

Distributed Coding of Actual and Hypothetical Outcomes in the Orbital and Dorsolateral Prefrontal Cortex

Hiroshi Abe¹ and Daeyeol Lee^{1,2,3,*}

¹Department of Neurobiology

²Kavli Institute for Neuroscience, Yale University School of Medicine

³Department of Psychology

Yale University, New Haven, CT 06510, USA

*Correspondence: daeyeol.lee@yale.edu

DOI 10.1016/j.neuron.2011.03.026

SUMMARY

Knowledge about hypothetical outcomes from unchosen actions is beneficial only when such outcomes can be correctly attributed to specific actions. Here we show that during a simulated rock-paper-scissors game, rhesus monkeys can adjust their choice behaviors according to both actual and hypothetical outcomes from their chosen and unchosen actions, respectively. In addition, neurons in both dorsolateral prefrontal cortex and orbitofrontal cortex encoded the signals related to actual and hypothetical outcomes immediately after they were revealed to the animal. Moreover, compared to the neurons in the orbitofrontal cortex, those in the dorsolateral prefrontal cortex were more likely to change their activity according to the hypothetical outcomes from specific actions. Conjunctive and parallel coding of multiple actions and their outcomes in the prefrontal cortex might enhance the efficiency of reinforcement learning and also contribute to their context-dependent memory.

INTRODUCTION

Human and animals can change their behaviors not only based on the rewarding and aversive consequences of their actions (Thorndike, 1911), but also by simulating the hypothetical outcomes that could have resulted from alternative unchosen actions (Kahneman and Miller, 1986; Lee et al., 2005; Hayden et al., 2009). The internal models about the animal's environment necessary for this mental simulation can be acquired without reinforcement (Tolman, 1948; Fiser and Aslin, 2001). In particular, the ability to incorporate simultaneously actual and hypothetical outcomes expected from chosen and unchosen actions can facilitate the process of finding optimal strategies during social interactions (Camerer, 2003; Gallagher and Frith, 2003; Lee, 2008; Behrens et al., 2009), since observed behaviors of other decision makers can provide the information about the

hypothetical outcomes from multiple actions. However, learning from both real and hypothetical outcomes is not trivial, because these two different types of information need to be linked to different actions correctly. For example, attributing the hypothetical outcomes from unchosen actions incorrectly to the chosen action would interfere with adaptive behaviors (Walton et al., 2010).

Although previous studies have identified neural signals related to hypothetical outcomes in multiple brain areas (Camille et al., 2004; Coricelli et al., 2005; Lohrenz et al., 2007; Chandrasekhar et al., 2008; Fujiwara et al., 2009; Hayden et al., 2009), they have not revealed signals encoding hypothetical outcomes associated with specific actions. Therefore, the neural substrates necessary for learning from hypothetical outcomes remain unknown. In the present study, we tested whether the information about the actual and hypothetical outcomes from chosen and unchosen actions is properly integrated in the primate prefrontal cortex. In particular, the dorsolateral prefrontal cortex (DLPFC) is integral to binding the sensory inputs in multiple modalities appropriately (Prabhakaran et al., 2000), including the contextual information essential for episodic memory (Baddeley, 2000; Mitchell and Johnson, 2009). DLPFC has also been implicated in processing hypothetical outcomes (Coricelli et al., 2005; Fujiwara et al., 2009) and in model-based reinforcement learning (Gläscher et al., 2010). Moreover, DLPFC neurons often change their activity according to the outcomes expected or obtained from specific actions (Watanabe, 1996; Leon and Shadlen, 1999; Matsumoto et al., 2003; Barraclough et al., 2004; Seo and Lee, 2009). Therefore, we hypothesized that individual neurons in the DLPFC might encode both actual and hypothetical outcomes resulting from the same actions and provide the substrate for learning the values of both chosen and unchosen actions. The orbitofrontal cortex (OFC) might be also crucial for behavioral adjustment guided by hypothetical outcome (Camille et al., 2004; Coricelli et al., 2005). However, how and whether OFC contributes to associating actual and hypothetical outcomes with their corresponding actions remains unclear (Tremblay and Schultz, 1999; Wallis and Miller, 2003; Kennerley and Wallis 2009; Padoa-Schioppa and Assad 2006; Tsujimoto et al., 2009; Walton et al., 2010). In the present study, we found that signals related to actual and hypothetical outcomes resulting from specific actions are encoded in both DLPFC and OFC,

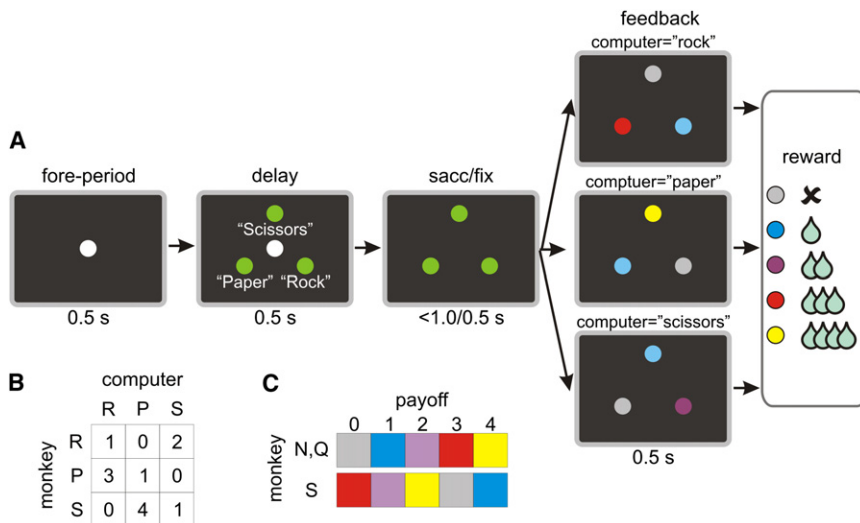


Figure 1. Behavioral Task and Payoffs

(A) Temporal sequence of a rock-paper-scissors task and example feedback displays. The positions of targets corresponding to rock (R), paper (P), scissors (S) are indicated in the frame for the delay period for an illustrative purpose (see also Figure S1).

(B) Payoff matrix in which each number refers to the amount of juice ($\times 0.2$ ml) delivered to the animal.

(C) Feedback colors used to indicate the payoffs.

although OFC neurons tend to encode such outcomes regardless of the animal's actions more than DLPFC neurons.

RESULTS

Effects of Actual and Hypothetical Outcomes on Animal's Behavior

Three monkeys were trained to perform a computer-simulated rock-paper-scissors game task (Figure 1A). In each trial, the animal was required to shift its gaze from the central fixation target toward one of three green peripheral targets. After the animal fixated its chosen target for 0.5 s, the colors of all three targets changed simultaneously and indicated the outcome of the animal's choice as well as the hypothetical outcomes that the animal could have received from the other two unchosen targets. These outcomes were determined by the payoff matrix of a biased rock-paper-scissors game (Figure 1B). For example, the animal would receive three drops of juice when it beats the computer opponent by choosing the "paper" target (indicated by the red feedback stimulus in Figure 1A, top). The computer opponent simulated a competitive player trying to minimize the animal's expected payoff by exploiting statistical biases in the animal's choice and outcome sequences (see Experimental Procedures). The optimal strategy for this game (Nash, 1950) is for the animal to choose "rock" with the probability of 0.5 and each of the remaining targets with the probability of 0.25 (see Supplemental Experimental Procedures available online). In this study, the positions of the targets corresponding to rock, paper, and scissors were fixed in a block of trials and changed unpredictably across blocks (Figure S1). The animal's choice behaviors gradually approached the optimal strategies after each block transition, indicating that the animals adjusted their behaviors flexibly (Figure S2A).

Theoretically, learning during an iterative game can rely on two different types of feedback. First, decision makers can adjust their choices entirely based on the actual outcomes of their previous choices. Learning algorithms exclusively relying on experienced outcomes are referred to as simple or model-free

reinforcement learning (RL) models (Sutton and Barto, 1998). Second, behavioral changes can be also driven by the simulated or hypothetical outcomes that could have resulted from unchosen actions. For example, during social interactions, hypothetical outcomes can be inferred from the choices of other players, and in game theory, this is referred to as belief learning (BL; Camerer, 2003; Gallagher and Frith, 2003; Lee et al., 2005). More generally, learning algorithms relying on simulated outcomes predicted by the decision maker's internal model about the environment are referred to as model-based reinforcement learning (Sutton and Barto, 1998).

Consistent with the predictions from both models, all the animals tested in our study were more likely to choose the same target again after winning than losing or tying in the previous trial (paired t test, $p < 10^{-13}$, for all sessions in each animal; Figure 2A). Moreover, as predicted by the BL model but not by the simple RL model, when the animals lost or tied in a given trial, they were more likely to choose in the next trial what would have been the winning target than the other unchosen target ($p < 10^{-7}$, for all sessions in each animal; Figure 2B), indicating that the animal's choices were also influenced by the hypothetical outcomes from unchosen actions. To quantify the cumulative effects of hypothetical outcomes on the animal's choices, we estimated learning rates for the actual (α_A) and hypothetical (α_H) outcomes from chosen and unchosen actions separately using a hybrid learning model that combine the features of both RL and BL (see Experimental Procedures). For all three animals, the learning rates for hypothetical outcomes were significantly greater than zero (two-tailed t test, $p < 10^{-27}$, for all sessions in each animal), although they were significantly smaller than the learning rates for actual outcomes (paired t test, $p < 10^{-48}$; see Table S1). According to the Bayesian information criterion (BIC), this hybrid learning model and BL model performed better than the RL model in more than 95% of the sessions for each animal. Therefore, animal's behavior was influenced by hypothetical outcomes, albeit less strongly than by actual outcomes. It should be noted that due to the competitive interaction with the computer opponent, the animals did not increase their reward rate by relying on such learning algorithms. In fact, for two monkeys (Q and S), average payoff decreased significantly as they were more strongly influenced by the actual outcomes from their previous choices (see Figure S2B and Supplemental Experimental Procedures). Average payoff was

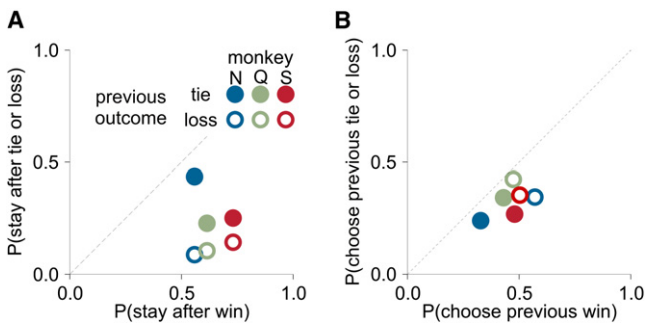


Figure 2. Learning from Actual and Hypothetical Payoffs

(A) Effects of actual payoffs on choices. All three animals were more likely to choose the same target again after winning (abscissa) than losing or tying (ordinate).

(B) Effects of hypothetical payoffs on choices. After both tie and loss trials, all three animals were more likely to choose the target that would have been a winning choice than the other unchosen target (tying and losing target in loss and tie trials, respectively). See also Figure S2 and Table S1.

not significantly related to the learning rates for hypothetical outcomes (Figure S2C).

Coding of Actual and Hypothetical Outcomes in DLPFC and OFC

To test whether and how neurons in different regions of the prefrontal cortex modulate their activity according to the hypothetical outcomes from unchosen actions, we recorded the activity of 308 and 201 neurons in the DLPFC and OFC, respectively, during a computer-simulated rock-paper-scissors game. For each neuron, its activity during the 0.5 s feedback period was analyzed by applying a series of nested regression models that included the animal's choice, actual payoff from the chosen target and hypothetical payoff from the unchosen winning target in a loss or tie trial as independent variables (see [Experimental Procedures](#)). Effects of actual and hypothetical payoffs were examined separately according to whether they were specific for particular actions or not, by testing whether the regressors corresponding to the actual or hypothetical outcomes from specific actions improve the model fit. In the present study, hypothetical outcomes were varied only for the winning targets during tie or loss trials. Therefore, to avoid the confounding of activity related to actual and hypothetical outcomes from different actions, their effects on neural activity was quantified as the activity changes related to the actual and hypothetical payoffs from winning targets only.

Overall, 127 (41.2%) and 91 (45.3%) neurons in DLPFC and OFC, respectively, encoded actual payoffs received by the animal (partial F-test, M3 versus M1, $p < 0.05$; see [Experimental Procedures](#); see Figure S3). In addition, 63 (20.5%) and 33 (16.4%) neurons in DLPFC and OFC significantly changed their activity related to actual outcomes differently according to the animal's chosen actions (M3 versus M2). Thus, the proportion of neurons encoding actual outcomes was not significantly different for DLPFC and OFC, regardless of whether activity related to outcomes from specific choices were considered separately or not (χ^2 test, $p > 0.25$).

Hypothetical payoffs from the winning targets during tie or loss trials were significantly encoded in 66 (21.4%) and 34 (16.9%) neurons in the DLPFC and OFC, respectively (M5 versus M3; see [Experimental Procedures](#)). The proportion of neurons encoding hypothetical outcomes was not significantly different for the two areas (χ^2 test, $p = 0.21$). On the other hand, the proportion of neurons significantly changing their activity related to hypothetical outcomes according to the position of the winning target was significantly higher in the DLPFC ($n = 53$, 17.2%) than in OFC ($n = 16$, 8.0%; χ^2 test, $p < 0.005$). For example, the DLPFC neuron illustrated in Figure 3A increased its activity during the feedback period according to the hypothetical payoff from the upper winning target (partial F-test, $p < 0.05$). This activity change was observed within a set of trials in which the animal's choice of a particular target led to loss or tie (Figure 3A, middle and bottom panels in the first column, respectively), and therefore was not due to the animal's choice of a particular action or its actual outcome. The OFC neuron illustrated in Figure 3B also changed its activity significantly according to the hypothetical winning payoffs, which was significantly more pronounced when the winning target was presented to the left (partial F-test, $p < 0.05$). Nevertheless, the activity related to the hypothetical outcome was qualitatively similar for all three positions of the winning target. The proportion of neurons with significant activity related to hypothetical outcomes was little affected when we controlled for several potential confounding factors, such as the winning payoff expected from the chosen target, the position of the target chosen by the animal in the next trial, and the parameters of saccade during the feedback period of loss trials (Table S2). The results were also largely unaffected when the data were analyzed after removing the first ten trials after each block transition, suggesting that the activity related to hypothetical outcomes were not due to unexpected changes in the payoffs from different target locations. In addition, there was no evidence for anatomical clustering of neurons that showed significant effects of actual or hypothetical outcomes (MANOVA, $p > 0.05$; Figure 4; Figure S4).

To compare the effect size of neural activity related to actual and hypothetical outcomes, the proportion of variance in the spike counts that can be attributed to different outcomes was computed using the coefficient of partial determination (CPD; see [Supplemental Experimental Procedures](#)). The effect size of activity related to actual outcome or hypothetical outcome was significantly larger in the OFC than in DLPFC, when the effects of outcomes from different targets were combined (two-tailed t test, $p < 0.01$; Figure 5A, AON and HON). By contrast, the effect size of activity related to actual or hypothetical outcomes from specific choices was not significantly different for two areas ($p > 0.6$; Figure 5A, AOC and HOC). For each area, we also examined whether the neural activity is more strongly related to a given type of outcomes (i.e., actual or hypothetical) associated with specific actions or not, using the difference in the CPD computed for all actions and those computed for specific actions. For actual outcomes, OFC neurons tended to encode actual outcomes similarly for all actions more than DLPFC (Figure 5B, AOC-AON; $p < 0.01$), whereas DLPFC neurons tended to encode hypothetical outcomes from specific actions more than

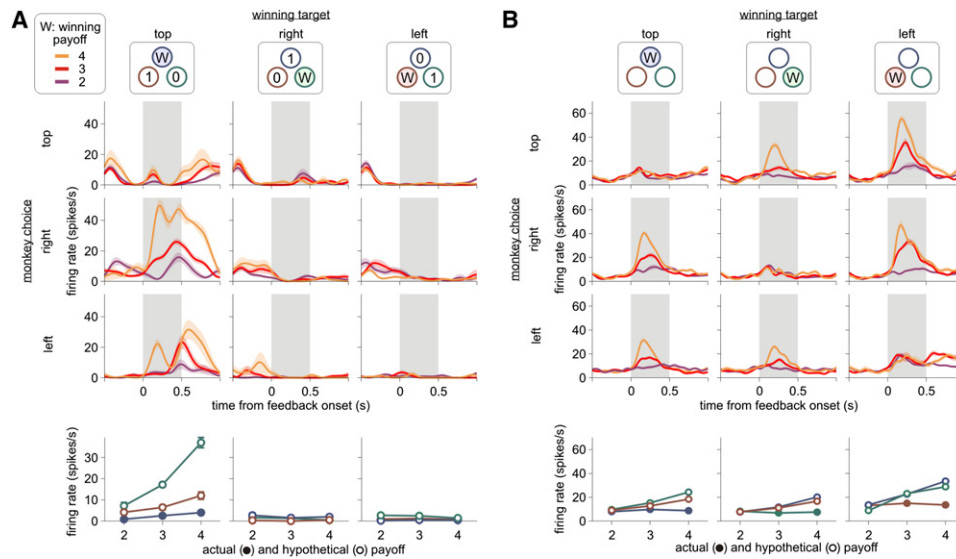


Figure 3. Example Neurons with Activity Related to Hypothetical Outcomes

(A) Spike density functions (SDF, convolved with a Gaussian filter with $\sigma = 40$ ms) of a DLPFC neurons estimated separately according to the position (columns) and payoff (line colors) of the winning target and the position of the target chosen by the animal (rows). Position of the winning target (W) and the payoffs from the other targets are indicated in the top panels, and the mean firing rates during the feedback period (gray shaded regions in the SDF plots) are shown at the bottom. Shaded areas and error bars, SEM.

(B) Activity of an OFC neuron in the same format as in (A). For the neuron in (A), the positions of losing and tying targets were fixed for each winning target (Experiment I), whereas for the neuron in (B), the positions of losing and tying targets were counter balanced (Experiment II). See also Figure S3.

OFC neurons (Figure 5B, HOC–HON; $p < 0.01$). This difference between DLPFC and OFC was statistically significant for both actual and hypothetical outcomes (2-way ANOVA, area \times choice-specificity interaction, $p < 0.05$). Taken together, these results suggest that both DLPFC and OFC play important roles in monitoring actual and hypothetical outcomes from multiple actions, although OFC neurons tend to encode actual and hypothetical outcomes from multiple actions more similarly than DLPFC neurons.

Congruency of Signals Related to Actual and Hypothetical Outcomes

To test whether prefrontal neurons tend to encode actual and hypothetical outcomes from the same action similarly, we estimated the effects of different outcomes separately for individual targets (924 and 603 neuron-target pairs or cases in DLPFC and OFC, respectively; see Experimental Procedures). Overall, 96 (10.4%) and 99 (16.4%) cases in the DLPFC and OFC, respectively, show significant effects of actual outcomes, whereas

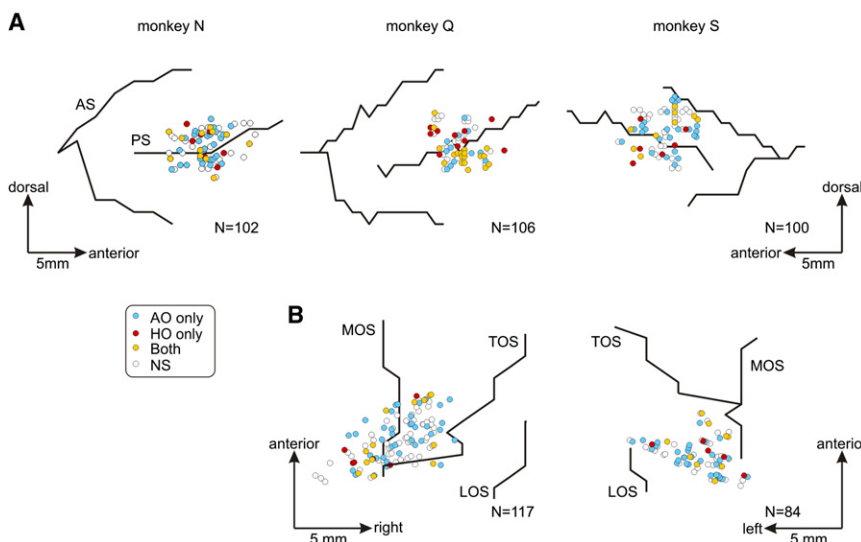


Figure 4. Anatomical Locations of Neurons with Outcome Effect

(A) Locations of DLPFC neurons that showed significant changes in their activity related to actual and hypothetical outcome irrespective of whether they were linked to specific choices or not.

(B) Locations of OFC neurons. The positions of the recorded neurons were estimated and plotted on a horizontal plane. The neurons shown medially from MOS were not in the ventral surface but in the fundus of MOS. MOS, medial orbital sulcus. LOS, lateral orbital sulcus. TOS, transverse orbital sulcus. The number of neurons recorded from each area is shown. See also Figure S4.

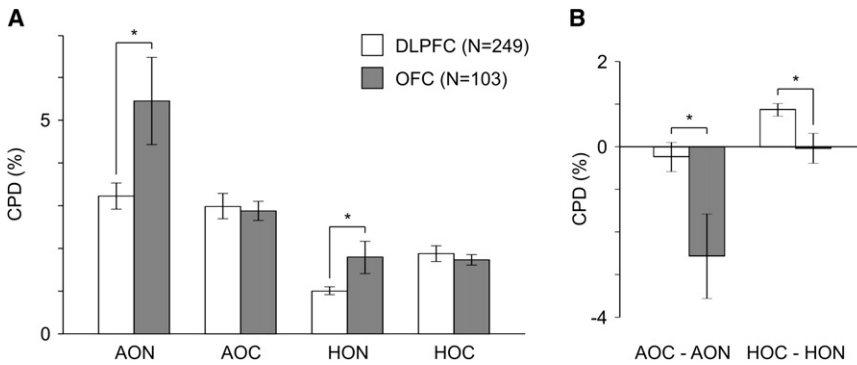


Figure 5. Effect Size for the Activity Related to Actual and Hypothetical Outcomes

(A) CPD related to actual and hypothetical outcomes are shown separately according to whether their effects vary across different actions (AOC/HOC) or not (AON/HON). (B) Difference between choice-specific and choice-unspecific CPD for actual (AOC-AON) and hypothetical (HOC-HON) outcomes. The legend shows the number of neurons included in this analysis. * $p < 0.01$ (two-tailed t test). Error bars, SEM. See also Table S2.

significant effects of hypothetical outcomes were found in 116 (12.6%) and 66 (11.0%) cases in the DLPFC and OFC. Activity increasing with actual winning payoffs was more common in both areas (63 and 69 cases in DLPFC and OFC, corresponding to 65.6% and 69.7%, respectively; binomial test, $p < 0.005$), whereas similar trends for the hypothetical outcomes (68 and 38 cases in DLPFC and OFC, corresponding to 58.6% and 57.6%) were not statistically significant. The effect size (standardized regression coefficients, M_6 ; see [Experimental Procedures](#)) of actual payoff was larger for the neurons increasing their activity with the winning payoff in both DLPFC (0.361 ± 0.010 versus 0.349 ± 0.011) and OFC (0.425 ± 0.016 versus 0.328 ± 0.017), but this was statistically significant only in the OFC (two-tailed t test, $p < 10^{-3}$). The effect size of the activity related to hypothetical outcome was also larger for the neurons increasing activity with the hypothetical winning payoff for DLPFC (0.282 ± 0.009 versus 0.253 ± 0.009) and OFC (0.283 ± 0.018 versus 0.248 ± 0.009), but this was significant only for DLPFC ($p < 0.05$). In addition, neurons in both DLPFC and OFC were significantly more likely to increase their activity with the actual outcomes from multiple targets than expected if the effect of outcomes from individual targets affected the activity of a given neuron independently (binomial test, $p < 0.05$; Table 1). OFC neurons also tended to increase their activity with the hypothetical outcomes from multiple targets ($p < 10^{-6}$; Table 1), whereas this tendency was not significant for DLPFC.

Neural activity leading to the changes in the value functions should change similarly according to the actual and hypothetical outcomes from the same action. Indeed, neurons in both DLPFC and OFC were significantly more likely to increase their activity with both actual and hypothetical outcomes from the same target than expected when the effects of actual and hypothetical outcomes were combined independently (χ^2 test, $p < 10^{-3}$; Table S3). Similarly, the standardized regression coefficients related to the actual and hypothetical outcomes estimated separately for the same target were significantly correlated for the neurons in both areas that showed significant choice-dependent effects of hypothetical outcomes ($r = 0.307$ and 0.318 for DLPFC and OFC, respectively; $p < 0.05$). These neurons also tended to change their activity according to the hypothetical outcomes from a given target similarly regardless of the target chosen by the animal, when tested using the standardized regression coefficient for the hypothetical outcome estimated separately for the

two remaining choices ($r = 0.381$ and 0.770 , for DLPFC and OFC, $p < 0.001$; Figure S5).

For neurons encoding hypothetical outcomes from specific actions, we also estimated the effects of the hypothetical outcomes from two different targets using a set of trials in which the animal chose the same target (see Figure S5). For DLPFC, the correlation coefficient for these two regression coefficients was not significant ($r = -0.042$, $p = 0.64$) and significantly lower than the correlation coefficient computed for the effects of hypothetical outcomes from the same target but with different choices (z-test, $p < 10^{-3}$). By contrast, activity related to the hypothetical outcomes from different choices was significantly correlated for OFC neurons ($r = 0.612$, $p < 10^{-5}$). This correlation coefficient was significantly higher than in DLPFC (z-test, $p < 10^{-4}$), and was not significantly different from the correlation coefficient computed for the effects of hypothetical outcomes

Table 1. Number of Neuron-Target Pairs Showing the Significant Effects of Actual and Hypothetical Outcomes from Different Targets

	DLPFC			OFC		
AO versus AO						
	AO+	A \bar{O}	NS	AO+	A \bar{O}	NS
AO+	9	1	107	31	1	75
AO-	-	8	49	-	9	41
NS	-	-	750	-	-	446
HO versus HO						
	HO+	H \bar{O}	NS	HO+	H \bar{O}	NS
HO+	7	3	119	14	1	47
HO-	-	4	85	-	6	43
NS	-	-	706	-	-	492
AO versus HO						
	HO+	H \bar{O}	NS	HO+	H \bar{O}	NS
AO+	8	6	112	21	6	111
AO-	10	3	53	3	5	52
NS	118	87	1451	52	45	911

For either AO or HO, the total number of cases is 3N, where N is the number of neurons, whereas for AO versus HO, this is 6N, since the effects of AO and HO estimated for two different targets are not symmetric (see also Table S3).

from the same target but with different choices for OFC (z -test, $p = 0.08$). We also found that the actual outcomes from a given target and hypothetical outcomes from the other targets were encoded independently in the DLPFC (Table 1). By contrast, OFC neurons tended to change their activity similarly according to actual and hypothetical outcomes from different targets (χ^2 test, $p < 0.001$).

The fact that DLPFC activity related to the hypothetical outcomes was correlated only for the same target makes it unlikely that such effect arose simply from the visual responses of DLPFC neurons. This is because the geometric relationship between the positions of chosen and unchosen targets in trials used to estimate the activity changes related to hypothetical outcomes was identical, except rotation, when they were compared for the same winning target and for the same choice of the animal (see Figure S5). We also tested whether the activity in DLPFC and OFC tends to change monotonically with hypothetical outcomes. To isolate the effect of hypothetical outcomes, this was tested separately for a set of trials in which the position of winning target as well as the animal's choice and its actual outcome were fixed (2448 and 2412 cases for DLPFC and OFC, respectively; see Experimental Procedures). Among 215 and 219 cases showing significant effects of hypothetical outcomes in the DLPFC and OFC (1-way ANOVA, $p < 0.05$), respectively, the proportion of cases in which activity increased monotonically was 32.1% and 27.9%. This was significantly higher than the chance level (1/6) in both areas (binomial test, $p < 0.001$).

Time Course of Outcome Information

We also found that the information about actual and hypothetical outcomes was processed with a similar time course in both cortical areas. In both areas, neurons tended to display changes in their activity related to actual and hypothetical outcomes within approximately 200 ms from the feedback onset (e.g., Figure 3; Figure S3 and Supplemental Experimental Procedures). The time course of CPD related to the actual and hypothetical outcomes also peaked almost simultaneously after the feedback onset (Figure 6). Moreover, we did not find any statistically significant differences in the latencies of neural activity related to actual and hypothetical outcomes for either cortical area, regardless of whether choice-dependent outcome effects were considered separately or not (Kolmogorov-Smirnov test, $p > 0.3$; Table S4). Consistent with the previous findings (Wallis and Miller, 2003), the latencies for the signals related to actual outcomes in the OFC were significantly shorter than in the DLPFC ($p < 0.05$), whereas the latencies for the signals related to hypothetical was not significantly different for the two areas ($p > 0.7$). The latency of choice-dependent outcome-related activity was not significantly different between the two areas ($p > 0.2$; Table S4).

Feedback-Related Activity and Subsequent Choices

We examined whether the activity of a given neuron during the feedback period was significantly related to the animal's choice in the next trial, after the effects of actual and hypothetical outcomes were accounted for. The number of neurons showing such effects was 15 (4.9%) and 13 (6.5%) in DLPFC and OFC,

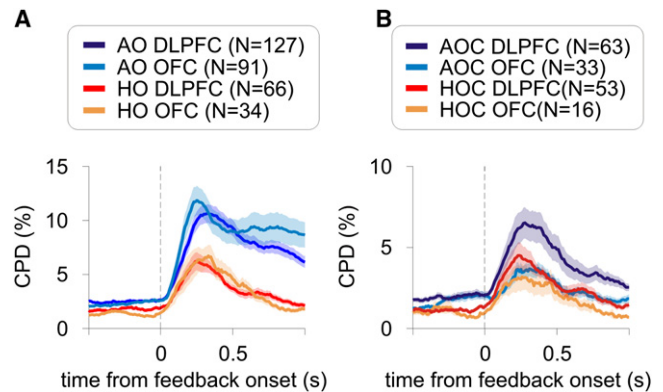


Figure 6. Time Course of Outcome-Related Activity

Time courses of CPD for actual and hypothetical outcomes plotted separately according to whether their effects on neural activity changed significantly for different actions (B) or not (A). Numbers in the parentheses refer to the number of neurons with significant effects in each cortical area. Shaded areas, SEM. See also Table S4.

respectively, and was not significantly higher than expected by chance (binomial test, $p > 0.4$). The proportion of such neurons was not significantly higher even for the neurons that showed significant effect of hypothetical outcomes (χ^2 test, $p > 0.1$, for both cortical areas). Despite the lack of direct linkage between random fluctuation in the activity during the feedback period and the animal's choice in the next trial, neurons in DLPFC and OFC showing outcome-related activity during the feedback period tended to show choice-related activity in other epochs. During the delay period, 34 (11.0%) and 13 (6.5%) neurons in DLPFC and OFC, respectively, changed their activity significantly according to the animal's choice in the same trial, whereas this increased to 179 (58.1%) and 52 (25.9%) during the prefeedback period (Table 2). The difference in the proportion of choice-related activity was significantly different for the two areas during the prefeedback period (χ^2 test, $p < 10^{-12}$), but not during the delay period ($p = 0.08$). DLPFC neurons showing choice-specific effects of actual outcomes during the feedback period were significantly more likely to encode the animal's choice during these two periods (22.2% and 69.8%, respectively; χ^2 test, $p < 0.05$). The number of neurons encoding the animal's choice during the fore-period was relatively low and not significantly different from expected by chance (21 and 10 neurons in DLPFC and OFC, respectively). Nevertheless, OFC neurons encoding actual outcomes or hypothetical outcomes associated with specific actions were significantly more likely to encode the animal's choice during the fore-period (Table 2; $p < 0.05$).

DISCUSSION

Prefrontal Cortex and Reinforcement Learning

Previous studies on the neurobiological substrate of reinforcement learning in animals have almost entirely focused on the behavioral and neural changes associated with actual outcomes, namely reinforcement and punishment. These studies have implicated multiple brain areas including the basal ganglia as the substrates for such learning (Schultz et al., 1997; O'Doherty

Table 2. Number of Neurons Classified by Their Outcome Effect during Feedback and Choice Effect during the Fore and Delay Periods

		Fore Period		Delay		Prefeedback	
		Choice	No Choice	Choice	No Choice	Choice	No Choice
DLPFC	AOC	3	60	14 ^a	49	44 ^a	19
	no AOC	18	227	20	225	135	110
	HOC	3	50	7	46	30	23
	no HOC	18	237	27	228	149	106
OFC	AOC	4 ^a	29	1	32	13	20
	no AOC	6	162	12	156	39	129
	HOC	3 ^a	13	0	16	6	10
	no HOC	7	178	13	172	46	139

Choice effect was tested with one-way ANOVA ($p < 0.05$).

^a $p < 0.05$ (χ^2 test). See also Table S5.

et al., 2004; Daw et al., 2005; Hikosaka et al., 2006; Matsumoto et al., 2007; Graybiel, 2008; Lee, 2008; Seo and Lee, 2009; Kim et al., 2009; Sul et al., 2010). However, actual outcomes represent only a small proportion of information that can be gained after performing an action in real life. In particular, the information about hypothetical outcomes from unchosen alternative actions can be used to revise the animal's internal model of its environment. The results from the present study demonstrate that the neurons in the prefrontal cortex rapidly process the information about the hypothetical outcomes from unchosen actions in addition to the actual outcomes from the animal's chosen action, so that both types of information can be used to update the animal's behavioral strategies (Lee, 2008; Behrens et al., 2009). This suggests a more flexible learning mechanism often referred to as model-based reinforcement learning than a simple, model-free reinforcement learning (Sutton and Barto, 1998; Daw et al., 2005, 2011; Pan et al., 2008; Gläscher et al., 2010).

In the present study, we found that the proportion of the neurons encoding the signals related to actual and hypothetical outcomes was similar for DLPFC and OFC. For actual outcomes, this was true, regardless of whether the signals differentially modulated by the outcomes from specific actions were considered separately or not. By contrast, for hypothetical outcomes, DLPFC neurons were more likely to encode the hypothetical outcomes related to specific actions. The effect size of the signals related to both actual and hypothetical outcomes were larger in the OFC than in the DLPFC, suggesting that OFC might play a more important role in monitoring both actual and hypothetical outcomes. Nevertheless, the difference between these two areas was less pronounced when the activity modulated differentially by the outcomes from different actions was considered separately. In particular, the effect size of the signals related to the hypothetical outcomes from specific choices was not different for the two areas. Thus, the contribution of DLPFC in encoding actual and hypothetical outcomes tends to focus on outcomes from specific choices.

The bias for DLPFC to encode hypothetical outcomes from specific actions is consistent with the previous findings that DLPFC neurons are more likely to encode the animal's actions than OFC neurons. This was true regardless of whether the chosen action was determined by the external stimuli (Tremblay and Schultz, 1999; Ichihara-Takeda and Funahashi, 2008) or

freely by the animal (Wallis and Miller, 2003; Padoa-Schioppa and Assad, 2006; Seo et al., 2007). In addition, DLPFC neurons often encode the specific conjunction of the animal's actions and their outcomes (Barraclough et al., 2004; Seo and Lee, 2009). Nevertheless, the interplay between DLPFC and OFC is likely to contribute to multiple aspects of decision making. For example, neurons in the OFC tend to encode the information about the animal's action and expected outcomes during the time of feedback, and might play an important role in updating the values of different actions (Tsujimoto et al., 2009; Sul et al., 2010). The results from the present study suggest that signals related to the actual and hypothetical outcomes might be combined with those related to the animal's actions, not only in DLPFC but also in OFC. In addition, neurons in both areas often encoded the actual or hypothetical outcomes independent of the animal's action, suggesting that they might contribute to real and fictive reward prediction errors, respectively (Sul et al., 2010; Daw et al., 2011). Neurons coding the animal's action and its actual outcomes have been also found in the medial frontal cortex (Matsumoto et al., 2003; Sohn and Lee, 2007; Seo and Lee, 2009), including the anterior cingulate cortex (Hayden and Platt, 2010). Previous studies have also found that ACC activity during the feedback period tends to be predictive of the animal's subsequent behavior (Shima and Tanji, 1998; Hayden et al., 2009), whereas the present study did not find such activity in DLPFC or OFC. This might be due to the fact that the task used in the present study did not provide any information about the optimal choice in the next trial. Nevertheless, it is also possible that ACC plays a more important role in switching the animal's behavioral strategies than DLPFC and OFC. In addition, neurons in DLPFC and OFC might provide the information about hypothetical outcomes from different actions more specifically than ACC neurons, since ACC neurons respond similarly to the actual and hypothetical outcomes (Hayden et al., 2009), and seldom display multiplicative interactions between actions and hypothetical outcomes (Hayden and Platt, 2010).

Time Course of Outcome Signals in Prefrontal Cortex

Many events in our daily lives, such as the announcement of winning lottery numbers, provide the information about the actual outcomes from chosen actions and hypothetical outcomes from other unchosen actions together. Similarly, the information about

the actual and hypothetical outcomes from chosen and unchosen actions was revealed simultaneously during the behavioral task used in the present study. We found that the information about actual and hypothetical outcome was processed almost simultaneously in the DLPFC and OFC. In contrast, previous studies have shown that in the anterior cingulate cortex, signals related to actual outcomes are processed earlier than those related to hypothetical outcomes (Hayden et al., 2009). This suggests that the information about the actual outcomes is processed immediately in multiple areas of the frontal cortex, while the information about hypothetical outcomes might be processed initially in the DLPFC and OFC and transferred to the anterior cingulate cortex. However, the time course of neural activity related to hypothetical outcomes might be also affected by the behavioral task. In particular, during the task used in the present study, outcomes were revealed following a short delay after the animal's behavioral response, whereas in the previous study on the ACC, the feedback was delivered without any delay after the behavioral response (Hayden et al., 2009). Therefore, the processing of signals related to hypothetical outcomes might be delayed by transient eye movement-related changes in attention (Golomb et al., 2008).

Implications for Episodic Memory and Counterfactual Thinking

OFC lesions lead to deficits during reversal learning in which the subjects are required to learn changing stimulus-reward associations (Izquierdo et al., 2004; Hornak et al., 2004; Tsuchida et al., 2010; Walton et al., 2010), and also impair the abilities to consider anticipated regret during decision making (Camille et al., 2004). Although DLPFC lesions produce more subtle effects on decision making than OFC lesions, DLPFC might be still important for binding various pieces of information in multiple modalities and establish memory traces about the animal's choices and their outcomes in a specific context (Wheeler et al., 1997). Lesions in the prefrontal cortex impair source memory, namely, the ability to recall the context of specific facts and events (Janowsky et al., 1989). In addition, patients with schizophrenia display impaired source memory (Rizzo et al., 1996a; Waters et al., 2004) and difficulties in correctly binding multiple perceptual features (Rizzo et al., 1996b; Burglen et al., 2004), as well as reduced abilities to distinguish between internally and externally generated responses (Bentall et al., 1991), suggesting that such deficits might arise from prefrontal dysfunctions. Therefore, the tendency for neurons in DLPFC to combine the animal's actions and their potential consequences conjunctively (Tanji and Hoshi, 2001; Barraclough et al., 2004; Tsujimoto and Sawaguchi, 2005) might underlie the role of this region in episodic memory (Baddeley, 2000).

Prefrontal cortex, including both DLPFC and OFC, might provide the anatomical substrates for counterfactual thinking, namely, the ability to simulate the potential outcomes of their actions without directly experiencing them. In the present study, hypothetical outcomes were indicated explicitly by visual cues. Nevertheless, prefrontal cortex, especially DLPFC, might be generally involved in updating the animal's decision-making strategies based on the outcomes predicted from the animal's previous experience through analogy and other abstract rules

(Miller and Cohen, 2001; Pan et al., 2008). In fact, patients with prefrontal lesions or schizophrenia tend to display less counterfactual thinking compared to control subjects (Hooker et al., 2000; Gomez Beldarrain et al., 2005) and are impaired in forming intentions based on counterfactual thinking (Roese et al., 2008). Thus, DLPFC might play a comprehensive role in monitoring the changes in the environment of decision makers resulting from their own actions and using this information to optimize decision-making strategies (Knight and Grabowecy, 1995).

EXPERIMENTAL PROCEDURES

Animal Preparation and Data Collection

Three male rhesus monkeys (N, Q, and S, body weight = 10~11 kg) were used. The animal's eye position was sampled at 225 Hz with an infrared eye tracker system (ET49, Thomas Recording, Germany). Single-unit activity was recorded from the DLPFC (monkeys N and Q, right hemisphere; monkey S, left hemisphere) and OFC (monkey Q, right hemisphere; monkey S, left hemisphere) using a multielectrode recording system (Thomas Recording, Germany) and a multichannel acquisition processor (Tucker-Davis Technologies, FL, or Plexon, TX). All isolated single-neuron activity was recorded without screening them for task-related activity. For the OFC recording, a cannula was used to guide the electrodes, and neurons were recorded largely from Walker's area 11 and 13. The 3D positions of the recorded neurons were estimated according to the depth of the electrode tip and the position and tilt of the recording chamber. This reconstruction was aided by MR images with an electrode inserted through the recording chamber and sulcal landmarks identified while recording from DLPFC. The center (AP, ML in mm) of the chamber was (37, 22), (32, 17) and (34, 16.5) for monkey N, Q, and S, respectively. All the procedures for animal care and experiments were approved by the Institutional Animal Care and Use Committee at Yale University.

Behavioral Task: Rock-Paper-Scissors Game

The animals performed an oculomotor free-choice task simulating a biased rock-paper-scissors game, in which the actual outcomes of the animal's chosen actions and hypothetical outcomes for unchosen winning targets were manipulated separately (Figure 1). Each trial began when the animal fixated a small white disk (1° diameter) at the center of a computer screen. After a 0.5 s fore-period, three peripheral targets (green disks, 1° diameter) were presented (eccentricity = 5°), and the animal was required to shift its gaze toward one of the peripheral targets within 1 s when the central target was extinguished 0.5 s later. Once the animal fixated its chosen target for 0.5 s, all peripheral targets simultaneously changed their colors to indicate their corresponding payoffs according to the payoff matrix of a biased rock-paper-scissors game (Figure 1B). The animal was required to maintain fixation on its chosen target for additional 0.5 s before juice reward was delivered. The amount of juice was determined by the values in the payoff matrix ($\times 0.2$ ml). The positions of targets corresponding to rock, paper, and scissors were fixed within a block, and changed across blocks. In both experiments I and II, each neuron was tested at least for six blocks. In experiment I (102 and 106 neurons from the DLPFC of the monkeys N and Q), each of three target configurations was tested in two separate blocks with their order randomized, whereas in experiment II (100 neurons from the DLPFC of monkey S, and 117 and 84 neurons from the OFC of monkeys Q and S), each of six target configurations was tested at least once (see Figure S1). For some neurons recorded in experiment II (41 DLPFC and 103 OFC neurons), each of six configuration was tested in two separate blocks (a total of 12 blocks). The average number of trials tested for each neuron was 422.1 ± 6.1 and 494.8 ± 12.2 for DLPFC and OFC, respectively. The results from the DLPFC in the two experiments did not show any qualitative differences, and were combined.

In experiment I, the number of trials in each block was given by $50 + N_e$, where $N_e \sim \text{exp}(0.05)$, truncated at 50, resulting in 67.1 trials/block on average. In experiment II, each recording session consisted of 6 or 12 blocks, and the number of trials in a block was given by $50 + N_e$, where $N_e \sim \text{exp}(0.2)$ truncated at 20, resulting in 53.9 trials/block on average. The feedback colors

corresponding to different payoffs were counterbalanced across monkeys (Figure 1C). In 20% of tie trials in experiment II, both of the unchosen targets (corresponding to win and loss) changed their colors to red (corresponding to zero payoff) during the feedback period. The results from these control trials were included in all the analyses by assigning 0 to hypothetical payoffs from the winning target. All other aspects of experiment I and II were identical. In both experiments, the computer opponent saved and analyzed the animal's choice and outcome history online and exploited any statistical biases in the animal's behavioral strategy significantly deviating from the optimal (Nash-equilibrium) strategy (analogous to algorithm 2 in Lee et al., 2005; see Supplemental Experimental Procedures). The experimental task was controlled and all the data stored using a Windows-based custom application.

Analysis of Behavioral Data

Choice data from each animal were analyzed with a series of learning models (Sutton and Barto, 1998; Lee et al., 2004, 2005). In all of these models, the value function $V(x)$ for action x was updated after each trial according to (real or hypothetical) reward prediction error, namely the difference between $V(x)$ and (real or hypothetical) reward for the same action, $R(x)$, namely, $V(x) \leftarrow V(x) + \alpha \{R(x) - V(x)\}$, where α is the learning rate. In a simple reinforcement learning (RL) model, the value function was updated only for the chosen action according to the actual payoff received by the animal. By contrast, in a hybrid learning (HL) model, the value functions were updated simultaneously for both chosen and unchosen actions, but with different learning rates for actual and hypothetical outcomes (α_A and α_H , respectively). Finally, a belief learning (BL) model learns the probability for each choice of the opponent, and uses this information to compute the expected payoff from the decision maker's own choice. Formally, this is equivalent to adjusting the value functions for both chosen and unchosen actions according to their actual and hypothetical payoffs, respectively, using the same learning rate (Camerer, 2003). Therefore, both RL and BL are special cases of HL (i.e., $\alpha_H = 0$ and $\alpha_A = \alpha_H$ for RL and BL, respectively).

For all three models, the probability of choosing action x , $p(x)$, was given by the softmax transformation, namely, $p(x) = \exp\{\beta V(x)\} / \sum_y \exp\{\beta V(y)\}$, where $y = \text{top, right, or left}$, and β is the inverse temperature. In addition, for each of these models, we tested the effect of adding a set of fixed choice biases. In these models, $p(x) = \exp\{\beta V(x) + b_x\} / \sum_y \exp\{\beta V(y) + b_y\}$, where b_R and b_P measure the biases for rock and paper relative to scissors, and $b_x = b_R, b_P$ and 0 for $x = \text{rock, paper, and scissors}$, respectively. The likelihood of each model was defined as the product of predicted probabilities for the targets chosen by the animal in each session. The maximum likelihood estimates for model parameters were estimated using fminsearch in Matlab (Mathworks). To compare model performance, we used the Bayesian information criterion (BIC), which is defined as $-2 \ln L + k \ln N$, where L is the likelihood of the model, k the number of model parameters (2, 2, and 3 for RL, BL, and HL models, respectively, which increased to 4, 4, and 5 for the models with choice bias terms), and N the number of trials in a given session. All the results are presented in means \pm SEM, unless indicated otherwise.

Analysis of Neural Data

The firing rates during the 0.5 s feedback period of each neuron were analyzed by applying a series of nested regression models that included various terms related to the animal's choice (CH), actual outcomes (AO), and hypothetical outcomes (HO). Effects of actual and hypothetical outcomes on neural activity were evaluated separately according to whether such effects change with the animal's choices (AO_C and HO_C) or not (AO_N and HO_N). Specifically, these terms were defined as follows.

$$\begin{aligned} CH &= a_0 + a_R C_R + a_L C_L \\ AO_N &= b_{\text{tie}} O_{\text{tie}} + b_{\text{win}} O_{\text{win}} + b_{\text{WP}} (O_{\text{win}} \times P_{\text{win}}), \\ AO_C &= b_{\text{tie}/R} (O_{\text{tie}} \times C_R) + b_{\text{win}/R} (O_{\text{win}} \times C_R) + b_{\text{WP}/R} (O_{\text{win}} \times P_{\text{win}} \times C_R) \\ &\quad + b_{\text{tie}/L} (O_{\text{tie}} \times C_L) + b_{\text{win}/L} (O_{\text{win}} \times C_L) + b_{\text{WP}/L} (O_{\text{win}} \times P_{\text{win}} \times C_L), \\ HO_N &= c_{\text{loss}} (O_{\text{loss}} \times P_{\text{win}}) + c_{\text{tie}} (O_{\text{tie}} \times P_{\text{win}}), \\ HO_C &= c_{\text{loss}/R} (O_{\text{loss}} \times P_{\text{win}} \times W_R) + c_{\text{tie}/R} (O_{\text{tie}} \times P_{\text{win}} \times W_R) \\ &\quad + c_{\text{loss}/L} (O_{\text{loss}} \times P_{\text{win}} \times W_L) + c_{\text{tie}/L} (O_{\text{tie}} \times P_{\text{win}} \times W_L) \end{aligned}$$

where C_X and O_Y denote a series of dummy variables indicating the animal's choice and its outcome ($C_X = 1$ when target X was chosen, and 0 otherwise,

where $X = T, R,$ or L , corresponding to top, right, or left; $O_Y = 1$ when the outcome was Y , and 0 otherwise, where $Y = \text{win, tie, or loss}$), and W_X a dummy variable indicating the winning target ($W_X = 1$ when X was the winning target, and 0 otherwise, where $X = T, R,$ or L). Since there were three choice targets and the intercept (a_0) is included in the regression models, coefficients associated with two choice variables (C_R and C_L) measures the changes in neural activity when the animal chooses the right or left target, compared to when the animal chooses the upper target. P_{win} denotes the payoff from the winning target in each trial ($P_{\text{win}} = 2, 3,$ or 4). Accordingly, the regression coefficient for the interaction term $O_{\text{win}} \times P_{\text{win}}$ in AO_N measures the effect of actual payoff from the winning target, whereas the regression coefficient for $O_{\text{loss}} \times P_{\text{win}}$ in HO_N measures the effect of hypothetical payoff from the winning target in a loss trial. Similarly, the coefficient for $O_{\text{win}} \times P_{\text{win}} \times C_X$ quantifies the effect of actual payoff from the target X in a winning trial, whereas the coefficients for $O_{\text{tie}} \times P_{\text{win}} \times W_X$ and $O_{\text{loss}} \times P_{\text{win}} \times W_X$ measure the effect of hypothetical payoff from the winning target in tie and loss trials, respectively. Using these five different groups of regressors, a set of nested regression models (M1 through M5) was constructed to analyze the firing rate, y .

$$\begin{aligned} M1 : & y = CH \\ M2 : & y = CH + AO_N \\ M3 : & y = CH + AO_N + AO_C \\ M4 : & y = CH + AO_N + AO_C + HO_N \\ M5 : & y = CH + AO_N + AO_C + HO_N + HO_C. \end{aligned}$$

None of the variables related to the actual outcome of the animal's choice were included in M1, whereas all of them were included in M3. Therefore, a given neuron was considered encoding actual outcomes, if the neural activity was better accounted for by M3 than by M1 (partial F-test, $p < 0.05$; Kutner et al., 2005). Similarly, a neuron was considered encoding hypothetical outcomes if M5 accounted for the firing rates better than M3. Whether a given neuron differentially modulated their activity according to the actual outcomes from specific targets was tested by comparing M2 and M3, whereas the effects of hypothetical outcomes related to specific targets were evaluated by comparing M4 and M5 (partial F-test, $p < 0.05$).

In the analyses described above (M1 through M5), the regressors related to actual or hypothetical outcomes and their conjunctions with the animal's choice were introduced separately to test whether neural activity was differentially modulated by the outcomes from different actions. To estimate the effect of actual winning payoff from each target on neural activity, we applied the following model separately to a set of winning trials in which the animal chose a particular target.

$$M6 : y = b_0 + b_q Q_{\text{win}},$$

where Q_{win} denotes the winning payoff from the chosen target ($Q_{\text{win}} = 2, 3,$ or 4). Similarly, the effect of the hypothetical payoff from a given target was estimated by applying the following model to a different subset of trials in which the animal chose one of the remaining two targets and did not win (lost or tied).

$$M7i : y = b_0 + b_u U + b_h H_{\text{win}},$$

where U is the dummy variable indicating which of the two remaining targets was chosen by the animal (e.g., $U = 0$ and 1 for the left and right targets, respectively, when analyzing the trials with the winning target at the top), and H_{win} now denotes the hypothetical payoff from the unchosen winning target (2, 3, or 4). For experiment I, it was not necessary to introduce a separate regressor for the actual outcome in this model (M7i), because the animal's choice also determined the actual payoff (see the top panels in Figure 3). In contrast, for experiment II, it is necessary to factor out the changes in neural activity related to the animal's choice and its actual outcome separately. Therefore, the following model was applied to estimate the effect of the hypothetical payoff in experiment II.

$$M7ii : y = U_1 \times (b_{\text{loss}1} O_{\text{loss}} + b_{\text{tie}1} O_{\text{tie}}) + U_2 \times (b_{\text{loss}2} O_{\text{loss}} + b_{\text{tie}2} O_{\text{tie}}) + b_h H_{\text{win}},$$

where U_1 and U_2 are the dummy variables indicating animals' choice which resulted in loss or tie. The effect size for the activity related to actual and hypothetical outcomes are estimated using the standardized regression

coefficients. To see how the activity changes related to the actual and hypothetical outcomes are related, we calculated the correlation coefficient between the standardized regression coefficients for b_q (M6) and b_h (M7i or M7ii) across a set of neuron-target pairs.

To test whether the activity related to the hypothetical outcomes from a particular target changed with the animal's choice (Figure S5), the following model was applied separately for each combination of chosen and unchosen targets in loss and tie trials for experiment I.

$$M8i: y = b_o + b_h H_{win}$$

For experiment II, another regressor was included to factor out the effect of actual outcome from the chosen target.

$$M8ii: y = b_o + b_{loss} O_{loss} + b_h H_{win}$$

Then, the correlation coefficient between the standardized regression coefficients (b_h) estimated for two different choices was calculated for the same unchosen winning target. As a control analysis, we also calculated the correlation coefficient between the regression coefficients associated with the same chosen target but two different unchosen winning targets. The angular difference in the retinal positions of the unchosen targets during the feedback period was matched for these two analyses (Figure S5). Therefore, if the activity related to hypothetical outcome merely reflected the properties of visual receptive fields, these two correlation coefficients would be similar.

To test whether the neurons significantly modulating their activity according to a particular factor (e.g., AOC or HO) are anatomically segregated from the remaining neurons, MANOVA was applied to their anatomical locations with the statistical significance as the factor (Figure 3; Figure S3). For this analysis, neurons recorded in all the animals were combined separately for the DLPFC and OFC.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, four tables, and Supplemental Experimental Procedures and can be found with this article online at doi: 10.1016/j.neuron.2011.03.026.

ACKNOWLEDGMENTS

We thank Irina Bobeica, Mark Hammond, and Patrice Kurnath for their technical assistance. This work was supported by Kavli Institute for Neuroscience at Yale University and US National Institute of Health grants (DA029330 and EY000785).

Accepted: March 30, 2011

Published: May 25, 2011

REFERENCES

- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423.
- Barracough, D.J., Conroy, M.L., and Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neurosci.* 7, 404–410.
- Behrens, T.E.J., Hunt, L.T., and Rushworth, M.F.S. (2009). The computation of social behavior. *Science* 324, 1160–1164.
- Bentall, R.P., Baker, G.A., and Havers, S. (1991). Reality monitoring and psychotic hallucinations. *Br. J. Clin. Psychol.* 30, 213–222.
- Burglen, F., Marczewski, P., Mitchell, K.J., van der Linden, M., Johnson, M.K., Danion, J.-M., and Salamé, P. (2004). Impaired performance in a working memory binding task in patients with schizophrenia. *Psychiatry Res.* 125, 247–255.
- Camerer, C.F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, NJ: Princeton Univ. Press).
- Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.R., and Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304, 1167–1170.
- Chandrasekhar, P.V.S., Capra, C.M., Moore, S., Noussair, C., and Berns, G.S. (2008). Neurobiological regret and rejoice functions for aversive outcomes. *Neuroimage* 39, 1472–1484.
- Coricelli, G., Critchley, H.D., Joffily, M., O'Doherty, J.P., Sirigu, A., and Dolan, R.J. (2005). Regret and its avoidance: a neuroimaging study of choice behavior. *Nat. Neurosci.* 8, 1255–1262.
- Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.* 8, 1704–1711.
- Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 1204–1215.
- Fiser, J., and Aslin, R.N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* 12, 499–504.
- Fujiwara, J., Tobler, P.N., Taira, M., Iijima, T., and Tsutsui, K. (2009). A parametric relief signal in human ventrolateral prefrontal cortex. *Neuroimage* 44, 1163–1170.
- Gallagher, H.L., and Frith, C.D. (2003). Functional imaging of 'theory of mind'. *Trends Cogn. Sci.* 7, 77–83.
- Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* 66, 585–595.
- Golomb, J.D., Chun, M.M., and Mazer, J.A. (2008). The native coordinate system of spatial attention is retinotopic. *J. Neurosci.* 28, 10654–10662.
- Gomez Beldarrain, M., Garcia-Monco, J.C., Astigarraga, E., Gonzalez, A., and Grafman, J. (2005). Only spontaneous counterfactual thinking is impaired in patients with prefrontal cortex lesions. *Brain Res. Cogn. Brain Res.* 24, 723–726.
- Graybiel, A.M. (2008). Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* 31, 359–387.
- Hayden, B.Y., Pearson, J.M., and Platt, M.L. (2009). Fictive reward signals in the anterior cingulate cortex. *Science* 324, 948–950.
- Hayden, B.Y., and Platt, M.L. (2010). Neurons in anterior cingulate cortex multiplex information about reward and action. *J. Neurosci.* 30, 3339–3346.
- Hikosaka, O., Nakamura, K., and Nakahara, H. (2006). Basal ganglia orient eyes to reward. *J. Neurophysiol.* 95, 567–584.
- Hooker, C., Roese, N.J., and Park, S. (2000). Impoverished counterfactual thinking is associated with schizophrenia. *Psychiatry* 63, 326–335.
- Hornak, J., O'Doherty, J., Bramham, J., Rolls, E.T., Morris, R.G., Bullock, P.R., and Polkey, C.E. (2004). Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *J. Cogn. Neurosci.* 16, 463–478.
- Ichihara-Takeda, S., and Funahashi, S. (2008). Activity of primate orbitofrontal and dorsolateral prefrontal neurons: effect of reward schedule on task-related activity. *J. Cogn. Neurosci.* 20, 563–579.
- Izquierdo, A., Suda, R.K., and Murray, E.A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.* 24, 7540–7548.
- Janowsky, J.S., Shimamura, A.P., and Squire, L.R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia* 27, 1043–1056.
- Kahneman, D., and Miller, D.T. (1986). Norm theory: comparing reality to its alternatives. *Psychol. Rev.* 93, 136–153.
- Kennerley, S.W., and Wallis, J.D. (2009). Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. *Eur. J. Neurosci.* 29, 2061–2073.
- Kim, H., Sul, J.H., Huh, N., Lee, D., and Jung, M.W. (2009). Role of striatum in updating values of chosen actions. *J. Neurosci.* 29, 14701–14712.
- Knight, R.T., and Grabowecky, M. (1995). Escape from linear time: prefrontal cortex and conscious experience. In *The Cognitive Neuroscience*, M.S. Gazzaniga, ed. (Cambridge, MA: MIT Press), pp. 1357–1371.

- Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005). *Applied Linear Statistical Models* (New York: McGraw-Hill).
- Lee, D. (2008). Game theory and neural basis of social decision making. *Nat. Neurosci.* *11*, 404–409.
- Lee, D., Conroy, M.L., McGreevy, B.P., and Barraclough, D.J. (2004). Reinforcement learning and decision making in monkeys during a competitive game. *Brain Res. Cogn. Brain Res.* *22*, 45–58.
- Lee, D., McGreevy, B.P., and Barraclough, D.J. (2005). Learning and decision making in monkeys during a rock-paper-scissors game. *Brain Res. Cogn. Brain Res.* *25*, 416–430.
- Leon, M.I., and Shadlen, M.N. (1999). Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* *24*, 415–425.
- Lohrenz, T., McCabe, K., Camerer, C.F., and Montague, P.R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proc. Natl. Acad. Sci. USA* *104*, 9493–9498.
- Matsumoto, K., Suzuki, W., and Tanaka, K. (2003). Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* *301*, 229–232.
- Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.* *10*, 647–656.
- Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* *24*, 167–202.
- Mitchell, K.J., and Johnson, M.K. (2009). Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? *Psychol. Bull.* *135*, 638–677.
- Nash, J.F. (1950). Equilibrium points in n -person games. *Proc. Natl. Acad. Sci. USA* *36*, 48–49.
- O’Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* *304*, 452–454.
- Padoa-Schioppa, C., and Assad, J.A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature* *441*, 223–226.
- Pan, X., Sawa, K., Tsuda, I., Tsukada, M., and Sakagami, M. (2008). Reward prediction based on stimulus categorization in primate lateral prefrontal cortex. *Nat. Neurosci.* *11*, 703–712.
- Prabhakaran, V., Narayanan, K., Zhao, Z., and Gabrieli, J.D.E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nat. Neurosci.* *3*, 85–90.
- Rizzo, L., Danion, J.-M., van der Linden, M., and Grangé, D. (1996a). Patients with schizophrenia remember that an event has occurred, but not when. *Br. J. Psychiatry* *168*, 427–431.
- Rizzo, L., Danion, J.-M., van der Linden, M., Grangé, D., and Rohmer, J.-G. (1996b). Impairment of memory for spatial context in schizophrenia. *Neuropsychology* *10*, 376–384.
- Roese, N.J., Park, S., Smallman, R., and Gibson, C. (2008). Schizophrenia involves impairment in the activation of intentions by counterfactual thinking. *Schizophr. Res.* *103*, 343–344.
- Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. *Science* *275*, 1593–1599.
- Seo, H., and Lee, D. (2009). Behavioral and neural changes after gains and losses of conditioned reinforcers. *J. Neurosci.* *29*, 3627–3641.
- Seo, H., Barraclough, D.J., and Lee, D. (2007). Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cereb. Cortex* *17*, i110–i117.
- Sohn, J.-W., and Lee, D. (2007). Order-dependent modulation of directional signals in the supplementary and presupplementary motor areas. *J. Neurosci.* *27*, 13655–13666.
- Shima, K., and Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* *282*, 1335–1338.
- Sul, J.H., Kim, H., Huh, N., Lee, D., and Jung, M.W. (2010). Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron* *66*, 449–460.
- Sutton, R.S., and Barto, A.G. (1998). *Reinforcement Learning: An Introduction* (Cambridge, MA: MIT Press).
- Tanji, J., and Hoshi, E. (2001). Behavioral planning in the prefrontal cortex. *Curr. Opin. Neurobiol.* *11*, 164–170.
- Thorndike, E.L. (1911). *Animal Intelligence: Experimental Studies* (New York: Macmillan).
- Tolman, E.C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* *55*, 189–208.
- Tremblay, L., and Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature* *398*, 704–708.
- Tsuchida, A., Doll, B.B., and Fellows, L.K. (2010). Beyond reversal: a critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. *J. Neurosci.* *30*, 16868–16875.
- Tsujimoto, S., and Sawaguchi, T. (2005). Context-dependent representation of response-outcome in monkey prefrontal neurons. *Cereb. Cortex* *15*, 888–898.
- Tsujimoto, S., Genovesio, A., and Wise, S.P. (2009). Monkey orbitofrontal cortex encodes response choices near feedback time. *J. Neurosci.* *29*, 2569–2574.
- Wallis, J.D., and Miller, E.K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* *18*, 2069–2081.
- Walton, M.E., Behrens, T.E.J., Buckley, M.J., Rudebeck, P.H., and Rushworth, M.F.S. (2010). Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* *65*, 927–939.
- Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. *Nature* *382*, 629–632.
- Waters, F.A.V., Maybery, M.T., Badcock, J.C., and Michie, P.T. (2004). Context memory and binding in schizophrenia. *Schizophr. Res.* *68*, 119–125.
- Wheeler, M.A., Stuss, D.T., and Tulving, E. (1997). Toward a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychol. Bull.* *121*, 331–354.