement learning, the signals related to the previous choices and their outcomes must be integrated properly to update the value functions, and these signals are present in multiple areas of the brain, including the prefrontal cortex and basal ganglia [24, 44–48]. Similarly, signals related to previous reinforcement and punishment can be decoded from the blood-oxygen-level-dependent (BOLD) signals from almost the entire brain of human participants during competitive games against a computer opponent [49]. These outcome-related BOLD signals are attenuated when the participants face a different opponent compared to when they face the same opponent as in the previous trial, suggesting that they are specifically linked to the opponent identity and do not merely reflect the passive decay in the BOLD signals [50].

For optimal performance, the rate of learning should be adjusted according to the uncertainty or volatility of the decision maker's environment. When the environment is highly unstable or uncertain, the learning rate should be set high so that the memory about the remote past can be replaced more rapidly by the new information about the present [51, 52]. Neuroimaging studies found that the activity in the anterior cingulate cortex was related to the volatility of the environment, suggesting that this area might play an important role in regulating the rate of learning according to the level of uncertainty in the decision maker's environment [51]. Neurophysiological studies have also found that signals encoded by the individual neurons in the anterior cingulate cortex and other regions of the primate prefrontal cortex encode the animal's previous reward history in multiple time scales, and thus might provide the flexible memory signals necessary to regulate the rate of

reinforcement learning adaptively [44, 53, 54]. There is also some evidence that the ACC sulcus and gyrus might be specialized for monitoring the outcomes of individual and social decision making, respectively [55–59].

Compared to the neural mechanisms for model-free reinforcement learning, how various types of model-based reinforcement learning are implemented in the brain is less well known. In addition, as representations and computations for model-based reinforcement learning must be tailored to the specific decision-making problems, functions localized in a given brain area for such learning might also vary. Nevertheless, recent studies have begun to provide important insights into the candidate neural correlates of model-based reinforcement learning, including those involved in social interactions. For example, changes in the environment that do not match the predictions based on the decision maker's knowledge produce the so-called state prediction errors, and this leads to increased activation in the intraparietal sulcus and lateral prefrontal cortex, suggesting that these brain areas might be involved in updating the probabilities of state transitions [39, 43]. When the predictions about the behaviors of others are violated, similar error signals can be generated. These so-called simulated-other's action prediction errors (SAPE) were also reflected in the BOLD signals in the dorsolateral prefrontal cortex (Figure 1A) [36]. Therefore, errors in the knowledge of the decision makers in both social and non-social domains might be processed by common or overlapping neural systems. However, the social environment that includes the intentions and preferences of others is typically more complex than the physical environment, and therefore, the acquisition and revision of its knowledge might involve additional areas of the brain. For example, signals related to SAPE were found not only in the lateral prefrontal cortex, but also in the dorsomedial prefrontal cortex (dmPFC; Figure 1A) [36] as well as the temporal parietal junction (TPJ) and the posterior superior temporal sulcus (pSTS). Similarly, when choices are made based on the advice from another person, the activity in the dmPFC reflected the prediction error for the accuracy of the advice [60]. In addition, predictions about the behaviors of others can be improved not only based on the simulated-other's action prediction errors, but also by simulating the reward prediction errors computed by others. These so-called simulated reward prediction error signals were found in the ventromedial prefrontal cortex [36].

The dorsomedial and ventromedial prefrontal cortex might also play an important role in learning and encoding the preference of others during social interactions. However, the precise anatomical areas encoding the signals related to the preference of self and others might vary according to the nature of choices and the personal traits of decision makers. For example, previous studies examined the nature of value signals encoded in these two brain areas while the decision makers chose for themselves and others alternately [60]. When the choices were for the decision makers themselves, the value difference between the chosen and unchosen options of the decision makers and their partners were reflected in the activity of the vmPFC and dmPFC, respectively. The locations of the value difference signals for the self and others were exchanged, when the choices were made for the partners, suggesting that the activity in the vmPFC and dmPFC reflected the value signals for executed and simulated choices, respectively, rather than the preferences of self vs. others [60]. Consistent with this hypothesis, the activity in the dmPFC increased when human participants made accurate predictions about the preference of others [61]. Furthermore, a

recent study showed that pro-sociality of the decision makers might further influence the extent to which the values of others influence the activity in the vmPFC (Figure 1B) [62].

The process of estimating the value functions from the information about the likely actions of others and their preference might be implemented in multiple brain areas. For example, neural recording in rodents as well as lesion and neuroimaging studies in human subjects have suggested that the of simulated outcomes might rely on the hippocampus [63–66]. In particular, the information decoded from the ensemble activity in the hippocampus might be related to the internal simulation of the animal's future actions [67–69]. Simulating the outcomes from alternative actions during social decision making might also rely on a set of cortical areas often associated with the theory of mind, such as the dmPFC and TPJ [6, 7, 13, 70, 71]. Although it has been suggested that the theory of mind is uniquely human, the middle superior temporal sulcus (STS) in the macaque brain displays a pattern of resting-state functional connectivity that is similar to that of the human TPJ [72]. Therefore, understanding the nature of information encoded by the neurons in the macaque middle STS might provide important clues about the contribution of TPJ in social decision making. In addition, the depth of recursive reasoning during social decision making was reflected in the activity of medial and lateral prefrontal cortex, suggesting that these areas might play an important role in predicting outcomes of social decision making according to the information about the likely behaviors of others [22, 73].

During repeated social interactions, decision makers might update not only their internal models about the behaviors of others, but also the value functions for their own actions based on counterfactual inferences. Lesion and neuroimaging studies have previously implicated the orbitofrontal cortex in counterfactual reasoning during individual decision making when the decision makers receive the information about the outcomes from unchosen actions [74, 75]. Signals related to the values of unchosen outcome have also been localized in the frontopolar cortex and dmPFC [76, 77]. The brain areas responsible for integrating counterfactual outcome or counterfactual reward prediction error signals into value functions during social interactions might at least partially overlap with those involved in counterfactual outcome processing for non-social decision making. For example, during repeated strategic interactions, value functions are updated not only according to the previous outcomes from the same action and counterfactual outcomes from unchosen action, but also based on the changes in the behaviors of others expected to occur in response to the participant's own actions. These so-called influence update signals were found in the superior temporal sulcus and medial prefrontal cortex [33]. In addition, signals related to belief prediction errors, namely, prediction errors related to belief learning or model-based reinforcement learning during iterative social decision making were found in the rostral anterior cingulate cortex [35]. On the other hand, activity in the striatum reflects both reward prediction error and belief prediction error during social decision making [35], consistent with the findings from studies on non-social multi-stage and financial decision making [37, 78]. Single neuron recording studies have also found that the signals related to actual outcomes from chosen actions and hypothetical outcomes from unchosen actions are often encoded by the same neurons in the prefrontal cortex [34] and anterior cingulate cortex [79].

Neural mechanisms for arbitration between learning algorithms

Many computations performed by the brain involve combining multiple sources of information, as when trying to estimate the location of an object based on multiple sensory cues [80]. For optimal performance, it is necessary to adjust the weights for different types of information according to their uncertainty [81]. Similarly, when multiple learning algorithms are available for a particular problem of decision making, the value functions computed by different algorithms must be combined with the weights determined by their accuracy [25]. During a multi-stage decision making, signals related to the accuracy of model-free reinforcement learning algorithm were found in the striatum, whereas those related to the accuracy of model-based algorithm were found in multiple cortical areas, including the dmPFC [39]. In addition, multiple regions of the prefrontal cortex might be involved in arbitrating between the two learning algorithms [39, 82]. Given that model-free reinforcement learning algorithms are more efficient computationally and require a fewer free parameters than model-based algorithms, the brain might rely on model-free algorithms as a default, especially when the decision makers lack accurate knowledge about their environment, and switch away from model-free algorithms when appropriate [83]. In fact, the correlation in the activity of the prefrontal cortex and striatum was inversely related to the reliability of model-based algorithm, and might be related to this switching process [39].

Neurophysiological studies in non-human primates have also identified switching signals that might be involved in the suppression of habitual automatic behaviors. For example, activity of the neurons in the medial prefrontal cortex often reflected a switching between individual motor responses or between different sequences of actions [84, 85]. This brain area might be also involved in switching away from a model-free reinforcement learning algorithm that can be exploited by a more sophisticated opponent during competitive social interactions [86]. In particular, during a computer-simulated biased matching pennies game, activity of neurons in the dmPFC displayed switching signals whose strength was correlated with the animal's tendency to switch away from the use of model-free reinforcement learning algorithm (Figure 2) [86].

Conclusions

The research on the neural basis of strategic decision making during social interactions poses a number of challenges due to its complexity and diversity. Therefore, computational tools developed in economics and machine learning play an important role in precisely characterizing the nature of various algorithms that underlie social decision making. In addition, the complexity of human social interactions occurring in the natural environment makes it difficult to study them in animal models, whereas the currently available non-invasive methods of measuring brain activity in human subjects do not provide the necessary spatial and temporal resolutions (Box 1). Therefore, studies on the neural basis of strategic decision making in humans would benefit by integrating the results from animal studies, especially given that there is a large overlap between the neural systems involved in social and non-social decision making.

Acknowledgement

We thank Drs. Hackjin Kim and Sunhae Sul for their helpful discussion. This work was supported by the National Institute of Health (R01 DA029330 and R21 MH104460).

Glossary

Beauty contest game	A game in which each player chooses a number between 0 and 100 and the winners are the players with their choice closest to a particular fraction (commonly 2/3) of the average of all choices.
Model-free reinforcement learning	A type of reinforcement learning in which the value functions are updated only for chosen actions and current states exclusively based on the reward prediction error, namely, the difference between the actual and expected rewards.
Model-based reinforcement learning	A type of reinforcement learning in which the value functions are calculated from the knowledge or internal model of the decision maker's environment.
Nash equilibrium	a set of strategies defined for each player in a game such that no individual players can improve their payoffs by deviating from it unilaterally.
Reinforcement learning theory	A theory that provides a collection of algorithms that can be used by an agent trying to maximize their future reward within an environment that has a Markov property, namely, an environment in which the probability of transition to a particular state is entirely determined by the current state and action chosen by the agent.
Stag-hunt game	A game in which each player chooses between a small constant payoff ("hare") and a large but uncertain payoff ("stag") which is earned only when both players choose it and hence cooperate.

References

1. von Neumann, J.; Morgenstern, O. Theory of games and economic behavior. Princeton University Press; 1944.
2. Sutton, RS.; Barto, AG. Reinforcement learning: an introduction. MIT Press; 1998.
3. Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. *Econometrica*. 1979; 47:263–292.
4. Kalenscher T, Pennartz CMA. Is a bird in the hand worth two in the future? The neuroeconomics of intertemporal decision making. *Prog. Neurobiol*. 2008; 84:284–315. [PubMed: 18207301]
5. Cai X, et al. Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron*. 2011; 69:170–182. [PubMed: 21220107]
6. Behrens TEJ, et al. The computation of social behavior. *Science*. 2009; 324:1160–1164. [PubMed: 19478175]
7. Seo H, Lee D. Neural basis of learning and preference during social decision making. *Curr. Opin. Neurobiol*. 2012; 22:990–995.
8. Stanley DA, Adolphs R. Toward a neural basis for social behavior. *Neuron*. 2013; 80:816–826. [PubMed: 24183030]

9. Ruff CC, Fehr E. The neurobiology of rewards and values in social decision making. *Nat. Rev. Neurosci.* 2014; 15:549–562. [PubMed: 24986556]
10. Kurzban R, et al. The evolution of altruism in humans. *Annu. Rev. Psychol.* 2015; 66:575–599. [PubMed: 25061670]
11. Du E, Chang SWC. Neural components of altruistic punishment. *Front. Neurosci.* 2015; 9:26. [PubMed: 25709565]
12. Yoshida W, et al. Game theory of mind. *PLoS Comput. Biol.* 2008; 4:e1000254. [PubMed: 19112488]
13. Griessinger T, Coricelli G. The neuroeconomics of strategic interaction. *Curr. Opin. Behav. Sci.* 2015; 3:73–79.
14. Nash JF. Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. U.S.A.* 1950; 36:48–49. [PubMed: 16588946]
15. Camerer, CF. Behavioral game theory: experiments in strategic interaction. Princeton University Press; 2003.
16. Lee D. Game theory and neural basis of social decision making. *Nat. Neurosci.* 2008; 11:404–409. [PubMed: 18368047]
17. Camerer CF, et al. A psychological approach to strategic thinking in games. *Curr. Opin. Behav. Sci.* 2015; 3:157–162.
18. Kahneman, D.; Tversky, A., editors. Choices, values, and frames. Cambridge Univ. Press; 2000.
19. Gigerenzer, G.; Selton, R., editors. Bounded rationality: the adaptive toolbox. MIT Press; 2001.
20. Stahl DO, Wilson PW. On players' models of other players: theory and experimental evidence. *Games Econ. Behav.* 1995; 10:218–254.
21. Camerer CF, et al. A cognitive hierarchy model of games. *Quart. J. Econ.* 2004; 119:861–898.
22. Coricelli G, Nagel R. Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 2009; 106:9163–9168. [PubMed: 19470476]
23. Puterman, ML. Markov decision processes: discrete stochastic dynamic programming. Wiley; 1994.
24. Lee D, et al. Neural basis of reinforcement learning and decision making. *Annu. Rev. Neurosci.* 2012; 35:287–308. [PubMed: 22462543]
25. Daw ND, et al. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 2005; 8:1704–1711. [PubMed: 16286932]
26. Dezfouli A, Balleine BW. Habits, action sequences and reinforcement learning. *Eur. J. Neurosci.* 2012; 35:1036–1051. [PubMed: 22487034]
27. Doll BB, Daw ND. The ubiquity of model-based reinforcement learning. *Curr. Opin. Neurobiol.* 2012; 22:1075–1081. [PubMed: 22959354]
28. O'Doherty JP, et al. The structure of reinforcement-learning mechanisms in the human brain. *Curr. Opin. Behav. Sci.* 2015; 1:94–100.
29. Erev I, Roth AE. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* 1998; 88:848–881.
30. Lee D, et al. Reinforcement learning and decision making in monkeys during a competitive game. *Cogn. Brain Res.* 2004; 22:45–58.
31. Camerer C, Ho T-H. Experience-weighted attraction learning in normal form games. *Econometrica.* 1999; 67:827–874.
32. Lee D, et al. Learning and decision making in monkeys during a rock-paper-scissors game. *Cogn. Brain Res.* 2005; 25:416–430.
33. Hampton AN, et al. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc. Natl. Acad. Sci. U. S. A.* 2008; 105:6741–6746. [PubMed: 18427116]
34. Abe H, Lee D. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron.* 2011; 70:731–741. [PubMed: 21609828]
35. Zhu L, et al. Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proc. Natl. Acad. Sci. U. S. A.* 2012; 109:1419–1424. [PubMed: 22307594]

36. Suzuki S, et al. Learning to simulate others' decisions. *Neuron*. 2012; 74:1125–1137. [PubMed: 22726841]
37. Daw ND, et al. Model-based influences on humans' choices and striatal prediction errors. *Neuron*. 2011; 69:1204–1215. [PubMed: 21435563]
38. Eppinger B, et al. Of goals and habits: age-related and individual differences in goal-directed decision making. *Front. Neurosci*. 2013; 7:253. [PubMed: 24399925]
39. Lee SW, et al. Neural computations underlying arbitration between model-based and model-free learning. *Neuron*. 2014; 81:687–699. [PubMed: 24507199]
40. Jordan JS. Bayesian learning in normal form games. *Games Econ. Beh.* 1991; 3:60–81.
41. Kalai E, Lehrer E. Rational learning leads to Nash equilibrium. *Econometrica*. 1993; 61:1019–1045.
42. Brown, GW. Iterative solution of games by fictitious play. In: Koopmans, TC., editor. *Activity Analysis of Production and Allocation*. Wiley; 1951. p. 374–376.
43. G006Cäscher J, et al. State versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*. 2010; 66:585–595. [PubMed: 20510862]
44. Seo H, Lee D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci*. 2007; 27:8366–8377. [PubMed: 17670983]
45. Seo H, Lee D. Behavioral and neural changes after gains and losses of conditioned reinforcers. *J. Neurosci*. 2009; 29:3627–3641. [PubMed: 19295166]
46. Kim H, et al. Role of striatum in updating values of chosen actions. *J. Neurosci*. 2009; 29:14701–14712. [PubMed: 19940165]
47. Sul JH, et al. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron*. 2010; 66:449–460. [PubMed: 20471357]
48. Donahue CH, et al. Cortical signals for rewarded actions and strategic exploration. *Neuron*. 2013; 80:223–234. [PubMed: 24012280]
49. Vickery TJ, et al. Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron*. 2011; 72:166–177. [PubMed: 21982377]
50. Vickery TJ, et al. Opponent identity influences value learning in simple games. *J. Neurosci*. 2015; 35:11133–11143. [PubMed: 26245974]
51. Behrens TEJ, et al. Learning the value of information in an uncertain world. *Nat. Neurosci*. 2007; 10:1214–1221. [PubMed: 17676057]
52. Nassar MR, et al. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci*. 2010; 30:12366–12378. [PubMed: 20844132]
53. Bernacchia A, et al. A reservoir of time constants for memory traces in cortical neurons. *Nat. Neurosci*. 2011; 14:366–372. [PubMed: 21317906]
54. Murray JD, et al. A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci*. 2014; 17:1661–1663. [PubMed: 25383900]
55. Rudebeck PH, et al. A role of the macaque anterior cingulate gyrus in social valuation. *Science*. 2006; 313:1310–1312. [PubMed: 16946075]
56. Behrens TEJ, et al. Associative learning of social value. *Nature*. 2008; 456:245–249. [PubMed: 19005555]
57. Chang SW, et al. Neuronal reference frames for social decisions in primate frontal cortex. *Nat. Neurosci*. 2013; 16:243–250. [PubMed: 23263442]
58. Apps MAJ, Ramnani N. The anterior cingulate gyrus signals the net value of others' rewards. *J. Neurosci*. 2014; 34:6190–6200. [PubMed: 24790190]
59. Apps MAJ, et al. Vicarious reinforcement learning signals when instructing others. *J. Neurosci*. 2015; 35:2904–2913. [PubMed: 25698730]
60. Nicolle A, et al. An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron*. 2012; 75:1114–1121. [PubMed: 22998878]
61. Kang P, et al. Dorsomedial prefrontal cortex activity predicts the accuracy in estimating others' preferences. *Front. Hum. Neurosci*. 2013; 7:686. [PubMed: 24324419]

62. Sul S, et al. Spatial gradient in value representation along the medial prefrontal cortex reflects individual differences in prosociality. *Proc. Natl. Acad. Sci. U. S. A.* 2015; 112:7851–7856. [PubMed: 26056280]
63. Hassabis D, et al. Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:1726–1731. [PubMed: 17229836]
64. Hassabis D, Maguire EA. Deconstructing episodic memory with construction. *Trends Cogn. Sci.* 2007; 11:299–306. [PubMed: 17548229]
65. Simon DA, Daw ND. Neural correlates of forward planning in a spatial decision task in humans. *J. Neurosci.* 2011; 31:5528–5539.
66. Doll BB, et al. Model-based choices involve prospective neural activity. *Nat. Neurosci.* 2015; 18:767–772. [PubMed: 25799041]
67. Johnson A, Redish AD. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* 2007; 27:12176–12189. [PubMed: 17989284]
68. Pfeiffer BE, Foster DJ. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature.* 2013; 497:74–79. [PubMed: 23594744]
69. Pezzulo G, et al. Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* 2014; 18:647–657. [PubMed: 25156191]
70. Bhatt MA, et al. Neural signatures of strategic types in a two-person bargaining game. *Proc. Natl. Acad. Sci. U. S. A.* 2010; 107:19720–19725. [PubMed: 21041646]
71. Denny BT, et al. A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J. Cogn. Neurosci.* 2012; 24:1742–1752. [PubMed: 22452556]
72. Mars RB, et al. Connectivity profiles reveal the relationship between brain areas for social cognition in human and monkey temporoparietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 2013; 110:10806–10811. [PubMed: 23754406]
73. Yoshida W, et al. Neural mechanisms of belief inference during cooperative games. *J. Neurosci.* 2010; 30:10744–10751. [PubMed: 20702705]
74. Camille N, et al. The involvement of the orbitofrontal cortex in the experience of regret. *Science.* 2004; 304:1167–1170. [PubMed: 15155951]
75. Coricelli G, et al. Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. Neurosci.* 2005; 8:1255–1262. [PubMed: 16116457]
76. Boorman ED, et al. How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron.* 2009; 62:733–743. [PubMed: 19524531]
77. Boorman ED, et al. Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biol.* 2011; 9:e1001093. [PubMed: 21738446]
78. Lohrenz T, et al. Neural signature of fictive learning signals in a sequential investment task. *Proc. Natl. Acad. Sci. U. S. A.* 2007; 104:9493–9498. [PubMed: 17519340]
79. Hayden BY, et al. Fictive reward signals in the anterior cingulate cortex. *Science.* 2009; 324:948–950. [PubMed: 19443783]
80. Fetsch CR, et al. Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nat. Rev. Neurosci.* 2013; 14:429–442. [PubMed: 23686172]
81. Pouget A, et al. Probabilistic brains: knowns and unknowns. *Nat. Neurosci.* 2013; 16:1170–1178. [PubMed: 23955561]
82. Donoso M, et al. Foundations of human reasoning in the prefrontal cortex. *Science.* 2014; 344:1481–1486. [PubMed: 24876345]
83. Gigerenzer G, Brighton H. Homo heuristicus: why biased minds make better inferences. *Top. Cogn. Sci.* 2009; 1:107–143. [PubMed: 25164802]
84. Shima K, et al. Role for cells in the presupplementary motor area in updating motor plans. *Proc. Natl. Acad. Sci. U. S. A.* 1996; 93:8694–8698. [PubMed: 8710933]
85. Isoda M, Hikosaka O. Switching from automatic to controlled action by monkey medial frontal cortex. *Nat. Neurosci.* 2007; 10:240–248. [PubMed: 17237780]
86. Seo H, et al. Neural correlates of strategic reasoning during competitive games. *Science.* 2014; 346:340–343. [PubMed: 25236468]

87. Chang SWC, Platt ML. Oxytocin and social cognition in rhesus macaques: implications for understanding and treating human psychopathology. *Brain Res.* 2014; 1580:57–68. [PubMed: 24231551]
88. Crockett MJ, Fehr E. Social brains on drugs: tools for neuromodulation in social neuroscience. *Soc. Cogn. Affect. Neurosci.* 2014; 9:250–254. [PubMed: 23887816]
89. Buckner RL, Carroll DC. Self-projection and the brain. *Trends Cogn. Sci.* 2007; 11:49–57. [PubMed: 17188554]
90. Mars RB, et al. On the relationship between the “default mode network” and the “social brain”. *Front. Hum. Neurosci.* 2012; 6:189. [PubMed: 22737119]

Box 1**Outstanding Questions**

- Do the regions in the human brain identified as key for the arbitration between model-free and model-based reinforcement learning algorithms, such as the inferior lateral prefrontal cortex and frontopolar cortex [39], also play a similar role during social interactions? If so, how do these cortical areas interact with those implicated in social cognition, such as TPJ?
- How are the neural signals related to the value functions for alternative actions and predicted behaviors of others influenced by more subjective social cues, such as facial features and personal traits of others, during social interactions?
- Are the weights given to different algorithms of strategic decision making and their parameters under the influence of neuromodulators and hormones, such as oxytocin and testosterone [87, 88]?
- What are the neural substrates of mental simulation during social decision making? Does the recursive reasoning and mental simulation during social decision making depend on the default mode network, including the posterior cingulate cortex [89, 90]?

Trends

- Social decision making relies on recursive inferences about the behaviors of others as well as learning from previous social interactions.
- Multiple areas in the association cortex and basal ganglia contribute to both individual and social decision making.
- Temporal parietal junction and medial prefrontal cortex are specifically involved in accurately predicting the behaviors of others during social interactions.
- Medial prefrontal cortex plays an important role in the arbitration between multiple strategies and learning algorithms during individual and social decision making.

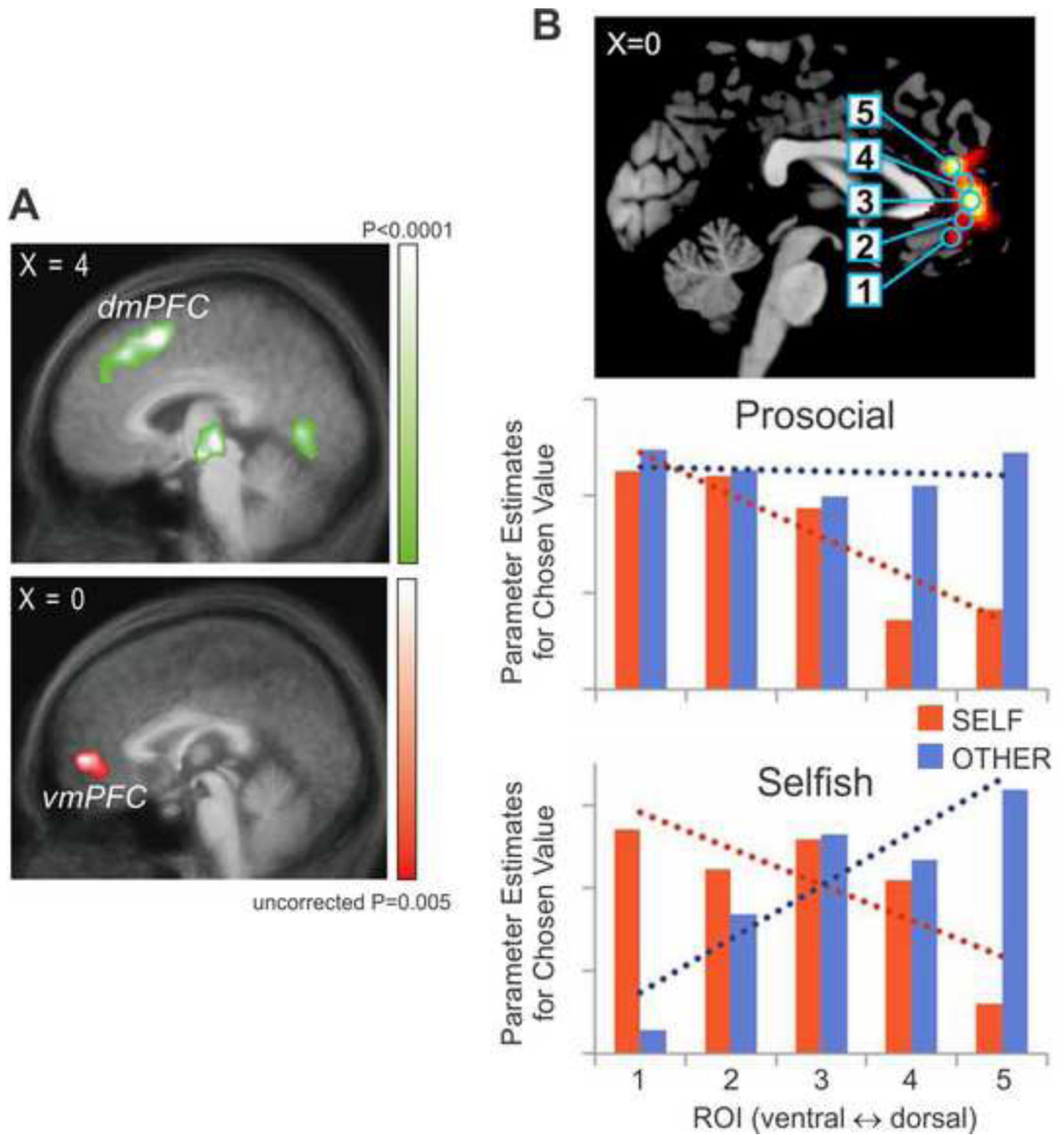


Figure 1. Activity in the medial prefrontal cortex related to learning and simulating the actions of others

A. Activity in the dorsomedial prefrontal cortex correlated with the simulated other’s action prediction errors (top) and activity in the ventromedial prefrontal cortex correlated with the simulated reward prediction error (bottom) [36]. **B.** Spatial gradient in the activity within a set of regions of interests (top) related to the value of chosen option for self vs. others in prosocial (middle) and selfish (bottom) individuals [62].

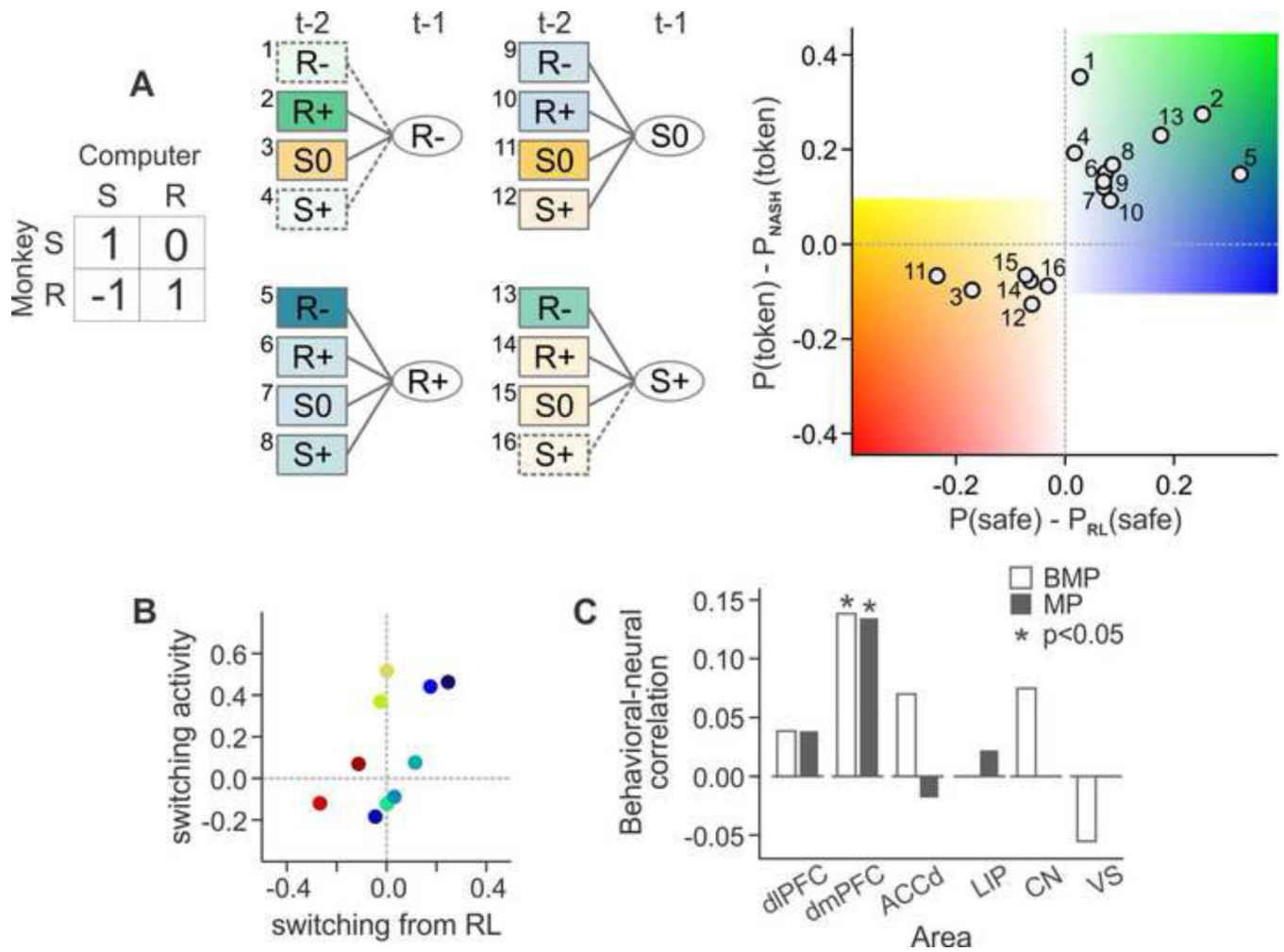


Figure 2. Neuronal activity in the dorsomedial prefrontal cortex related to strategic decision making

A. Monkeys deviated systematically from model-free reinforcement learning to avoid the exploitation by the computer opponent during a biased matching pennies task, in which the payoff to the animal was determined jointly by the choices of the monkey and computer (left). S and R, safe and risky target, respectively, and the number in the payoff matrix indicates the number of tokens received by the animal. The color of each box in the decision tree (middle) and the position of each circle in the scatter plot (right) indicate how much the probability of choosing the safe target following a particular choice–outcome history deviated from a model-free reinforcement learning algorithm, $P(\text{safe}) - P_{RL}(\text{safe})$, and how this increased or decreased the probability of winning compared to the equilibrium strategy, $P(\text{token}) - P_{NASH}(\text{token})$. R–, R+, denote loss and gain from the risky target, whereas S0 and S+ indicate neutral outcome and gain from the safe target. **B.** The strength of neural signals (ordinate) related to the tendency to switch away from a model-free reinforcement learning (abscissa) across different choice–outcome history (shown in **A**) for a single neuron in the dorsomedial prefrontal cortex. **C.** Same neural–behavioral correlation for switching for the population of neurons recorded in the dorsolateral prefrontal cortex (dlPFC), dorsomedial prefrontal cortex (dmPFC), dorsal anterior cingulate cortex (ACCd), lateral intraparietal area

(LIP), caudate nucleus (CN), and ventral striatum (VS). White and gray bars indicate the results obtained from a token-based biased matching pennies task [45, 86] and a symmetric matching pennies task [44, 48], respectively.