Lateral Intraparietal Cortex and Reinforcement Learning during a Mixed-Strategy Game

Hyojung Seo,^{1*} Dominic J. Barraclough,^{2*} and Daeyeol Lee¹

¹Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06510, and ²Department of Brain and Cognitive Sciences, Center for Visual Science, University of Rochester, Rochester, New York 14627

Activity of the neurons in the lateral intraparietal cortex (LIP) displays a mixture of sensory, motor, and memory signals. Moreover, they often encode signals reflecting the accumulation of sensory evidence that certain eye movements might lead to a desirable outcome. However, when the environment changes dynamically, animals are also required to combine the information about its previously chosen actions and their outcomes appropriately to update continually the desirabilities of alternative actions. Here, we investigated whether LIP neurons encoded signals necessary to update an animal's decision-making strategies adaptively during a computer-simulated matching-pennies game. Using a reinforcement learning algorithm, we estimated the value functions that best predicted the animal's choices on a trial-by-trial basis. We found that, immediately before the animal revealed its choice, ~18% of LIP neurons changed their activity according to the difference in the value functions for the two targets. In addition, a somewhat higher fraction of LIP neurons displayed signals related to the sum of the value functions, which might correspond to the state value function or an average rate of reward used as a reference point. Similar to the neurons in the prefrontal cortex, many LIP neurons also encoded the signals related to the animal's previous choices. Thus, the posterior parietal cortex might be a part of the network that provides the substrate for forming appropriate associations between actions and outcomes.

Introduction

Choosing an action can be considered rational if a set of numerical quantities can be associated with alternative actions such that this quantity is maximized by the chosen action. These hypothetical quantities are referred to as utility functions (von Neumann and Morgenstern, 1944). In reinforcement learning theory, value functions, defined as the estimate of expected rewards, play a similar role, but they are continually adjusted according to reward prediction errors, namely, the discrepancy between actual and expected rewards (Sutton and Barto, 1998). It is also assumed that the decision maker sometimes chooses suboptimal actions for exploration. Because of these additional features, reinforcement learning theory successfully accounts for various choice behaviors in humans and animals (Samejima et al., 2005; Daw et al., 2006), including strategic interactions in social settings (Camerer, 2003; Lee, 2008). Moreover, many neuroimaging and single-neuron recording studies have investigated the neural basis of reinforcement learning (O'Doherty, 2004; Knutson and Cooper, 2005; Daw and Doya, 2006; Lee, 2006; Montague et al., 2006). For example, midbrain dopamine neurons encode the reward prediction errors (Schultz, 1998), which have also been identified in neuroimaging studies (McClure et al., 2003;

O'Doherty et al., 2003; D'Ardenne et al., 2008). Neurons encoding the value functions have been found in many brain areas, including the prefrontal cortex (Watanabe, 1996; Leon and Shadlen, 1999; Barraclough et al., 2004), amygdala (Paton et al., 2006), and basal ganglia (Samejima et al., 2005; Lau and Glimcher, 2008).

Despite the widespread signals related to value functions, their functional roles still remain incompletely understood, in part because neurons carrying signals related to value functions tend to encode multiple variables. Therefore, understanding the relationship among the value-related signals and other variables encoded by the same neuron might provide important insights regarding their origin and functions. For example, neurons in the lateral prefrontal cortex often encode not only the value functions but also the information about the animal's previous choices and their outcomes (Barraclough et al., 2004). Similarly, many neurons in the lateral intraparietal cortex (LIP) encode the value functions for eye movements (Platt and Glimcher, 1999; Dorris and Glimcher, 2004; Sugrue et al., 2004), typically in conjunction with the animal's chosen eye movement. Although neurons in the prefrontal cortex and LIP display similar properties (Chafee and Goldman-Rakic, 1998), the possibility that LIP neurons might also encode the history of the animal's choices and their outcomes has not been tested. For direct comparisons with the activity in the prefrontal cortex (Barraclough et al., 2004; Seo et al., 2007; Seo and Lee, 2008), in the present study, we recorded the activity of LIP neurons while rhesus monkeys played an oculomotor matching-pennies task. This task is well suited for neurobiological studies of decision making, because it reduces the serial correlation among the animal's choices (Lee and Seo,

Received March 28, 2009; revised May 4, 2009; accepted May 6, 2009.

This study was supported by National Institutes of Health Grant MH073246. We thank L. Carr and J. Swanson for their technical assistance and H. Abe, M. W. Jung, and T. J. Vickery for their helpful comments on this manuscript.

*H.S. and D.J.B. contributed equally to this work.

Correspondence should be addressed to Dr. Daeyeol Lee, Department of Neurobiology, Yale University School of Medicine, 333 Cedar Street, Sterling Hall of Medicine B404, New Haven, CT 06510. E-mail: daeyeol.lee@yale.edu.

DOI:10.1523/JNEUROSCI.1479-09.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/297278-12\$15.00/0

Figure 1. Behavioral tasks and performance. **A**, Memory saccade task. **B**, Free-choice task that simulated a matching-pennies game. **C**, Average regression coefficients associated with the choices of the animal (left) and the computer opponent (right) in multiple previous trials. Large symbols indicate that the corresponding values were significantly different from 0 (two-tailed t test, $p < 0.05$). Histograms indicate the fraction of daily sessions in which a given regression coefficient was significantly different from 0 (two-tailed t test, $p < 0.05$). Sacc/Fix, Saccade/fixation.

2007). We found that, similar to the neurons in the lateral prefrontal cortex, LIP neurons encoded signals related to the animal's choice and reward history in addition to the value functions for the alternative actions.

Materials and Methods

Animal preparations. Three rhesus monkeys (H and I, male; K, female; body weight, 5–11 kg) were used. The animal's head was fixed using a halo system that was attached to the skull via a set of four titanium head posts (Lee and Quessy, 2003). Eye movements were monitored at the sampling rate of 225 Hz with a high-speed video-based eye tracker (ET49; Thomas Recording). All the procedures used in this study were approved by the University of Rochester Committee on Animal Research and conformed to the Public Health Service Policy on Humane Care and Use of Laboratory Animals and the *Guide for the Care and Use of Laboratory Animals*.

Behavioral tasks. The results described in this study were obtained from two different oculomotor tasks: a memory saccade task (Fig. 1A) and a free-choice task (Fig. 1B). The present study focuses on the results from the free-choice task, whereas the memory saccade task was used to examine the directional tuning property of each neuron. Trials in both tasks began in the same manner with the animal fixating a small yellow square [$0.9^\circ \times 0.9^\circ$; CIE (Commission Internationale de l'Eclairage, or International Commission on Illumination) coordinates, $x = 0.432$, $y = 0.494$, $Y = 62.9$ cd/m 2] in the center of a computer screen for a 0.5 s fore

period. In both tasks, the animal was also required to maintain its fixation until the central square was extinguished. If the animal broke its fixation prematurely, the trial was aborted immediately without any reward.

During the memory saccade task, a green disk (radius, 0.6° ; CIE coordinates, $x = 0.286$, $y = 0.606$, $Y = 43.2$ cd/m 2) was presented during a 0.5 s cue period in one of eight different locations 5° away from the central fixation target. The target location in each trial was chosen pseudorandomly. When the central fixation target was extinguished after a 1 s memory period, the animal was required to shift its gaze into a circular region (3° radius) around the location previously occupied by the peripheral target within 1 s to receive a drop of juice reward. For each neuron tested in this study, the animal first performed 80 trials of the memory saccade task (10 trials per target) before switching to the free-choice task.

The free-choice task was modeled after a two-player zero-sum game, known as the matching pennies, and has been described previously in detail (Barracough et al., 2004; Lee et al., 2004; Seo and Lee, 2007). During this task, two identical green disks (radius, 0.6° ; CIE coordinates, $x = 0.286$, $y = 0.606$, $Y = 43.2$ cd/m 2) were presented 5° away in diametrically opposed locations along the horizontal meridian. Targets were always presented along the horizontal meridian to make it possible to compare the results from the present study directly with the previous findings in the prefrontal cortex (Barracough et al., 2004) and dorsal anterior cingulate cortex (ACCd) (Seo and Lee, 2007). The central target was extinguished after a 0.5 s delay period, and the animal was required to shift its gaze toward one of the targets within 1 s. After the animal maintained the fixation on its chosen peripheral target for 0.5 s, a red ring (radius, 1° ; CIE coordinates, $x = 0.632$, $y = 0.341$, $Y = 17.6$ cd/m 2) appeared around the target selected by the computer. The animal

was rewarded only if it chose the same target as the computer that simulated a rational decision maker during a matching-pennies game by minimizing the animal's expected payoff (Lee et al., 2004). Accordingly, the animal's optimal strategy during this task, also known as the Nash equilibrium in game theory, was to choose the two targets equally frequently regardless of the animal's previous choices and its outcomes. Any deviations from this optimal strategy could be potentially exploited by the computer opponent. The computer opponent was programmed to exploit statistical biases in the animal's choice behavior by applying a set of statistical tests to the animal's entire choice and reward history during a given recording session (corresponding to algorithm 2 by Lee et al., 2004).

Neurophysiological recording. Single-unit activity was recorded from the neurons in the lateral bank of the intraparietal sulcus of three monkeys (H and I, left hemisphere; K, right hemisphere), using a five-channel multielectrode recording system (Thomas Recording). The placement of the recording chamber was guided by magnetic resonance (MR) images in all animals. For two animals (monkeys I and K), the location of the intraparietal sulcus was further confirmed by MR images obtained with a microelectrode inserted into the brain at known locations inside the recording chamber. All the neurons in area LIP were recorded at least 2.5 mm below the cortical surface down the sulcus (Barash et al., 1991). In the present study, all the neurons localized in LIP were tested for the two tasks described above without prescreening and were included in the analysis if they remained isolated for 80 trials of the memory saccade task and at least 131 trials of the free-choice task.

Analysis of behavioral data. To test whether and how the animal's choice during the free-choice task was influenced by its choices in the previous trials and their outcomes, the following logistic regression model was applied:

$$\text{logit } P_t(R) \equiv \log P_t(R)/(1 - P_t(R)) \\ = a_0 + A_c[c_{t-1} c_{t-2} \dots c_{t-10}] + A_o[o_{t-1} o_{t-2} \dots o_{t-10}], \quad (1)$$

where c_t and o_t indicate the choice of the animal and the choice of the computer opponent in trial t , respectively (1 for rightward choices and -1 otherwise). The choice behavior was also analyzed with a reinforcement learning model (Sutton and Barto, 1998). In this model, the value function for choosing target x_t (R or L for rightward and leftward choices, respectively) in trial t , $Q_t(x_t)$, was updated for the target chosen by the animal according to the reward prediction error, as follows:

$$Q_{t+1}(x_t) = Q_t(x_t) + \alpha[r_t - Q_t(x_t)], \quad (2)$$

where r_t denotes the reward received by the animal in trial t (0 and 1 for unrewarded and rewarded trials, respectively), and α denotes the learning rate. The value function for the unchosen target was not updated. The reward prediction error, $[r_t - Q_t(x_t)]$, corresponds to the discrepancy between the actual reward and the expected reward. The probability that the animal would choose the rightward target in trial t , $P_t(R)$, was determined from these value functions according to the so-called softmax transformation as follows:

$$P_t(R) = 1/[1 + \exp - \beta \{Q_t(R) - Q_t(L)\}], \quad (3)$$

where β , referred to as the inverse temperature, determines the randomness of the animal's choices. Model parameters were estimated separately for each recording session as well as for the entire behavioral dataset obtained from each animal, using a maximum-likelihood procedure (Pawitan, 2001; Lee et al., 2004; Seo and Lee, 2007). Denoting the animal's choice in trial t as c_t ($= R$ or L), the likelihood for the animal's choices in a given session was given by:

$$L = \prod_i P_i(c_i) = P_1(c_1) P_2(c_2) \dots P_N(c_N), \quad (4)$$

where N denotes the number of trials. The maximum-likelihood procedure was implemented using the `fminsearch` function in Matlab 7.0 (MathWorks).

As described in Results, the animals displayed a significant tendency to make their choices according to the so-called win-stay-lose-switch (WSLS) strategy (Lee et al., 2004). In other words, they were more likely to choose the same target rewarded in the previous trial and switched to the other target otherwise. If the animal selected its targets based on a fixed probability of adopting the win-stay-lose-switch strategy, p_{WSLS} , the likelihood that the animal would choose the target x would be $P_t(x) = p_{\text{WSLS}}$, when the computer opponent chose x in the previous trial, and $P_t(x) = (1 - p_{\text{WSLS}})$, otherwise. The maximum-likelihood estimate for p_{WSLS} was obtained by calculating the frequency of the trials in which the animal chose the same target that was chosen by the computer in the previous trial. Whether the animal's choice behavior in a given session was better accounted for by the reinforcement learning model or by the probabilistic win-stay-lose-switch strategy was determined by the Bayesian information criterion (BIC) (Burnham and Anderson, 2002):

$$\text{BIC} = -2 \log L + k \log N, \quad (5)$$

where L is the likelihood, k the number of model parameters (1 for the win-stay-lose-switch strategy model and 2 for the reinforcement learning model), and N the number of trials.

Analysis of neural data. Whether the activity of a given LIP neuron during the memory period of the memory saccade task changed according to the direction of the saccade target was tested by applying the following circular regression model (Georgopoulos et al., 1982):

$$y_t = b_0 + b_1 \sin \theta_t + b_2 \cos \theta_t,$$

where y_t and θ_t denote the spike rate during the memory period and the direction of the saccade target in trial t , respectively, and b_0 through b_2

denote the regression coefficients. A neuron was considered directionally tuned when b_1 or b_2 was significantly different from 0 after correcting for the multiple comparisons (Bonferroni's correction, t test, $p < 0.05$). In this case, the preferred direction θ^* was given by $\text{atan}(b_1/b_2)$.

To test whether the activity of a given LIP neuron was significantly influenced by the value functions of individual targets, we applied the following regression model to the rate of spikes during the 0.5 s delay period, y_t :

$$y_t = b_0 + b_1 c_t + b_2 Q_t(L) + b_3 Q_t(R), \quad (6)$$

where c_t denotes the animal's choice in trial t , $Q_t(x_t)$ is the value function for target x_t , and b_0 through b_3 denote the regression coefficients. The value functions used in this model were obtained by fitting the reinforcement learning described above (Eqs. 2, 3) to the entire behavioral data obtained from each animal. Although this model would be useful in estimating the effect of the value functions for individual targets on neural activity, it is the difference in the value functions for the two targets that ultimately determines the animal's choice in the reinforcement learning model (Eq. 3). Therefore, the regression coefficients in the above regression model provide a limited amount of information about whether the value-related signals in a given LIP neuron can be used for choosing a particular action. Therefore, another regression model was applied to the same data to test whether the activity during the same period was more closely related to the sum of the two value functions or their difference:

$$y_t = b_0 + b_1 c_t + b_s \{Q_t(R) + Q_t(L)\} + b_d \{Q_t(R) - Q_t(L)\}. \quad (7)$$

It should be noted that the regressors of these two models span the same linear space. Therefore, these two models accounted for the variance in the neural activity equally well.

Some previous studies in area LIP have tested whether the activity of LIP neurons was related to the value of the reward expected from a particular choice normalized by the sum of the reward values for both targets (Platt and Glimcher, 1999; Sugrue et al., 2004). It should be noted that the difference in the value functions was highly correlated with the fraction of the value function for one of the targets divided by the sum of the value functions estimated in the present study (but see Corrado et al., 2005). The correlation coefficient between these two quantities averaged across all the sessions was 0.917 ± 0.002 in the present study. Therefore, it was difficult to test whether the LIP activity was more closely related to the difference in the value functions or the fraction of the value function for a particular target. We chose to use the difference in the value functions, because this determines the probability of choosing each of the two alternative actions in the reinforcement learning algorithm used to analyze the behavioral data and also because this quantity was mostly uncorrelated with the sum of the value functions. We also tested whether the activity of LIP neurons might be related to the value function for the target chosen by the animal, using the following regression model:

$$y_t = b_0 + b_1 c_t + b_s \{Q_t(R) + Q_t(L)\} + b_d \{Q_t(R) - Q_t(L)\} \\ + b_c Q_t(c_t). \quad (8)$$

To investigate how the neural activity related to the value functions was influenced by the animal's choice, the choice of the computer opponent, and the reward received by the animal in the previous trial, we added each of these variables separately to the above regression model (Eq. 7). Another regression model in which all of these three variables were included together was also tested.

The standard method to evaluate the statistical significance of a regression coefficient is a t test. However, when the variables in a regression model are serially correlated, this violates the independence assumption regarding the errors and may bias the results of t test (Snedecor and Cochran, 1989). Therefore, the statistical significance of each regression coefficient in the above models was evaluated using a permutation test. For this, we randomly shuffled all the trials in a given session and recalculated the value functions according to the shuffled sequences of the animal's choices and rewards. We then recalculated the regression coefficients for the same regression model. This procedure was repeated 1000

Table 1. The average probability that the animal would choose the rightward target, make its choice according to the WSLS strategy, or choose the target with the maximum value function estimated from a reinforcement learning (RL) algorithm

Monkey	$p(\text{right})$	$p(\text{WSLS})$	$p(\text{RL})$
H	0.498 (24.1%)	0.568 (75.9%)	0.591 (83.8%)
I	0.490 (18.8%)	0.512 (25.0%)	0.553 (90.6%)
K	0.511 (25.9%)	0.556 (66.7%)	0.565 (63.0%)
All	0.499 (22.7%)	0.544 (54.6%)	0.569 (79.6%)

Numbers in the parentheses indicate the percentage of the sessions in which these percentages were significantly higher than 0.5 (binomial test, $p < 0.05$).

times, and the p value for each regression coefficient was given by the frequency of shuffles in which the magnitude of the original regression coefficient was exceeded by the regression coefficients obtained for the shuffled data.

We also analyzed whether the activity of a given neuron was influenced by the animal's choices, the choices of the computer opponent, and the animal's rewards in the current and previous trials, using the following multiple regression model:

$$y_t = B[1 \ u_t \ u_{t-1} \ u_{t-2} \ u_{t-3}]', \quad (9)$$

where u_t is a row vector consisting of three dummy variables corresponding to the animal's choice (0 and 1 for leftward and rightward choices, respectively), the choice of the computer (coded in the same way as the animal's choice), and the reward (0 and 1 for unrewarded and rewarded trials, respectively) in trial t , and B is a vector of 13 regression coefficients.

This analysis was performed separately for the spike rates in a series of nonoverlapping 0.5 s bins defined relative to the time of target onset or feedback onset. Finally, to investigate how the signals related to the choices of the two players and their outcomes are related to the value functions, the sum of the value functions and their difference were added to this regression model as the following:

$$y_t = B[1 \ u_t \ u_{t-1} \ u_{t-2} \ u_{t-3}]' + b_s \{Q_t(R) + Q_t(L)\} + b_d \{Q_t(R) - Q_t(L)\}. \quad (10)$$

For the purpose of illustration, the spike density functions were calculated using a Gaussian filter ($\sigma = 50$ ms) separately for various subset of trials as necessary. Average values shown in Results correspond to mean $\pm$ SEM.

Results

Behavior during the matching-pennies game

During the free-choice task, the animal's choice was rewarded according to the payoff matrix of the matching-pennies game, while the computer opponent analyzed and exploited systematic biases in the animal's choice sequences and their tendencies to choose the targets that were rewarded recently (Barraclough et al., 2004; Lee et al., 2004). Behavioral data obtained from a total of 39,363 trials during the free-choice task in 88 sessions (29, 32, and 27 sessions for monkeys H, I, and K, respectively) were analyzed to investigate the animal's decision-making strategies. For the free-choice task used in the present study, the optimal strategy was to choose the two targets equally frequently and independently across successive trials. Indeed, averaged across all sessions, the overall frequency of choosing the rightward target was 49.9%, and this was not significantly different from 50% (t test, $p = 0.768$). In addition, for the majority of recording sessions (68 of 88 sessions, 77.3%), the null hypothesis that the two targets were chosen equally frequently could not be rejected (binomial test, $p > 0.05$) (Table 1). However, the animals tended to choose the same target chosen by the computer in the previous trial. In other words, the animal tended to choose the same target again after a rewarded trial ("win-stay") and to choose the other target

otherwise ("lose-switch"). On average, the probability that the animal would choose its target according to the win-stay-lose-switch strategy was 0.544 ± 0.005 , and this was significantly larger than 0.5 (t test, $p < 10^{-14}$). In addition, the null hypothesis that the animal chose its target according to this so-called win-stay-lose-switch strategy with the probability of 0.5 was rejected in more than half of the sessions (54.6%) (Table 1). The bias to use the win-stay-lose-switch strategy was also reflected in the positive regression coefficient associated with the computer's choice in the previous trial in a logistic regression analysis (Eq. 1). This analysis also revealed that the tendency to choose the same target chosen by the computer opponent was significant for the previous trials up to seven trials before the current trial (Fig. 1C). In contrast, the animals showed relatively weak or little biases to choose the same target as in the previous trials (Fig. 1C).

We also found that a simple reinforcement learning algorithm (Sutton and Barto, 1998) accounted for the animal's choices during the free-choice task better than the model with a fixed probability of WSLS strategy. On average, the probability that the animal's choice was correctly predicted by the value functions in the reinforcement learning model was 0.569 ± 0.004 (Table 1), and this was significantly higher than the probability that the animal's choice was predicted by the win-stay-lose-switch strategy (t test, $p < 10^{-7}$). In addition, the value of BIC was smaller for the reinforcement learning model than the WSLS-strategy model in 68 of 88 sessions (77.3%). The average learning rate in the reinforcement learning model was 0.468, 0.295, and 0.633 for monkeys H, I, and K, respectively, and therefore revealed some individual variability (Fig. 2A). Nevertheless, the fact that the reinforcement learning model performed better than the WSLS-strategy model indicates that, consistent with the results from the logistic regression model (Fig. 1C), the animal's choices were influenced by the outcomes of the animal's choices in multiple trials. These two models are still related, however, because the reinforcement learning model with the learning rate of 1 is equivalent to the WSLS-strategy model. Indeed, across different sessions, there was a significant positive correlation between the learning rate in the reinforcement learning model and the probability that the animal's choice was determined by the WSLS strategy (Fig. 2B), indicating that the WSLS strategy arises as a special case from the same adaptive process described by the reinforcement learning model.

LIP signals related to values and choices

A total of 198 neurons were recorded from the lateral bank of the intraparietal sulcus in three rhesus monkeys and tested during both the memory saccade and free-choice tasks. All well isolated neurons were tested and included in the analysis. For the free-choice task, each neuron was tested for 131 to 846 trials and on average for 441.7 ± 14.3 trials. The results from the circular regression analysis showed that the activity of 85 neurons (42.9%) showed significant directional tuning during the 1 s memory period of the memory saccade task (supplemental Figs. 1, 2, available at www.jneurosci.org as supplemental material). For 53 of these neurons, the preferred directions were within 45° of the horizontal meridian, and they are referred to as horizontally tuned neurons below. To test whether the activity of LIP neurons reflected the value functions of alternative choices during the free-choice task, we applied a multiple linear regression model that included the value functions estimated from the animal's choices. To distinguish between the signals related to the value functions and the animal's upcoming choice, this model also included a dummy variable corresponding to the animal's choice.

The results from this analysis showed that a substantial number of neurons in area LIP significantly modulate their activity during the delay period of free-choice trials according to the value functions of one or both targets. For example, the activity of the neuron illustrated in Figure 3A increased its activity with the value function for the leftward target, whereas the effect of the value function for the rightward target was not statistically significant. In contrast, the neurons shown in Figure 3, B and C, significantly increased their activity with the value functions of both targets.

To estimate how often the activity of LIP neurons was influenced by the value functions for at least one of the two targets, we used the Bonferroni's correction to account for the fact that the significance test was applied separately for the two targets. The overall percentage of neurons that showed significant effects of value functions was 22.7% (permutation test, $p < 0.05$). Among the horizontally tuned neurons, the percentage of neurons that significantly modulated their activity according to the value functions was somewhat higher (28.3%). However, this was not significantly higher compared with the proportion of such neurons in the entire population (z test, $p > 0.1$). Therefore, whether the targets were presented along the axis of the preferred direction of the neuron or not did not strongly influence the LIP activity related to the value functions. Next, we constructed a 2×2 contingency table by counting the neurons that significantly modulated their activity according to the value function of each target and used a χ^2 test to test whether the activity of a given LIP neuron tends to be influenced by the value functions of the two targets independently or not. The percentage of neurons that modulated their activity according to the value functions of both targets was 6.6%, which was significantly higher than expected for independent effects (χ^2 test, $p < 0.05$). Therefore, the activity of individual LIP neurons tends to be affected by the value functions for both targets.

Although the results described above indicate that the activity of LIP neurons is often modulated by the value functions of alternative actions, they do not directly address the question of whether such activity can contribute to the process of choosing one of the two targets. This is because, in the reinforcement learning model, the choice between the two alternative targets is determined by the difference in the value functions of the two targets rather than by the value function of either target alone. Therefore, the fact that the activity of a given LIP neuron reflects the value function for one of the targets does not necessarily indicate that the same neuron contributes to the process of choosing a particular target based on the difference in the value functions. For example, the activity of neurons shown in Figure 3, B and C, was not systematically related to the difference in the value functions of the two targets, because the activity of these neurons increased similarly with the value functions of both targets. To investigate how the neural activity related to the value functions might be used for the purpose of decision making, we applied a multiple regression model that included the sum of the value functions and their difference in addition to the animal's upcoming choice. This analysis showed that 17.7% of LIP neurons significantly modulated their activity according to the difference in the value functions (Figs. 3A, 4). In addition, 21.7% of

Figure 2. Parameters of reinforcement learning model. **A**, Inverse temperature (β) plotted against the learning rate (α) estimated from the same testing session. **B**, The probability that the animal would use the win–stay–lose–switch strategy is plotted against the learning rate (α).

the neurons displayed significant modulations in their activity related to the sum of the value functions (Fig. 4). Whether the activity of a given neuron was affected by the difference of value function did not influence the likelihood that the activity of the same neuron would be affected by the sum of the value function (χ^2 test, $p > 0.05$). In other words, the sum and difference of value function influenced the activity of each LIP neuron independently.

We also tested whether the directional tuning of LIP neurons during the memory period of the memory saccade and activity related to the value functions might be systematically related. To this end, we examined the relationship between the standardized regression coefficient related to the horizontal position of the target in the memory saccade task and the standardized regression coefficient related to the difference in the value functions during the free-choice task. We found that these two coefficients were significantly and positively correlated ($r = 0.189$; $p < 0.01$) (Fig. 5), suggesting that the LIP activity related to the value functions during the free-choice task can be read out in the same coordinate system used to decode the directionally tuned activity during the memory period of the memory saccade task. Nevertheless, this relationship was relatively weak, and some LIP neurons did not modulate their activity consistently according to the remembered position of the target during the memory period and the difference in the value functions during the free-choice task (Fig. 5).

Not surprisingly, the neurons that displayed changes in their activity according to the value functions of individual targets were more likely to contribute signals related to the sum of the value functions or their differences compared with the neurons that did not show any significant effects of value functions for either target. We first selected all the neurons showing significant modulations in their activity related to the value functions of one or both targets without using the Bonferroni's correction, because the goal of this analysis was not to determine the percentage of LIP neurons with significant value-related signals but to characterize the general nature of value-related signals in LIP. Among such neurons ($n = 62$), 46.8 and 58.1% of LIP neurons significantly modulated their activity according to the difference and sum of the value functions, respectively (Fig. 4). Similar results were obtained and the percentages were even higher when the Bonferroni's correction was applied (48.9 and 66.7% among 45

Figure 3. Three example LIP neurons showing significant changes in their activity according to value functions (A–C). For each neuron, the leftmost column shows the difference between the value functions for the two targets (black) and the fraction of trials in which the animal chose the rightward target estimated by the moving average of 10 successive trials. The next two columns show the average spike rates during the delay period of the free-choice task for each decile of trials sorted by the value function for the leftward or rightward target. This was computed separately according to the target chosen by the animal (open circles, leftward target; filled circles, rightward target). The last two columns show the activity of the same neuron for the sum of the value functions and their difference in the same format. Light and dark gray histograms show the distribution of trials in which the animal chose the leftward and rightward targets, respectively.

Figure 4. Effect of value functions on LIP activity. Raw (left) and standardized (right) regression coefficients associated with the value functions for leftward (abscissa) and rightward (ordinate) targets. Circles correspond to the neurons in which the effect of the value function was significant for at least one target, whereas squares indicate the neurons in which the effect of value function was not significant for either target. Green and blue symbols indicate the neurons in which the activity was significantly influenced by either the sum of the value functions or their difference, respectively, whereas the red symbols indicate the neurons in which the activity was significantly influenced by both. Empty symbols indicate the neurons in which neither variable had a significant effect.

neurons). In contrast to the signals related to the sum and difference of value functions, we found that the value function for the target chosen by the animal in a given trial influenced the activity of LIP neurons relatively infrequently. The percentage of neurons showing the significant effect of value function for the chosen target was relatively low during the delay period (8.6%), although this was significantly higher than the 5% significance level (binomial test, $p < 0.05$). The percentage of such neurons increased

somewhat to 11.1% during the 0.5 s fixation period before feedback onset, but this difference was not statistically significant (χ^2 test, $p > 0.4$).

LIP signals related to choices and outcomes

To test whether the activity of LIP neurons was influenced by the choices of the animal and the computer opponent in the preceding trials and the outcomes of the animal's previous choices, the firing rate of each neuron was analyzed with a multiple regression model including these variables in the current and three preceding trials (Eq. 9). To examine the time course of such neural signals, this analysis was performed for a series of 0.5 s windows aligned at the time of target or feedback onset. The results from this analysis showed that LIP neurons often modulated their activity according to multiple variables. However, which of these multiple variables influenced the activity of a given

neuron and their time courses varied substantially across the population of LIP neurons. For example, the neuron illustrated in Figure 3A was tuned along the horizontal axis and tended to increase its activity with the value function of the leftward target. The activity of the same neuron during the delay period also increased significantly when the animal chose the leftward target in the same trial compared with when the animal chose the right-

Figure 5. Relationship between the spatial tuning during the memory period of the memory saccade task and the activity related to the difference in the value functions. For each LIP neuron, standardized regression coefficient for the difference in the value functions (ordinate) is plotted against standardized regression coefficient for the horizontal target position during the memory saccade task. These coefficients were estimated from the delay period of the free-choice task and the memory period of the memory saccade task, respectively. Filled symbols indicate the neurons that are horizontally tuned, whereas green (red) symbols indicate the neurons in which the coefficient for the difference in the value functions, $Q_t(R) - Q_t(L)$, was significantly positive (negative).

ward target (Figs. 3A; 6, top, Trial Lag = 0). In addition, the same neuron significantly increased its activity during the delay period when the animal had chosen the rightward target in the previous trial (Fig. 6, top, Trial Lag = 1). Although the activity of this neuron during the delay period was not significantly affected by the choice of the computer opponent in the previous trial (Fig. 6, middle, Trial Lag = 1), its activity increased significantly when the computer opponent had chosen the leftward target two or three trials before the current trial compared with when the rightward target had been chosen by the computer opponent (Fig. 6, middle, Trial Lag = 2 and 3). Finally, immediately after the feedback period, this neuron increased its activity immediately when the animal was rewarded (Fig. 6, bottom, Trial Lag = 0) but decreased its activity when the animal had been rewarded in the previous trial (Fig. 6, bottom, Trial Lag = 1). During the delay period, the activity of the same neuron was higher if the animal was rewarded in the previous trial. In summary, the activity of this neuron during the delay period was influenced by the animal's choice and its outcome as well as the choice of the computer opponent in multiple trials.

Two other example neurons in which the activity was significantly influenced by the outcome of the animal's choice in the previous trial are illustrated in Figures 7 and 8. One of these neurons was vertically tuned during the memory period of the memory saccade task with its preferred direction in the upward direction (Fig. 7). The activity of this neuron was not significantly modulated by the animal's choice in the same trial during either the delay or feedback period, whereas its activity during the delay period increased slightly but significantly when the animal's choice in the previous trial was the leftward target (Fig. 7, top, Trial Lag = 1). The activity of the same neuron showed a robust increase during the feedback period of unrewarded trials (Fig. 7, bottom, Trial Lag = 0). Its activity, however, increased significantly during the delay and feedback periods, when the animal was rewarded in the previous trial (Fig. 7, bottom, Trial Lag = 1). Therefore, the activity of this neuron during the feedback period was influenced oppositely by the outcomes of the animal's choices in the current and previous trial. The neuron illustrated

in Figure 8 did not show a significant directional tuning during the memory period of the memory saccade task. However, its activity was slightly but significantly higher during the delay period when the animal would choose the rightward target in the same trial (Fig. 8, top, Trial Lag = 0). After the fixation offset, the activity of this neuron was significantly enhanced when the animal chose the leftward target. In addition, the activity of the same neuron increased significantly during the feedback period when the choice of the computer was the leftward target (Fig. 8, middle, Trial Lag = 0). This neuron also displayed a higher level of activity during the feedback period of rewarded trials than in unrewarded trials (Fig. 8, bottom, Trial Lag = 0), and this difference was maintained through the delay and feedback period of the next trial (Fig. 8, bottom, Trial Lag = 1). Thus, in contrast to the neuron shown in Figure 7, the activity of this neuron was influenced consistently by the positive outcomes of the animal's choices in the current and previous trials.

To investigate the time course of signals related to these multiple variables at the population level, we calculated the percentage of neurons that showed statistically significant modulations in their activity according to each of these variables (Fig. 9A). Overall, 12.6 and 25.8% of the neurons significantly changed their activity during the fore period and delay period, respectively, according to the animal's upcoming choice in the same trial (Fig. 9A, top, Trial Lag = 0), whereas the corresponding percentage during the 0.5 s fixation period immediately before feedback onset was 82.3%. Therefore, for a large majority of LIP neurons tested in the present study, postsaccadic activity was significantly influenced by the direction of the eye movement, although they were not prescreened for any saccade-related activity. During the feedback period, many LIP neurons significantly changed their activity not only according to the animal's choice (69.7%) but also according to the choice of the computer opponent (53.5%) and the outcome of the animal's choice (84.3%).

Activity of many LIP neurons during the delay period and feedback period also reflected the choices of the animal and computer opponent in the previous trial and its outcome (Fig. 9A, Trial Lag = 1). Overall, 29.3 and 29.8% of the neurons displayed significant modulations in their activity during the delay period according to the animal's choice (Fig. 9A, top, Trial Lag = 1) and its outcome (Fig. 9A, bottom, Trial Lag = 1) in the previous trial, whereas a smaller but still statistically significant number of neurons changed their activity according to the choice of the computer opponent in the previous trial (11.6%) (Fig. 9A, middle, Trial Lag = 1). The percentages of neurons that changed their activity significantly according to the animal's choice, the choice of the computer, and choice outcome two trials before the current trial were 8.6, 9.1, and 8.6%, respectively, and all were still significantly higher than the 5% significance level (binomial test, $p < 0.05$). Activity during the feedback period was also significantly influenced by the animal's choice and its outcome in the previous trial for 14.1 and 23.7% of the neurons, respectively. The percentage of neurons that modulated their activity during the feedback period according to the choice of the computer in the previous trial was 7.6%. Although this was significantly lower than the percentages for the animal's choice in the previous trial (χ^2 test, $p < 0.05$), it was still significantly higher than the 5% significance level.

In the present study, the two alternative targets were always presented along the horizontal meridian during the free-choice task, regardless of the directional tuning properties of individual neurons during the memory saccade task. Therefore, we tested

Figure 6. Activity of the same neuron illustrated in Figure 3A that showed significant modulation in their activity during the delay period according to the animal's choice and reward in the previous trial. A pair of spike density functions in each panel indicates the average activity sorted by the animal's choice (top), the choice of the computer opponent (middle), or the reward received by the animal (bottom) in the current (Trial Lag = 0) or previous three (Trial Lag = 1 to 3) trials. The activity during the trials in which the rightward (leftward) target was chosen or in which the animal was rewarded (unrewarded) is indicated by the green (black) line. Red symbols indicate the regression coefficient associated with each variable during each 0.5 s window used in the regression analysis. Large symbols indicate that the regression coefficient was significantly different from 0 (t test, $p < 0.05$). Gray background corresponds to the delay period (left columns) or feedback period (right columns).

Figure 7. Activity of an LIP neuron that showed vertically tuned activity during the memory period of the memory saccade trials. Same format as in Figure 6.

whether the activity of LIP neurons was affected differently by the animal's current and previous choices and their outcomes depending on whether the targets were presented along the axis of the preferred direction of the neuron or not. For this analysis, neurons were divided according to whether their activity during the memory period of the memory saccade task was significantly tuned and, if so, whether their preferred directions were closer to the horizontal or vertical meridian. We found that the horizontally tuned neurons were more likely to modulate their activity

during the delay period according to the animal's upcoming choice. The percentage of such neurons was 41.5% for the horizontally tuned neurons, whereas the corresponding percentages were 20.4 and 18.8% for the untuned and vertically tuned neurons. This difference among three groups of neurons was statistically significant (χ^2 test, $p < 0.01$). During the feedback period, directionally tuned neurons in LIP were more likely to change their activity according to the choice of the computer opponent (χ^2 test, $p < 0.005$). The percentage of the untuned neurons with significant effects of the choice of the computer during the feedback period was 43.4%, whereas the corresponding percentages for the horizontally and vertically tuned neurons were 66.6 and 68.8%, respectively. Overall, however, the directional tuning properties had a relatively minor influence on the frequency and time course of the signals related to the previous choices of the two players and their outcomes (Fig. 9B). For example, during the delay period, the percentages of neurons showing significant modulations in their activity related to the animal's choice in the previous trial were 41.5, 25.0, and 24.8% for the horizontally tuned, vertically tuned, and untuned neurons, respectively. This difference among the three groups of neurons was not statistically significant (χ^2 test, $p > 0.05$). Similarly, the percentage of neurons that modulated their activity according to the choice of the computer opponent or the outcome of the animal's choice in the previous trial did not differ among these three groups of neurons (Fig. 9B).

In summary, the neural signals related to the animal's choice built up in area LIP during the fore period and delay period and decayed gradually in the next two trials. The signals related to the choice of the computer opponent and choice outcome that emerged during the feedback period also dissipated gradually in the next two trials. Similarly, the magnitude of regression coefficients related to the same variables decayed gradually (supplemental Fig. 3, available at www.jneurosci.org as supplemental material). It should be noted that, during the feedback period, the outcome of the animal's choice, namely, whether the animal would be rewarded or not at the end of the feedback period, was indicated by the presence or absence of a red feedback ring presented around the peripheral target chosen by the animal. Therefore, the outcome-related activity observed during the feedback period might reflect the visual responses of LIP neurons. Regardless of its origin, however, the outcome-related activity in area LIP was sustained for multiple trials, unrelated to the animal's choice, and was primarily independent of the directional tuning properties of

LIP neurons. Therefore, it is unlikely that the outcome-related activity in area LIP was entirely visually driven or attributable to the processes related to saccade preparations. In addition, signals related to the animal's upcoming choice were more likely to emerge in the neurons tuned along the horizontal axis, whereas the signals related to the choice of the computer opponent emerged more frequently for the directionally tuned neurons regardless of their preferred directions.

Decomposition of value-related signals in LIP

During the matching-pennies game used in this study, the animal's choice was rewarded when it matched the choice of the opponent. Therefore, in the reinforcement learning model used to analyze the animal's choice behaviors, the value function for a given target would increase with the frequency in which the same target was chosen by the computer opponent. Similarly, the difference in the value functions for the two targets would correspond approximately to the relative frequency in which each of the two targets was chosen by the computer opponent. Moreover, the value function for the chosen target increases or decreases depending on whether the animal is rewarded or not, whereas the value function for the unchosen target remains unchanged. Therefore, the sum of the value functions for the two targets would be primarily determined by the frequency of rewarded trials in the recent past. This suggests that the changes in LIP activity related to the computer's choices and the outcomes of the animal's choices in the previous trials might provide a substrate for computing the value functions and their linear combinations, such as their sum and difference. To test this possibility, we added the sum and difference of the value functions to the multiple linear regression model that includes the choices of the two players and their outcomes in the previous trials (Eq. 10). Compared with the results obtained from the regression model without the value functions, the percentage of neurons that significantly changed their activity according to the animal's choices in the previous trials did not change significantly (Fig. 9A, top, gray symbols), suggesting that activity changes in LIP resulting from the animal's previous choices were not related to the value functions. In contrast, the percentage of neurons that significantly changed their activity during the fore period according to the choice of the computer opponent in the previous trial decreased significantly when the value functions were included in the model. During the fore period, the percentage of neurons that significantly changed their activity according to the choice of the computer opponent in the previous trial was 17.7% when the model did not include the value functions, and this decreased to 8.6% when the value functions were included in the regression model (χ^2 test, $p < 0.005$) (Fig. 9A, middle). The percentage of such neurons also decreased from 11.6 to 9.1% during the delay period, but this difference was not significantly significant (χ^2 test, $p > 0.4$). The percentage of neurons that significantly changed their activity during the fore period or delay period according to the reward in the previous trial did not change significantly when the value functions were included in the regression model, although the difference was significant during the 0.5 s

Figure 8. Activity of an LIP neuron that was not directionally tuned during the memory period of the memory saccade trials. Same format as in Figure 6.

window immediately before the fore period and before the feedback period (Fig. 9A, bottom).

We also tested how the neural signals in LIP related to the choices of the two players in the previous trial and their outcomes might manifest as the sum and difference of value functions. When the animal's choice or the choice of the computer opponent was added individually to the regression model, the percentage of neurons that significantly changed their activity according to the sum of value functions was essentially unaffected (Fig. 10). In contrast, when the reward in the previous trial was added to the model, the percentage of such neurons decreased significantly from 21.7 to 10.1%. The inclusion of the animal's choice and the choice of the computer opponent in addition to the reward in the previous trial did not further reduce the percentage of such neurons (Fig. 10). Therefore, the reward in the previous trial was an important factor contributing to the signals related to the sum of value functions in LIP. When we added the animal's choice, the choice of the computer opponent, or the outcome of the previous trial individually to the regression model with the value functions, the percentage of neurons that significantly modulated their activity according to the difference of value functions did not change significantly (Fig. 10). In contrast, when all three variables were simultaneously added to the regression model, the percentage of neurons significantly decreased from 17.7 to 9.1% (χ^2 test, $p < 0.005$). Nevertheless, the percentage of neurons that showed significant changes related to the difference of value functions was still significantly higher than the 5% significance level, indicating that the signals related to the difference of value functions in area LIP did not entirely arise from the signals related to the choices of the two players and their outcome in the most recent trial.

Discussion

LIP encoding of value functions during a mixed-strategy game

During the matching-pennies task, the animals are required to choose the two targets equally frequently and independently across successive trials (Lee et al., 2004). Such stochastic behaviors are advantageous for identifying different types of neural

Figure 9. Population summary of LIP activity related to choices and outcomes. **A**, Fraction of LIP neurons that showed significant changes in their activity according to the animal's choice (top), the choice of the computer opponent (middle), and the reward in the current (Trial Lag = 0) or previous (Trial Lag = 1 to 3) trials. The statistical significance was determined using a linear regression model applied to the spike rates in a series of nonoverlapping 0.5 s window aligned to target onset (left columns) or feedback onset (right columns). Gray symbols show the results obtained from the regression model that also included the value functions, and the asterisks indicate that they are significantly different from the results obtained from the regression model without the value functions (black symbols; χ^2 test, $p < 0.05$). **B**, Same results shown in **A** from the regression model without the value functions, sorted separately according to the tuning property of each neuron. Asterisks indicate that the values among the horizontally tuned (green), vertically tuned (red), and untuned (blue) neurons were significantly different (χ^2 test, $p < 0.05$). Gray background corresponds to the delay period (left columns) or feedback period (right columns).

signals involved in reinforcement learning, such as the history of animal's previous choices and their outcomes (Lee and Seo, 2007; Seo and Lee, 2008), making competitive games an attractive paradigm for neurobiological studies of decision making (Barraclough et al., 2004; Dorris and Glimcher, 2004; Cui and Andersen, 2007; Seo and Lee, 2007, 2009; Thevarajah et al., 2009; Vickery and Jiang, 2009). Moreover, in both humans and animals, reinforcement learning algorithms can account for system-

atic deviations in the choice behaviors during the matching-pennies game, suggesting that the feedback from the previous choices steer decision makers toward the optimal strategies during competitive interactions (Mookherjee and Sopher, 1994; Erev and Roth, 1998; Lee et al., 2004; Soltani et al., 2006; Lee, 2008). Previous studies have suggested that the LIP plays an important role in adaptive decision making by accumulating the sensory evidence (Gold and Shadlen, 2007; Yang and Shadlen, 2007) or integrating the information about the previous reward history. For example, when the reward probability is unpredictably altered and thus needs to be estimated from the animal's experience, LIP activity tends to track the probability of reward from the movements toward their receptive fields normalized by the overall reward probability (Sugrue et al., 2004). In the formulation of the matching law (Herrnstein, 1997), this so-called fractional income determines the probability that a particular action would be chosen. When the reward probabilities change frequently, the fraction income can be computed locally within a limited temporal window (Sugrue et al., 2004; Kennerley et al., 2006), and the resulting estimates of incomes would resemble the value functions of the reinforcement learning theory (Sutton and Barto, 1998). Indeed, during the so-called inspection game, the activity of LIP neurons was closely related to the value function of the target in the receptive field that was estimated using a reinforcement learning algorithm (Dorris and Glimcher, 2004).

Consistent with these previous findings, we found in the present study that LIP activity was modulated by the difference in the value functions for the two alternative targets. These might be used directly to allow the animal to choose an optimal action. In the present study, the difference in the value functions and the quotient of the value function for a given target divided by the sum of the value functions were highly correlated. Therefore, it is possible that the activity seemingly related to the difference in the value functions might reflect the value function for a particular target normalized by the sum of the value functions (Dorris and

Glimcher, 2004; Sugrue et al., 2004). We also found that individual LIP neurons displayed a substantial degree of heterogeneity in the extent to which their activity was influenced by the value functions of individual targets or their linear transformations. In particular, many LIP neurons modulated their activity according to the sum of the value functions rather than their difference. In reinforcement learning theory (Sutton and Barto, 1998), the average of the action value functions weighted by the probabilities

Figure 10. Fraction of LIP neurons that significantly changed their activity according to the sum of the value functions (black) or their difference (gray). Base, The results from the regression model that only included the animal's choice and the value functions in the same trial; +M, +C, +R, the results obtained from the regression model that also include the choice of the animal (computer's choice, reward) in the previous trial; +All, the results from the model including all three variables. Solid and dotted horizontal lines correspond to the 5% significance level and the minimum value significantly higher than 5%.

of corresponding actions is referred to as the state value function. Given that the two targets were chosen almost equally frequently during the matching-pennies task, the LIP activity related to the sum of the value functions might encode the state value function (Belova et al., 2008). The state value functions would provide the information about the average rate of reward expected when the animal maintains its current decision-making strategy. This might be used as a reference point for evaluating the desirability of the outcome from a given action (Helson, 1948; Kahneman and Tversky, 1979; Frederick and Loewenstein, 1999; Seo and Lee, 2007). We also found that LIP neurons tend to modulate their activity according to the value functions for both targets more frequently than expected when such effects were combined independently. In contrast, the sum of value functions and their difference influenced the activity of different LIP neurons independently, raising the possibility that the signals related to these two variables might be carried separately by the inputs to the LIP.

LIP signals related to choices and outcomes

Consistent with the previous studies in area LIP (Platt and Glimcher, 1999; Coe et al., 2002; Dorris and Glimcher, 2004; Sugrue et al., 2004; Cui and Andersen, 2007), many neurons examined in the present study started to change their activity during the delay period according to the direction of the upcoming eye movement chosen by the animal. Moreover, neurons tuned horizontally during the memory saccade task were more likely to show such choice-related preparatory activity. Such choice-related activity peaked during the fixation period before the feedback onset and subsided gradually during the subsequent intertrial interval and throughout the next trial. In contrast to the preparatory signals related to the animal's upcoming choices, the frequency of signals related to the animal's previous choice was not significantly affected by the tuning properties of LIP neurons, suggesting that the signals related to the previous choice might be less sharply tuned. The signals related to the previous actions of the animal might provide the memory trace, known as the eligibility trace (Sutton and Barto, 1998), which might be used to form appropriate associations between a particular action and reward. We also found that many LIP neurons changed their activity differ-

ently depending on the feedback signal about the outcome of the animal's choice. The possibility that at least the initial component of this activity might be a pure visual signal cannot be excluded completely. However, similar to the signals related to the animal's previous choice, this outcome-related activity also diminished gradually over the course of two or three trials, suggesting that LIP neurons might encode the information about the animal's reward history. Moreover, many LIP neurons changed their activity differently depending on the choice of the computer opponent. Not surprisingly, when the animal's reward in the previous trial and the value functions were included in the same regression model, the proportion of the LIP neurons that showed significant effects for each variable was reduced but still higher than expected by chance, suggesting that the activity of LIP neurons related to the value functions was based on the animal's reward history in multiple trials.

Relation to the signals related to values and choices in other brain areas

Many of the signals identified in the present study for the LIP have been found in other brain areas, suggesting that the computation of value functions based on the animal's previous actions and their outcomes might take place in multiple regions of the brain. In particular, during the matching-pennies task, neurons in the dorsolateral prefrontal cortex (DLPFC) carry the signals related to the value functions, the current and previous choices of the animal and its opponent, as well as the animal's rewards in previous trials (Barraclough et al., 2004; Seo et al., 2007; Seo and Lee, 2008). The proportion of neurons with value-related activity as well as the time course of signals related to the animal's choice and reward history was similar for DLPFC and LIP (supplemental Figs. 4, 5). In contrast, the neurons in the ACCd primarily encoded the sum of the value functions and the animal's reward history (Seo and Lee, 2007) (supplemental Figs. 4, 5, available at www.jneurosci.org as supplemental material). Signals related to the animal's previous choice has been also found in the striatum (Kim et al., 2007; Lau and Glimcher, 2007), suggesting that the eligibility trace for chosen actions might be widespread in the brain. Similarly, signals related to past reward might be present in many areas, including the orbitofrontal cortex (Simmons and Richmond, 2008) and posterior cingulate cortex (Hayden et al., 2008). Neurons with signals related to value functions have been found in multiple areas in the frontal cortex, including the DLPFC (Watanabe, 1996; Leon and Shadlen, 1999; Kobayashi et al., 2002; Kim et al., 2008, 2009), orbitofrontal cortex (Padoa-Schioppa and Assad, 2006), and ACCd (Shidara and Richmond, 2002; Amiez et al., 2006). Similarly, neurons in the striatum often change their activity according to the value functions for specific actions (Samejima et al., 2005) as well as the value of the action chosen by the animal during a free-choice task (Lau and Glimcher, 2008). Signals related to the value of the chosen outcome have been also identified in the orbitofrontal cortex (Padoa-Schioppa and Assad, 2006). In the present study, we found that LIP neurons seldom encoded the value function for the chosen action. Nevertheless, the anatomical distribution of these multiple types of value functions and their time courses need to be examined more closely in the future studies. Such studies are likely to provide important insights into how the information about the previous actions and their outcomes can be used to control the animal's future actions appropriately.

References

- Amiez C, Joseph JP, Procyk E (2006) Reward encoding in the monkey anterior cingulate cortex. *Cereb Cortex* 16:1040–1055.

- Barash S, Bracewell RM, Fogassi L, Gnadt JW, Andersen RA (1991) Saccade-related activity in the lateral intraparietal area. I. Temporal properties; comparison with area 7a. *J Neurophysiol* 66:1095–1108.
- Barracough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. *Nat Neurosci* 7:404–410.
- Belova MA, Paton JJ, Salzman CD (2008) Moment-to-moment tracking of state value in the amygdala. *J Neurosci* 28:10023–10030.
- Burnham KP, Anderson DR (2002) Model selection and multimodel inference. A practical information-theoretic approach, Ed 2. New York: Springer.
- Camerer CF (2003) Behavioral game theory. Princeton: Princeton UP.
- Chafee MV, Goldman-Rakic PS (1998) Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J Neurophysiol* 79:2919–2940.
- Coe B, Tomihara K, Matsuzawa M, Hikosaka O (2002) Visual and anticipatory bias in three cortical eye fields of the monkey during an adaptive decision-making task. *J Neurosci* 22:5081–5090.
- Corrado GS, Sugrue LP, Seung HS, Newsome WT (2005) Linear-nonlinear-poisson models of primate choice dynamics. *J Exp Anal Behav* 84:581–617.
- Cui H, Andersen RA (2007) Posterior parietal cortex encodes autonomously selected motor plans. *Neuron* 56:552–559.
- D'Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* 319:1264–1267.
- Daw ND, Doya K (2006) The computational neurobiology of learning and reward. *Curr Opin Neurobiol* 16:199–204.
- Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441:876–879.
- Dorris MC, Glimcher PW (2004) Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44:365–378.
- Erev I, Roth AE (1998) Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am Econ Rev* 88:848–881.
- Frederick S, Loewenstein G (1999) Hedonic adaptation. In: *Well-being: the foundations of hedonic psychology* (Kahneman D, Diener E, Schwartz N, eds), pp 302–329. New York: Russell Sage Foundation.
- Georgopoulos AP, Kalaska JF, Caminiti R, Massey JT (1982) On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J Neurosci* 2:1527–1537.
- Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535–574.
- Hayden BY, Nair AC, McCoy AN, Platt ML (2008) Posterior cingulate cortex mediates outcome-contingent allocation of behavior. *Neuron* 60:19–25.
- Helson H (1948) Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol Rev* 55:297–313.
- Herrnstein RJ (1997) The matching law: papers in psychology and economics. Cambridge, MA: Harvard UP.
- Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–291.
- Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. *Nat Neurosci* 9:940–947.
- Kim S, Hwang J, Lee D (2008) Prefrontal coding of temporally discounted values during intertemporal choice. *Neuron* 59:161–172.
- Kim S, Hwang J, Seo H, Lee D (2009) Valuation of uncertain and delayed rewards in primate prefrontal cortex. *Neural Netw* 22:294–304.
- Kim YB, Huh N, Lee H, Baeg EH, Lee D, Jung MW (2007) Encoding of action history in the rat ventral striatum. *J Neurophysiol* 98:3548–3556.
- Knutson B, Cooper JC (2005) Functional magnetic resonance imaging of reward prediction. *Curr Opin Neurol* 18:411–417.
- Kobayashi S, Lauwereyns J, Koizumi M, Sakagami M, Hikosaka O (2002) Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex. *J Neurophysiol* 87:1488–1498.
- Lau B, Glimcher PW (2007) Action and outcome encoding in the primate caudate nucleus. *J Neurosci* 27:14502–14514.
- Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* 58:451–463.
- Lee D (2006) Neural basis of quasi-rational decision making. *Curr Opin Neurobiol* 16:191–198.
- Lee D (2008) Game theory and neural basis of social decision making. *Nat Neurosci* 11:404–409.
- Lee D, Quessey S (2003) Activity in the supplementary motor area related to learning and performance during a sequential visuomotor task. *J Neurophysiol* 89:1039–1056.
- Lee D, Seo H (2007) Mechanisms of reinforcement learning and decision making in the primate dorsolateral prefrontal cortex. *Ann NY Acad Sci* 1104:108–122.
- Lee D, Conroy ML, McGreevy BP, Barracough DJ (2004) Reinforcement learning and decision making in monkeys during a competitive game. *Brain Res Cogn Brain Res* 22:45–58.
- Leon MI, Shadlen MN (1999) Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24:415–425.
- McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38:339–346.
- Montague PR, King-Casas B, Cohen JD (2006) Imaging valuation models in human choice. *Annu Rev Neurosci* 29:417–448.
- Mookherjee D, Sopher B (1994) Learning behavior in an experimental matching pennies game. *Games Econ Behav* 7:62–91.
- O'Doherty JP (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr Opin Neurobiol* 14:769–776.
- O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. *Neuron* 38:329–337.
- Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441:223–226.
- Paton JJ, Belova MA, Morrison SE, Salzman CD (2006) The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* 439:865–870.
- Pawitan Y (2001) In all likelihood: statistical modelling and inference using likelihood. Oxford: Clarendon.
- Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400:233–238.
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337–1340.
- Schultz W (1998) Predictive reward signal of dopamine neurons. *J Neurophysiol* 80:1–27.
- Seo H, Lee D (2007) Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J Neurosci* 27:8366–8377.
- Seo H, Lee D (2008) Cortical mechanisms for reinforcement learning in competitive games. *Philos Trans R Soc Lond B Biol Sci* 363:3845–3857.
- Seo H, Lee D (2009) Behavioral and neural changes after gains and losses of conditioned reinforcers. *J Neurosci* 29:3627–3641.
- Seo H, Barracough DJ, Lee D (2007) Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cereb Cortex* 17:1110–1117.
- Shidara M, Richmond BJ (2002) Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science* 296:1709–1711.
- Simmons JM, Richmond BJ (2008) Dynamic changes in representations of preceding and upcoming reward in monkey orbitofrontal cortex. *Cereb Cortex* 18:93–103.
- Snedecor GW, Cochran WG (1989) Statistical methods, Ed 8. Ames, IA: Iowa State UP.
- Soltani A, Lee D, Wang XJ (2006) Neural mechanism for stochastic behaviour during a competitive game. *Neural Netw* 19:1075–1090.
- Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782–1787.
- Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.
- Thevarajah D, Mikulić A, Dorris MC (2009) Role of the superior colliculus in choosing mixed-strategy saccades. *J Neurosci* 29:1998–2008.
- Vickery TJ, Jiang YV (2009) Inferior parietal lobule supports decision making under uncertainty in humans. *Cereb Cortex* 19:916–925.
- von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton: Princeton UP.
- Watanabe M (1996) Reward expectancy in primate prefrontal neurons. *Nature* 382:629–632.
- Yang T, Shadlen MN (2007) Probabilistic reasoning by neurons. *Nature* 447:1075–1080.