Published in final edited form as:
Ann N Y Acad Sci. 2011 December ; 1239: 100–108. doi:10.1111/j.1749-6632.2011.06223.x.

Prefrontal cortex and hybrid learning during iterative competitive games

Hiroshi Abe[1], **Hyojung Seo**[2], and **Daeyeol Lee**[2]

[1]Laboratory of Neurobiology, The Rockefeller University, New York, New York

[2]Department of Neurobiology, Kavli Institute for Neuroscience, Yale University School of Medicine, New Haven, Connecticut

Abstract

Behavioral changes driven by reinforcement and punishment are referred to as simple or model-free reinforcement learning. Animals can also change their behaviors by observing events that are neither appetitive nor aversive, when these events provide new information about payoffs available from alternative actions. This is an example of model-based reinforcement learning, and can be accomplished by incorporating hypothetical reward signals into the value functions for specific actions. Recent neuroimaging and single-neuron recording studies showed that the prefrontal cortex and the striatum are involved not only in reinforcement and punishment, but also in model-based reinforcement learning. We found evidence for both types of learning, and hence hybrid learning, in monkeys during simulated competitive games. In addition, in both the dorsolateral prefrontal cortex and orbitofrontal cortex, individual neurons heterogeneously encoded signals related to actual and hypothetical outcomes from specific actions, suggesting that both areas might contribute to hybrid learning.

Keywords

belief learning; decision making; game theory; reinforcement learning; reward

Hybrid learning during iterative games

Difficulty of choosing an optimal action in a particular situation, namely, an action that produces the outcome most desirable to the decision maker, varies tremendously according to the uncertainty and stability of the decision-maker's environment. Since the environment changes constantly for all animals, and since the outcomes of alternative actions are seldom completely known, animals must always monitor the outcomes of their actions and adjust their estimates appropriately in order to improve their action selection strategies. The reinforcement learning theory characterizes the computational properties of algorithms that can be used to choose optimal actions in a dynamic environment.[1] As in many other models of decision making, reinforcement learning algorithms are based on a set of quantities, referred to as value functions, that correspond to the desirabilities of outcomes expected from a particular action or a particular state of the environment. In this framework, the probability of choosing a particular action increases with the value function associated with that action.

Corresponding author: Daeyeol Lee, Ph.D., Department of Neurobiology, Yale University School of Medicine, 333 Cedar Street, SHM B404, phone: +1 (203) 785-3527, fax: +1 (203) 785-5263, daeyeol.lee@yale.edu.

Conflicts of interest
The authors declare no conflicts of interest.

In the framework of reinforcement learning, specific learning rules describe how the value functions are adjusted through the decision maker's experience. Depending on the type of information utilized to update value functions, two different kinds of learning models can be distinguished. For example, simple or model-free reinforcement learning models utilize only the information about the value of reward or penalty directly received by the decision maker as a result of chosen actions. However, values of such direct reward or penalty correspond to a relatively small portion of information that can be utilized to predict the outcomes of future actions accurately. If a decision maker could learn from merely observing unexpected changes in his or her environment, then this information might be used to revise the value function for a particular action even before or without taking the corresponding action. This implies that the decision maker's newly revised model of his or her environment was used to simulate the hypothetical outcomes of alternative actions. Consequently, these simulated hypothetical outcomes can then be used to update the value functions. Accordingly, this is referred to as model-based reinforcement learning.[1] In both types of reinforcement learning models, the value functions are adjusted according to the discrepancies between the reward received by the decision maker and the reward expected from the current value functions. The difference between real and expected reward is referred to as reward prediction error. In model-based reinforcement learning model, value functions can be updated by the reward prediction error computed using the value functions estimated by the decision maker's model. Moreover, even for actions not chosen by the decision maker, their value functions can be revised according to the difference between hypothetical rewards and the rewards predicted by the current value functions. This is often referred to as fictive or counterfactual reward prediction error.[2,3]

Given that model-based reinforcement learning algorithms can allow decision makers to revise their strategies much more rapidly and flexibly, it is perhaps not surprising that choices in both humans and animals can be better accounted for by model-based reinforcement learning.[4–9] In particular, in social settings, behaviors of other intelligent decision makers can change more frequently and unpredictably than inanimate objects. Therefore, simple learning algorithms that depend only on the actual outcomes of previous actions would be insufficient to utilize a variety of social cues available to infer the intentions and likely behaviors of other individuals. Thus, the ability to deploy model-based reinforcement learning algorithms would be especially advantageous when the outcomes of decisions are determined by the behaviors of multiple decision makers in social settings.[7] During social interactions, decision makers might revise their strategies according to their beliefs about the likely choices of other decision makers, and this is referred to as belief learning in game theory.[10] Therefore, model-based reinforcement learning is analogous to belief learning. In model-based reinforcement learning or belief learning models, the value function for a given action would be updated by the payoff expected for that action, regardless of whether it was actually chosen by the decision maker or not. By contrast, previous studies have demonstrated that during iterative competitive games, behavioral changes in decision makers or players can be accounted for best by a learning model that incorporates the features of both model-free and model-based reinforcement learning.[6,11–13] While hypothetical outcomes associated with unchosen actions still influence the decision maker's future behaviors, actual payoffs from chosen actions tend to do so more strongly. For example, in the experience-weighted attraction (EWA) model of Camerer and Ho,[11] the value function for a given action is updated differently depending on whether that action was actually taken or not. For a chosen action, the value function is updated according to the actual payoff, whereas value functions of unchosen actions are updated according to hypothetical payoffs that would have been obtained from such actions. The learning rate, which controls how rapidly reward prediction errors are incorporated into value functions, can be set differently for actual and hypothetical outcomes. Thus, the EWA is a hybrid

learning model combining the features of both simple reinforcement learning and belief learning.

Neural substrates of simple reinforcement learning

For optimal decision making, two different types of computations are necessary. Prior to committing to a particular action, the desirability of outcomes expected from each action must be estimated. This process is likely to be distributed in multiple brain areas, since single-neuron or metabolic activity related to the subjective values of outcomes expected from different actions and objects has been observed in the prefrontal cortex,[14–20] medial frontal cortex,[21, 22] posterior parietal cortex,[23–26] basal ganglia,[27–35] and amygdala.[36] Once a chosen action is executed, then its outcomes must be monitored, and any discrepancies between the expected and actual outcomes must be taken into consideration to revise the decision-maker's behavioral strategies. Such reward prediction errors are used to update the value functions.[1] Dopamine neurons in the substantia nigra and ventral tegmental area encode the reward prediction errors,[37, 38] although some of them encode rectified reward prediction errors that might correspond to the saliency of sensory information.[39–41]

Many different behavioral tasks have been used to investigate the neural signals related to the subjective values of expected outcomes and reward prediction errors, including Pavlovian conditioning[36, 37] and dynamic foraging tasks.[24] Behavioral tasks simulating competitive games have been also used.[16, 25, 26, 42–46] For many simple competitive games, the optimal strategy is to choose multiple options stochastically and independently across trials. In game theory, a set of strategies is referred to as a Nash equilibrium when no individual players can increase their payoffs by changing their strategies unilaterally. For example, during a matching pennies game, each of the two players chooses one of two options (e.g., heads and tails), and one of them wins when the two choices match and loses otherwise. For this game, there exists a single Nash equilibrium strategy, which is to choose each of the two options with a 0.5 probability. However, the actual choices observed in humans and non-human primates during competitive games often display systematic deviations from Nash equilibrium strategies. For example, during the matching pennies task, both people and monkeys tend to repeat the same choice after winning more frequently than after losing.[11, 47–49] This so-called win-stay-lose-switch strategy is a hallmark of a model-free reinforcement learning algorithm, in which the value functions for actions leading to successes and failures are increased and decreased, respectively. Therefore, near-equilibrium behaviors observed in humans and animals during competitive games might result from a model-free reinforcement learning.

Consistent with these behavioral findings, single-neuron recording studies have found that neural signals related to the value functions for alternative actions estimated by a model-free reinforcement learning algoritihm are distributed in multiple brain areas, including the prefrontal cortex and posterior parietal cortex.[16, 25, 26, 50, 51] In most of these studies, the possibility that the animals might have also utilized model-based reinforcement learning algorithms was not tested. Nevertheless, in order to represent and update value functions for different actions, signals related to the animal's previous choices and their outcomes must be combined appropriately. Indeed, many neurons in the prefrontal cortex and posterior parietal cortex maintained the signals related to the animal's choice and its outcome during several trials (Figure 1).[26, 45, 50, 51] Moreover, it was found that the memory signals related to the animal's previous choices and their outcomes decay with a range of time constants in a population of neurons in multiple cortical areas, suggesting that they might provide the reservoir of different time constants that might be necessary for optimizing the rate of learning as the stability of the animal's environment changes.[52] The signals related to the animal's previous choices and their outcomes have been identified in other brain areas,

including the orbitofrontal cortex and striatum, suggesting that the process of updating the value functions during model-free reinforcement learning is broadly distributed.[53–57]

Prefrontal cortex and hybrid learning

Most animal behaviors can be understood as actions that are selected to maximize the overall desirabilities of expected outcomes. Accordingly, unexpected outcomes of the animal's chosen actions have powerful influence on its subsequent behaviors. However, this does not mean that the animal's behavioral strategies are modified only on the basis of actual outcomes of their actions. When humans and animals acquire a new piece of information about the possible changes in their environments, this can also lead to changes in their decision-making strategies even before they experience unexpected outcomes from their actions. Whereas these two different types of reinforcement learning can be clearly distinguished by their computational characteristics, they might be implemented by a common neural substrate. For example, during a multistep decision making used by Daw and his colleagues,[8] subjects adjusted their behaviors not only according to whether a particular action was followed by reward or not, but also according to whether such reward was obtained following expected sequence of events in the environment or not. These results suggest that a part of the error signals used to modify the subject's behavior was computed using a model-based reinforcement learning algorithm. Moreover, the blood-oxygen-level-dependent (BOLD) signals in the striatum were influenced by both model-free and model-based reward prediction errors.[2, 8] Similarly, subjects performing a navigation task in a dynamic virtual maze largely made their choices using a model-based reinforcement learning algorithm, and the value of the chosen actions estimated by the same model was localized in the striatum.[9]

In our recent study, we tested whether the activity of individual neurons in different regions of the primate prefrontal cortex encoded signals that can be utilized to implement model-based reinforcement learning or hybrid learning.[13] We focused on the dorsolateral prefrontal cortex (DLPFC) and orbitofrontal cortex (OFC, Figure 2A) for several reasons. First, the prefrontal cortex is often considered to play an important role in flexible, context-dependent action planning, which would be facilitated by model-based reinforcement learning,[58, 59] whereas the striatum, especially the dorsolateral striatum, might support model-free reinforcement learning.[60] Second, a number of studies have demonstrated that the prefrontal cortex is involved in evaluating the outcomes expected from different actions and choices[14–17, 61] as well as the actual outcomes resulting from chosen options.[45, 62, 63] In addition, unexpected state transitions in the environment lead to increased BOLD activity in the human prefrontal cortex, even when they are not associated with reinforcement or punishment.[64] In hybrid learning and model-based reinforcement learning, the value functions are updated by actual outcomes from chosen actions as well as hypothetical outcomes associated with unchosen actions. Therefore, these results raise the possibility that neurons encoding both actual and hypothetical outcomes might be involved in updating the value functions according to hybrid learning or model-based reinforcement learning algorithms. Finally, neuroimaging studies have also revealed BOLD activity in DLPFC and OFC related to the experience of regret and relief, namely, the discrepancy between the realized outcome and the outcome that could have been obtained by choosing a different action.[65, 66]

We examined the activity of individual neurons in the DLPFC (n=308) and OFC (n=201, Figure 2A) while the monkeys played a computer-simulated rock-paper-scissors task (Figure 2B).[13] During this task, the animal began each trial by fixating a small central target. After a brief (0.5 s) fore-period, three identical peripheral targets were presented, and the animal was required to shift its gaze towards one of these peripheral targets when the central target

was extinguished 0.5 s later. Once the animal fixated its chosen peripheral target for 0.5 s, all three peripheral targets changed their colors and revealed the amount of juice reward available to the animal from each target (Figure 2D). This was determined by the payoff matrix of a biased rock-paper-scissors game (Figure 2C), and the reward was delivered to the animal 0.5 s after the targets changed their colors. For example, in trials in which the computer chose the target corresponding to rock, the payoff for the animal was 0, 1, or 3 drops of juice (0.2 ml/drop) for choosing the targets corresponding to scissors, rock, and paper, respectively. The positions of the targets for the animals that correspond to rock, paper, and scissors were counter-balanced across blocks, and the computer opponent simulated a rational opponent in a zero-sum game.[6,13]

As described previously,[6, 13] the animal's choices during this rock-paper-scissors task are better accounted for by a hybrid learning model in which value functions for different choices are adjusted not only by the actual outcome of the animal's choice in each trial but also by the hypothetical outcomes from targets not chosen by the animal, better than simple reinforcement learning or belief learning models. Also, neurons in DLPFC and OFC often encoded the magnitude of reward obtained by the animal (Figure 3). For some neurons, activity related to this actual outcome varied significantly across different target locations, suggesting that they might contribute to updating the value functions of different actions according to a model-free reinforcement learning algorithm.[46, 50] The proportion of neurons encoding the actual outcomes was similar in the DLPFC and OFC. For example, 20.5% and 16.4% the neurons in the DLPFC and OFC, respectively, changed their activity significantly according to the actual outcomes differently depending on the position of the target chosen by the animal. In addition, neurons in both DLPFC and OFC modulated their activity according to the hypothetical payoff available from the winning target even during the trials in which the animal did not win (Figure 4). As in hybrid learning, neurons in both areas were more likely to encode the signals related to actual outcomes of the animal's choices than those related to hypothetical outcomes. The overall proportion of neurons encoding the hypothetical outcomes from winning targets was not significantly different for the DLPFC (21.4%) and OFC (16.9%). In addition, the proportion of neurons encoding hypothetical outcomes was significantly higher among those also encoding actual outcomes (χ^2-test, $p<0.001$) in both DLPFC (32.3%) and OFC (25.3%). Therefore, in both areas, individual neurons tended to process the information about both actual and hypothetical outcomes. Notably, DLPFC neurons were more likely to encode hypothetical outcomes differently according to the position of the winning target than OFC neurons (17.2% vs. 8.0%;χ^2-test, $p<0.005$).

Signals related to hypothetical or fictive reward[67] as well as those related to actual reward[45, 62, 68] had been also observed in the anterior cingulate cortex (ACC). Therefore, the information about both actual and hypothetical rewards might be widespread in multiple areas of the frontal cortex. Nevertheless, there is also some evidence that signals related to hypothetical outcomes in different cortical areas might contribute to different functions. For example, neurons in the ACC tended to encode both actual and hypothetical outcomes indiscriminately,[67] whereas neurons in the DLPFC and OFC often encoded actual and hypothetical outcomes that resulted or could have resulted from a particular action.[13] The tendency to encode hypothetical outcomes from specific actions, namely, conjunctions of actions and hypothetical outcomes, was stronger in the DLPFC than in the OFC,[13] suggesting that DLPFC might be particularly important in updating the value functions for various actions according to both actual and hypothetical outcomes, thereby subserving hybrid learning.

Neural basis of counterfactual thinking

Counterfactual thinking refers to the perception and imagination of alternative outcomes that could have been materialized by a course of actions different from the chosen one, and inferences related to such hypothetical outcomes.[69] The fact that neurons in the primate prefrontal cortex encode not only the actual outcomes resulting from the animal's actions but also hypothetical outcomes suggests that the same neural machinery might also underlie counterfactual thinking.[70] Indeed, disruption of prefrontal functions might impair counterfactual thinking.[71] Moreover, patients with schizophrenia also show reduced abilities to utilize counterfactual thinking to improve their behaviors.[72, 73] During the rock-paper-scissors task used to examine the activity of prefrontal cortex, both actual and hypothetical outcomes were revealed to the animal explicitly by visual cues.[13] Nevertheless, convergence of information about actual and hypothetical outcomes in the prefrontal cortex suggests that the prefrontal cortex might mediate counterfactual thinking by aligning and integrating the signals from heterogeneous sources and using them for updating the value functions for different actions and therefore planning future actions. Overall, neural signals related to actual and hypothetical outcomes were distributed similarly in the DLPFC and OFC. However, compared to the DLPFC neurons, OFC neurons were more likely to encode the hypothetical outcomes regardless of the nature of corresponding actions. Therefore, the DLPFC might be more involved in using the information about the hypothetical outcomes to guide the animal's future behaviors, whereas the OFC might contribute more to the emotional aspect of counterfactual thinking.

Acknowledgments

This research was supported by Kavli Institute for Neuroscience at Yale University and US National Institute of Health (DA029330).

References

1. Sutton, RS.; Barto, AG. Reinforcement learning: an introduction. MIT Press; Massachusetts: 1998.
2. Lohrenz T, McCabe K, Camerer CF, Montague PR. Neural signature of fictive learning signals in a sequential investment task. Proc Natl Acad Sci USA. 2007; 104:9493–9498. [PubMed: 17519340]
3. Boorman ED, Behrens TE, Rushworth MF. Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. PLoS Biol. 2011; 9:e1001093. [PubMed: 21738446]
4. Tolman EC. Cognitive maps in rats and men. Psychol Rev. 1948; 55:189–208. [PubMed: 18870876]
5. Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology. 1998; 37:407–419. [PubMed: 9704982]
6. Lee D, McGreevy BP, Barraclough DJ. Learning and decision making in monkeys during a rock-paper-scissors game. Cogn Brain Res. 2005; 25:416–430.
7. Lee D. Game theory and neural basis of social decision making. Nat Neurosci. 2008; 11:404–409. [PubMed: 18368047]
8. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. Neuron. 2011; 69:1204–1215. [PubMed: 21435563]
9. Simon DA, Daw ND. Neural correlates of forward planning in a spatial decision task in humans. J Neurosci. 2011; 31:5526–5539. [PubMed: 21471389]
10. Fudenberg, D.; Levine, DK. The theory of learning in games. MIT Press; Massachusetts: 1998.
11. Camerer C, Ho T-H. Experience-weighted attraction learning in normal form games. Econometrica. 1999; 67:827–874.
12. Camerer, CF. Behavioral game theory: experiments in strategic interaction. Princeton Univ. Press; New Jersey: 2003.
13. 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]

14. Watanabe M. Reward expectancy in primate prefrontal neurons. Nature. 1996; 382:629–632. [PubMed: 8757133]
15. Leon MI, Shadlen MN. Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. Neuron. 1999; 24:415–425. [PubMed: 10571234]
16. Barraclough DJ, Conroy ML, Lee D. Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci. 2004; 7:404–410. [PubMed: 15004564]
17. Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. Nature. 2006; 441:223–226. [PubMed: 16633341]
18. Padoa-Schioppa C. Neurobiology of economic choice: a good-based model. Annu Rev Neurosci. 2011; 34:333–359. [PubMed: 21456961]
19. Luhmann CC, Chun MM, Yi DJ, Lee D, Wang XJ. Neural dissociation of delay and uncertainty in intertemporal choice. J Neurosci. 2008; 28:14459–14466. [PubMed: 19118180]
20. Kim S, Hwang J, Seo H, Lee D. Valuation of uncertain and delayed rewards in primate prefrontal cortex. Neural Netw. 2009; 22:294–304. [PubMed: 19375276]
21. Gläscher J, Hampton AN, O'Doherty JP. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb Cortex. 2009; 19:483–495. [PubMed: 18550593]
22. Sohn J-W, Lee D. Order-dependent modulation of directional signals in the supplementary and pre-supplementary motor areas. J Neurosci. 2007; 27:13655–13666. [PubMed: 18077677]
23. Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. Nature. 1999; 400:233–238. [PubMed: 10421364]
24. Sugrue LP, Corrado GS, Newsome WT. Matching behavior and the representation of value in the parietal cortex. Science. 2004; 304:1782–1787. [PubMed: 15205529]
25. Dorris MC, Glimcher PW. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron. 2004; 44:365–378. [PubMed: 15473973]
26. Seo H, Barraclough DJ, Lee D. Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. J Neurosci. 2009; 29:7278–7289. [PubMed: 19494150]
27. Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol. 2000; 84:3072–3077. [PubMed: 11110834]
28. Knutson B, Adams CM, Fong GW, Hommer D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J Neurosci. 2001; 21:RC159. [PubMed: 11459880]
29. Tom SM, Fox CR, Trepel C, Poldrack RA. The neural basis of loss aversion in decision-making under risk. Science. 2007; 315:515–518. [PubMed: 17255512]
30. Samejima K, Ueda Y, Doya K, Kimura M. Representation of action-specific reward values in the striatum. Science. 2005; 310:1337–1340. [PubMed: 16311337]
31. Lau B, Glimcher PW. Value representations in the primate striatum during matching behavior. Neuron. 2008; 58:451–463. [PubMed: 18466754]
32. Kable JW, Glimcher PW. The neural correlates of subjective value during intertemporal choice. Nat Neurosci. 2007; 10:1625–1633. [PubMed: 17982449]
33. Kable JW, Glimcher PW. The neurobiology of decision: consensus and controversy. Neuron. 2009; 63:733–745. [PubMed: 19778504]
34. Pine A, Seymour B, Roiser JP, Bossaerts P, Friston KJ, Curran HV, Dolan RJ. Encoding of marginal utility across time in the human brain. J Neurosci. 2009; 29:9575–9581. [PubMed: 19641120]
35. Cai X, Kim S, Lee D. Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. Neuron. 2011; 69:170–182. [PubMed: 21220107]
36. Paton JJ, Belova MA, Morrison SE, Salzman CD. The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature. 2006; 439:865–870. [PubMed: 16482160]
37. Schultz W. Predictive reward signal of dopamine neurons. J Neurophysiol. 1998; 80:1–27. [PubMed: 9658025]

38. Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci. 2007; 10:1615–1624. [PubMed: 18026098]
39. Ravel S, Richmond BJ. Dopamine neuronal responses in monkeys performing visually cued reward schedules. Eur J Neurosci. 2006; 24:277–290. [PubMed: 16882024]
40. Joshua M, Adler A, Mitelman R, Vaadia E, Bergman H. Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probablistic classical conditioning trials. J Neurosci. 2008; 28:11673–11684. [PubMed: 18987203]
41. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature. 2009; 459:837–841. [PubMed: 19448610]
42. Cui H, Andersen RA. Posterior parietal cortex encodes autonomously selected motor plans. Neuron. 2007; 56:552–559. [PubMed: 17988637]
43. Thevarajah D, Mikulić A, Dorris MC. Role of the superior colliculus in choosing mixed-strategy saccades. J Neurosci. 2009; 29:1998–2008. [PubMed: 19228954]
44. Vickery TJ, Jiang YV. Inferior parietal lobule supports decision making under uncertainty in humans. Cereb Cortex. 2009; 19:916–925. [PubMed: 18728197]
45. 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]
46. Seo H, Lee D. Behavioral and neural changes after gains and losses of conditioned reinforcers. J Neurosci. 2009; 29:3627–3641. [PubMed: 19295166]
47. Mookherjee D, Sopher B. Learning behavior in an experimental matching pennies game. Games Econ Behav. 1994; 7:62–91.
48. 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.
49. Lee D, Conroy ML, McGreevy BP, Barraclough DJ. Reinforcement learning and decision making in monkeys during a competitive game. Cogn Brain Res. 2004; 22:45–58.
50. Seo H, Lee D. Cortical mechanisms for reinforcement learning in competitive games. Philos Trans R Soc B. 2008; 363:3845–3857.
51. Seo H, Barraclough DJ, Lee D. Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. Cereb Cortex. 2007; 17:i110–i117. [PubMed: 17548802]
52. Bernacchia A, Seo H, Lee D, Wang XJ. A reservoir of time constants for memory traces in cortical neurons. Nat Neurosci. 2011; 14:366–372. [PubMed: 21317906]
53. Simmons JM, Richmond BJ. Dynamic changes in representations of preceding and upcoming reward in monkey orbitofrontal cortex. Cereb Cortex. 2008; 18:93–103. [PubMed: 17434918]
54. Histed MH, Pasupathy A, Miller EK. Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. Neuron. 2009; 63:244–253. [PubMed: 19640482]
55. Kim H, Sul JH, Huh N, Lee D, Jung MW. Role of striatum in updating values of chosen actions. J Neurosci. 2009; 29:14701–14712. [PubMed: 19940165]
56. Sul JH, Kim H, Huh N, Lee D, Jung MW. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. Neuron. 2010; 66:449–460. [PubMed: 20471357]
57. Sul JH, Jo S, Lee D, Jung MW. Role of rodent secondary motor cortex in value-based action selection. Nat Neurosci. 2011 In press.
58. Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 2001; 24:167–202. [PubMed: 11283309]
59. Pan X, Sawa K, Tsuda I, Tsukada M, Sakagami M. Reward prediction based on stimulus categorization in primate lateral prefrontal cortex. Nat Neurosci. 2008; 11:703–712. [PubMed: 18500338]
60. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 2005; 8:1704–1711. [PubMed: 16286932]
61. Kennerley SW, Dahmubed AF, Lara AH, Wallis JD. Neurons in the frontal lobe encode the value of multiple decision variables. J Cogn Neurosci. 2009; 21:1162–1178. [PubMed: 18752411]

62. Kennerley SW, Wallis JD. Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. Eur J Neurosci. 2009; 29:2061–2073. [PubMed: 19453638]

63. Wallis JD, Kennerley SW. Heterogeneous reward signals in prefrontal cortex. Curr Opin Neurobiol. 2010; 20:191–198. [PubMed: 20303739]

64. Gläscher J, Daw N, Dayan P, O'Doherty JP. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron. 2010; 66:585–595. [PubMed: 20510862]

65. Coricelli G, Critchley HD, Joffily M, O'Doherty JP, Sirigu A, Dolan RJ. Regret and its avoidance: a neuroimaging study of choice behavior. Nat Neurosci. 2005; 8:1255–1262. [PubMed: 16116457]

66. Fujiwara J, Tobler PN, Taira M, Iijima T, Tsutsui KI. A parametric relief signal in human ventrolateral prefrontal cortex. Neuroimage. 2009; 44:1163–1170. [PubMed: 18992349]

67. Hayden BY, Pearson JM, Platt ML. Fictive reward signals in the anterior cingulate cortex. Science. 2009; 324:948–950. [PubMed: 19443783]

68. Matsumoto M, Matsumoto K, Abe H, Tanaka K. Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci. 2007; 10:647–656. [PubMed: 17450137]

69. Roese, NJ.; Olson, JM. What might have been: the social psychology of counterfactual thinking. Psychology Press; New York: 1995.

70. Barbey AK, Krueger F, Grafman J. Structured event complexes in the medial prefrontal cortex support counterfactual representations for future planning. Philos Trans R Soc B. 2009; 364:1291–1300.

71. Gomez Beldarrain M, Garcia-Monco JC, Astigarraga E, Gonzalez A, Grafman J. Only spontaneous counterfactual thinking is impaired in patients with prefrontal cortex lesions. Cogn Brain Res. 2005; 24:723–726.

72. Hooker C, Roese NJ, Park S. Impoverished counterfactual thinking is associated with schizophrenia. Psychiatry. 2000; 63:326–335. [PubMed: 11218555]

73. Roese NJ, Park S, Smallman R, Gibson C. Schizophrenia involves impairment in the activation of intentions by counterfactual thinking. Schizophr Res. 2008; 103:343–344. [PubMed: 17600684]

74. Carmichael ST, Price JL. Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. J Comp Neurol. 1994; 346:366–402. [PubMed: 7527805]

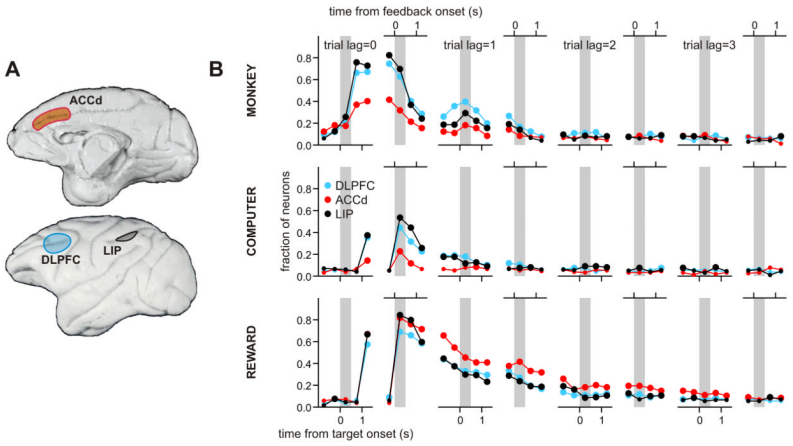

Figure 1.

A. Medial (top) and lateral (bottom) views of the rhesus monkeys' brain, showing the locations of recorded areas in dorsolateral prefrontal cortex (DLPFC),[16, 51] dorsal anterior cingulate cortex (ACCd),[45] and lateral intraparietal cortex (LIP).[26] **B.** Temporal changes in the fraction of neurons significantly modulating their activity according to the animal's choice (top), choice of the computer opponent (equivalent to action-outcome conjunction; middle), and the outcome of the animal's choice (bottom) in the current (trial lag=0) and 3 previous trials (trial lag=1 to 3) during a computer-simulated matching-pennies task. 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).

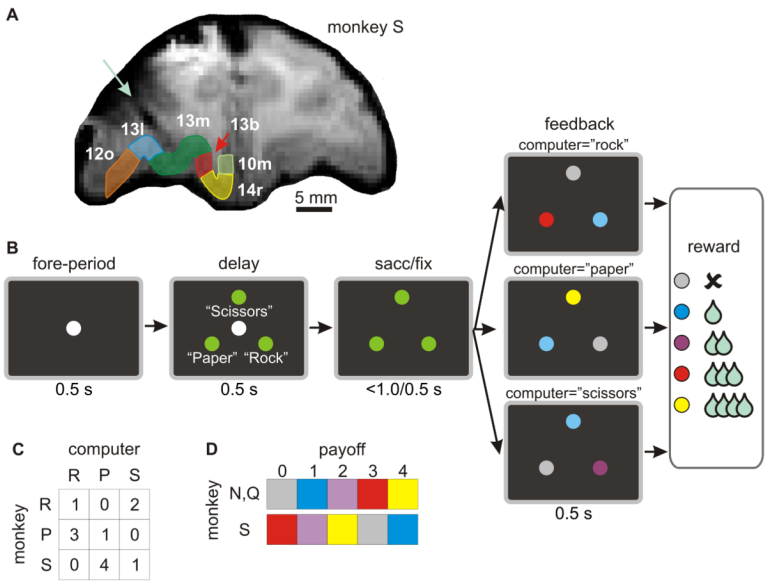

Figure 2.

A. Magnetic resonance image of a rhesus monkey used for neurophysiological recording experiments during a rock-paper-scissors task. Numbers indicate different cytoarchitectonic divisions of the orbitofrontal cortex.[74] A light blue arrow indicates an electrode track. **B.** Temporal sequence of a rock-paper-scissors task used to investigate neuronal signals related to hypothetical outcomes.[13] The amount of reward delivered 0.5 s after feedback onset was determined by the payoff matrix of a biased rock-paper-scissors task (**C**). **D.** Feedback colors used to indicate different payoffs. N, Q, S refer to the three monkeys trained on this task.

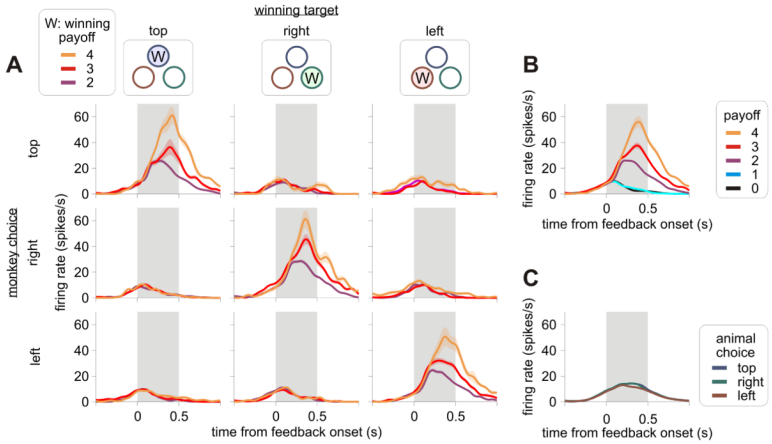

Figure 3.

An example OFC neuron that modulated its activity only according to the actual outcome of the animal's choice. **A.** 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. **B.** Average spike density functions shown as a function of actual payoffs. **C.** Average spike density function shown as a function of the animal's choice.

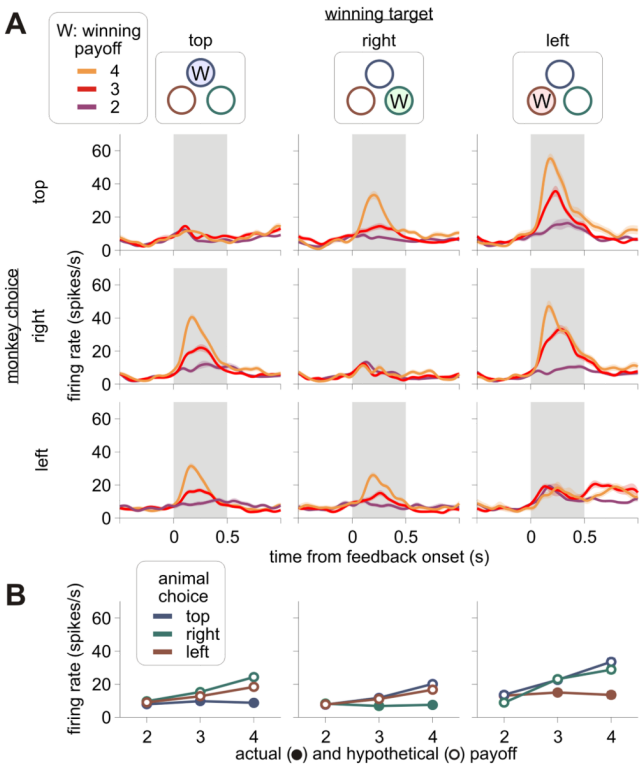

Figure 4.

An example OFC neuron that modulated its activity according to the hypothetical outcome from the winning target. **A.** Same format as in Figure 3A. **B.** The average spike rate estimated separately according to the position of the winning target (columns) and the position of the target chosen by the animal (colors).