Behavioral/Systems/Cognitive

Behavioral and Neural Changes after Gains and Losses of Conditioned Reinforcers

Hyojung Seo and Daeyeol Lee
Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06510

Human behaviors can be more powerfully influenced by conditioned reinforcers, such as money, than by primary reinforcers. Moreover, people often change their behaviors to avoid monetary losses. However, the effect of removing conditioned reinforcers on choices has not been explored in animals, and the neural mechanisms mediating the behavioral effects of gains and losses are not well understood. To investigate the behavioral and neural effects of gaining and losing a conditioned reinforcer, we trained rhesus monkeys for a matching pennies task in which the positive and negative values of its payoff matrix were realized by the delivery and removal of a conditioned reinforcer. Consistent with the findings previously obtained with non-negative payoffs and primary rewards, the animal's choice behavior during this task was nearly optimal. Nevertheless, the gain and loss of a conditioned reinforcer significantly increased and decreased, respectively, the tendency for the animal to choose the same target in subsequent trials. We also found that the neurons in the dorsomedial frontal cortex, dorsal anterior cingulate cortex, and dorsolateral prefrontal cortex often changed their activity according to whether the animal earned or lost a conditioned reinforcer in the current or previous trial. Moreover, many neurons in the dorsomedial frontal cortex also signaled the gain or loss occurring as a result of choosing a particular action as well as changes in the animal's behaviors resulting from such gains or losses. Thus, primate medial frontal cortex might mediate the behavioral effects of conditioned reinforcers and their losses.

Introduction

Behaviors are seldom rewarded immediately by primary reinforcers, such as food. Instead, strengths of many behaviors are enhanced or diminished by conditioned reinforcers that have been previously associated with primary reinforcements or punishments (Wolfe, 1936; Kelleher and Gollub, 1962; Kazdin, 1977). In humans, the loss of conditioned reinforcers as a result of a particular behavior can subsequently suppress the same behavior. Namely, the loss of conditioned reinforcers, such as money, is punishing, and this is referred to as response cost (Weiner, 1962; Kazdin, 1972, 1977). Although reinforcing effects of conditioned reinforcers are well documented for many animal species, how choices are controlled by the loss of conditioned reinforcers has been rarely studied in animals (Nader and Morgan, 2001; Pietras and Hackenberg, 2005). In the present study, we characterized how the gains and losses of conditioned reinforcers influence the choice behaviors of rhesus monkeys.

Previous studies on neural activity related to gains and losses of conditioned reinforcers were exclusively performed in humans. For example, studies based on scalp recordings and neuroimaging have shown that information about monetary gains and losses is rapidly processed (Gehring and Willoughby, 2002; Holroyd et al., 2004) and influences neural activity in multiple brain areas (Elliott et al., 1997; Delgado et al., 2000; Knutson et al., 2000; O'Doherty et al., 2001, 2003; Remijnse et al., 2005; Kim et al., 2006; Liu et al., 2007; Seymour et al., 2007; Wrase et al., 2007). In contrast, animal studies have mostly investigated the neural activity related to aversive outcomes using a pavlovian conditioning procedure in which the aversive outcomes are delivered regardless of the animal's behavior (Paton et al., 2006; Joshua et al., 2008; Matsumoto and Hikosaka, 2009), or using an avoidance task in which aversive outcomes could be almost entirely avoided (Nishijo et al., 1988; Koyama et al., 2001; Kobayashi et al., 2006; Hosokawa et al., 2007). As a result, it has remained difficult to elucidate the neural mechanisms responsible for adjusting the animal's behavioral strategy based on the aversive outcomes of its previous choices. Similarly, how the neural signals related to positive and negative outcomes influence subsequent choices oppositely is not known.

In the present study, we trained rhesus monkeys in a token-based binary choice task, in which tokens exchangeable with juice reward could be gained or lost. During this task, gains and losses are expressed in the same currency, which is critical for the comparison of neural activity related to gains and losses. We found that neurons modulating their activity according to the gains and losses of conditioned reinforcers were common in multiple regions of the prefrontal cortex. Furthermore, neurons in the dorsomedial frontal cortex were more likely to change their activity related to the animal's upcoming choice differently according to the previous choice and its outcome than those in the dorsal anterior cingulate cortex and dorsolateral prefrontal cortex. Thus, the dorsomedial frontal cortex might play a unique role in

Received Oct. 1, 2008; revised Feb. 16, 2009; accepted Feb. 19, 2009.

This work was supported by National Institutes of Health Grant MH073246. We are grateful to M. W. Jung for his helpful comments on this manuscript.

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

DOI:10.1523/JNEUROSCI.4726-08.2009

 0270-6474/09/293627-15$15.00/0

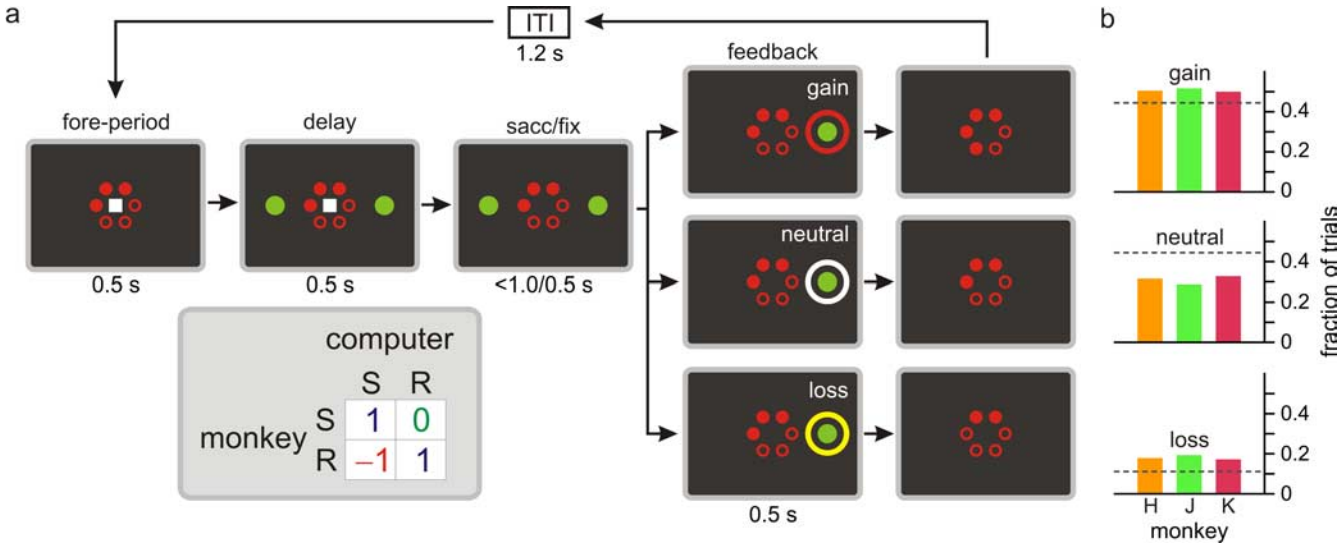

Figure 1. Behavioral task. ***a***, Spatiotemporal sequence of the task and payoff matrix used to determine the outcome of the animal's choice. S and R refer to safe and risky targets. ***b***, Frequencies of gain, neutral, and loss outcomes. Dotted lines indicate the outcome probabilities expected for the optimal strategy.

adjusting the animal's decision-making strategy based on the gains and losses of conditioned reinforcers.

Materials and Methods

Animal preparations and data acquisition. Two male (H and J; body weight, 9~11 kg) and one female (K; body weight, 6 kg) rhesus monkeys were used. Their eye positions were monitored at a sampling rate of 225 Hz with a high-speed eye tracker (ET49; Thomas Recording). Single-neuron activity was recorded from the dorsomedial frontal cortex (DMFC), dorsal anterior cingulate cortex (ACCd), or dorsolateral prefrontal cortex (DLPFC), using a five-channel multielectrode recording system (Thomas Recording) and a multichannel acquisition processor (Plexon). DMFC neurons were located in the supplementary eye field or its immediately surrounding areas (Tehovnik et al., 2000), whereas ACCd neurons were recorded from the dorsal bank of the cingulate sulcus (area 24c). DLPFC neurons were recorded anterior to the frontal eye field, as identified by eye movements evoked by electrical stimulation (Bruce et al., 1985). All the neurons encountered during the recording sessions were tested for the behavioral task described below without any prescreening. All the procedures used in this study were approved by the University of Rochester Committee on Animal Research and the Institutional Animal Care and Use Committee at Yale University, and conformed to the Public Health Service *Policy on Humane Care and Use of Laboratory Animals* and *Guide for the Care and Use of Laboratory Animals*.

Behavioral task. Animals were trained to perform a token-based binary choice task (Fig. 1*a*). After the animal fixated a central square during a 0.5 s foreperiod, two green peripheral targets were presented on a computer screen along the horizontal meridian. The animal was then required to shift its gaze toward one of the targets when the central square was extinguished at the end of a 0.5 s delay period. In addition, a series of red disks ("tokens") were displayed throughout the trial in a circular array at the center of the computer screen, and served as conditioned reinforcements, because they were exchanged with six drops of juice reward when the animal accumulated six of them. Once the animal fixated its chosen peripheral target for 0.5 s, a feedback ring appeared around the chosen target and its color indicated whether the number of tokens would increase (gain: red in all monkeys), decrease (loss: gray in monkey H; yellow in J and K), or remain unchanged (neutral: blue in monkey H; white in J and K). At the end of this 0.5 s feedback period, the number of tokens displayed on the screen was adjusted after gain or loss outcomes, and any remaining tokens were displayed continuously during the subsequent intertrial interval. To test whether the activity seemingly related to gains and losses reflected the color selectivity of the neurons, a subset of neurons recorded in two animals (monkeys J and K) were further tested in separate blocks in which the gain, neutral, and loss outcomes were signaled by gray, blue, and orange feedback rings, respectively. After juice delivery, the animal received two to four free tokens at the onset of the next trial.

The outcome of the animal's choice in a given trial was determined by the payoff matrix of a biased matching pennies game (Fig. 1*a*, inset) (Barraclough et al., 2004; Lee et al., 2004). During this game, the computer opponent simulated a rational decision maker who chooses its target so as to minimize the payoff of the animal, and the animal gained a token only when it chose the same target as the computer. When the animal's choice was different from that of the computer, its outcome was different for the two targets. For one target, referred to as "risky" target, the animal lost a token, whereas for the other target, referred to as "safe" target, the outcome was neutral, and the animal neither gained nor lost a token. The positions of the risky and safe targets were fixed in a block of trials, and their positions were changed without any cues with a 0.1 probability each trial after 40 trials. In game theory, a set of strategies from which none of the players can deviate individually to increase his or her payoff is referred to as a Nash equilibrium (Nash, 1950). The particular matching pennies game used in this study has a unique Nash equilibrium which corresponds to the animal and the computer opponent choosing the safe target with 2/3 and 1/3 probabilities, respectively. To show this, it should be noted that, at the equilibrium, the expected value of payoff should be equal for the two alternative targets. Denoting the probability of choosing the safe target for the animal and the computer opponent as p and q, this implies that the expected values of the animal's payoff from the safe and risky targets are q and $(-1)\ q + (1 - q)$, respectively. Therefore, $q = (-1)\ q + (1 - q)$, so $q = 1/3$ at the equilibrium. Similarly, $p + (-1)\ (1 - p) = (1 - p)$, so $p = 2/3$, because the expected payoffs from the two targets should be equal for the computer opponent. This implies that when both players make their choices according to the equilibrium strategies, the probabilities that the animal's choice would lead to the gain, neutral, and loss outcomes would be 4/9, 4/9, and 1/9, respectively. This implies that, against the competitive computer opponent, the animal would obtain on average 1/3 token each trial when it chose the risky and safe targets with 1/3 and 2/3 probabilities. If the animal's strategy in a given block deviated significantly from the equilibrium strategy, this was exploited by the computer opponent and would decrease the expected payoff for the animal. For example, if the animal chose the safe target more frequently than with 2/3 probability, the computer opponent always chose the risky target. In game theory, a strategy is referred to as pure when it corresponds to choosing a particular action exclusively, whereas a mixed strategy refers to a case in which multiple actions are chosen stochastically. Therefore, the equilibrium

strategy for the biased matching pennies task used in this study was mixed. An important advantage of using a game with a mixed equilibrium strategy is that it reduces the serial correlation between the successive choices of the animal and makes the outcome of each choice stochastic. This makes it possible to estimate the neural activity related to the animal's choice and its outcome in a given trial separately from those in the previous trial (Lee and Seo, 2007).

Analysis of behavioral data. To test how the animal's choice was affected by the gains and losses of tokens in previous trials, the following logistic regression model was applied (Lee et al., 2004):

$$\text{logit } p_t(\text{right}) \equiv \log p_t(\text{right})/p_t(\text{left}) = \mathbf{A_{beh}}[1\, U_{t-1} U_{t-2} \ldots U_{t-10}]', \quad (1)$$

where p_t(right) and p_t(left) refer to the probability of choosing the rightward and leftward targets in trial t, respectively; $U_t = [C_t\, G_t\, L_t\, R_t]$ is a row vector consisting of four separate regressors corresponding to the animal's choice ($C_t = 0$ and 1 for leftward and rightward choices, respectively), gain outcome ($G_t = -1$ and 1 for gaining a token from leftward and rightward choices, respectively, and 0 otherwise), loss outcome ($L_t = -1$ and 1 for losing a token from leftward and rightward choices, respectively, and 0 otherwise), and reward delivery ($R_t = -1$ and 1 if the animal obtains the sixth token and receives juice reward after choosing leftward and rightward targets, respectively, and 0 otherwise) in trial t; and $\mathbf{A_{beh}}$ is a vector of regression coefficients. A positive (negative) coefficient in this logistic regression model indicates that a particular outcome resulting from the choice of a given target reinforces (punishes) the same behavior. This was evaluated separately for multiple lags of trials. For example, a positive coefficient for gain with a lag of two trials indicates that when the animal's choice in a given trial (trial t) results in the gain of a token, the animal is more likely to choose the same target two trials later (trial $t + 2$). This model was fit to the animal's behavioral data in each session, and the regression coefficients related to the same variable and trial lag (e.g., G_{t-4}) were averaged separately.

Previous behavioral studies in humans and nonhuman primates have demonstrated that decision makers during competitive games might adjust their strategies dynamically according to a reinforcement learning algorithm (Sutton and Barto, 1998; Camerer, 2003; Lee et al., 2004, 2005). To test whether similar learning algorithms can account for the changes in the animal's choice behavior resulting from the gains and losses of tokens, we applied a reinforcement learning model, in which the value function was updated according to the following: $V_{t+1}(x) = V_t(x) + \alpha\{r_t - V_t(x)\}$, where $V_t(x)$ refers to the value function for choosing x in trial t, r_t is the outcome in trial t (−1, 0, and 1 for loss, neutral, and gain outcomes, respectively), and α is the learning rate. The value function for the unchosen target was not updated. The probability that the animal would choose the rightward target was given by the logistic (also known as softmax) transformation of the value functions as follows:

$$p_t(\text{right}) = \exp\beta V_t(\text{right})/\{\exp\beta V_t(\text{right}) + \exp\beta V_t(\text{left})\} = 1/[1 + \exp -\beta\{V_t(\text{right}) - V_t(\text{left})\}], \quad (2)$$

where β refers to the inverse temperature that controls the randomness of the animal's choice. The two parameters of this model (α and β) were estimated separately for each session according to a maximum-likelihood procedure (Pawitan, 2001; Seo and Lee, 2007).

To further test whether the neutral outcome was less reinforcing than the gain outcome and whether the loss outcome was more punishing than the neutral outcome, we also estimated the reward values of the neutral and loss outcomes as additional free parameters (r_{neutral} and r_{loss}) in the above reinforcement learning model, whereas the gain was still coded as +1. To avoid over-parametrization, the value of inverse temperature β was estimated and fixed for the entire dataset from each animal using the same reinforcement learning model described above in which the loss, neutral, and gain outcomes were coded as −1, 0, and +1, respectively.

To test whether the introduction of these additional parameters resulted in overfitting, we computed the Bayesian information criterion (BIC) for each model as follows (Burnham and Anderson, 2002): BIC = $-2 \log L + k \log N$, where L denotes the likelihood of a given model, k denotes the number of free parameters, and N denotes the number of trials. The second term in this formula penalizes the use of additional parameters, because the model with the smaller BIC is preferred.

Analysis of neural data. To investigate how neural activity was influenced by the gains and losses of tokens as well as the animal's choices, we focused our analyses on the spike rates during the delay period (0.5 s after target onset) and feedback period (0.5 s after feedback onset). The spike rate y_t during each of these intervals in trial t was then analyzed using the following multiple linear regression model:

$$y_t = a_0 + \mathbf{A_C}[C_t\, C_{t-1}]' + \mathbf{A_G}[G_t\, G_{t-1}]' + \mathbf{A_L}[L_t\, L_{t-1}]' + \mathbf{A_R}[R_t\, R_{t-1}]' + a_A \text{Asset}_t + a_R \text{Risk}_t + a_S \text{Sacc}_t + a_{\text{RT}} \text{RT}_t, \quad (3)$$

where C_t is the animal's choice in trial t ($C_t = -1$ and 1 for leftward and rightward choices, respectively), G_t is a dummy variable for the gain outcome ($G_t = 1$ for gain trials and 0 otherwise), L_t is a dummy variable for the loss outcome ($L_t = 1$ for loss trials and 0 otherwise), R_t is a dummy variable for the reward ($R_t = 1$ for rewarded trials and 0 otherwise), Asset_t is the number of tokens (0 to 5) possessed by the animal at the beginning of trial t, Risk_t is a dummy variable indicating whether the animal chose the risky target or not ($\text{Risk}_t = 1$ if the animal selects the risky target in trial t and 0 otherwise), Sacc_t is a dummy variable indicating the presence ($\text{Sacc}_t = 1$) or absence ($\text{Sacc}_t = 0$) of a saccade during the feedback period in trial t, and RT_t is the saccadic reaction time in milliseconds relative to feedback onset for trials in which the animal made a saccade during the feedback period ($\text{RT}_t = 0$, for the trials without saccades), and finally a_0, a_A, a_R, a_S, a_{RT}, $\mathbf{A_C}$, $\mathbf{A_G}$, $\mathbf{A_L}$, and $\mathbf{A_R}$ are regression coefficients. In addition to the variables related to the animal's choice and its outcome in the current trial t, this model included the corresponding variables from the previous trial $t - 1$, because it has been previously demonstrated that the neurons in some of the cortical areas tested in this present study often encode the signals related to the animal's choice and its outcome in the previous trial (Barraclough et al., 2004; Seo and Lee, 2007). We also tested whether the neural activity related to the gain and loss outcomes during the feedback period was influenced by the number of tokens owned by the animal. This was accomplished by adding to the above regression model (Eq. 3) the interaction term for gain and asset, $\text{Asset}_t \times G_t$, and the interaction term for loss and asset, $\text{Asset}_t \times L_t$.

Previous studies have found that reward-related activity of the neurons in the frontal cortex, such as the dorsolateral prefrontal cortex (Barraclough et al., 2004) and supplementary eye field (Uchida et al., 2007), might be influenced by the direction of the animal's eye movement. Such signals related to the conjunction of choice and outcome can be used to update the value function for a particular action (Seo et al., 2007; Seo and Lee, 2008). Therefore, to test whether the neural activity is influenced by specific conjunctions of the animal's choices and its outcomes, a set of interaction terms were added to the above regression model as follows:

$$\begin{aligned} y_t = a_0 &+ \mathbf{A_C}[C_t\, C_{t-1}]' + \mathbf{A_G}[G_t\, G_{t-1}]' + \mathbf{A_L}[L_t\, L_{t-1}]' + \\ &\mathbf{A_R}[R_t\, R_{t-1}]' + \mathbf{A_{G2}}[G_t \times C_t\, G_{t-1} \times C_t\, G_{t-1} \times C_{t-1}]' + \\ &\mathbf{A_{L2}}[L_t \times C_t\, L_{t-1} \times C_t\, L_{t-1} \times C_{t-1}]' + a_{C2} C_{t-1} \times C_t + \\ &a_{G3} G_{t-1} \times C_{t-1} \times C_t + a_{L3} L_{t-1} \times C_{t-1} \times C_t \\ &+ a_A \text{Asset}_t + a_R \text{Risk}_t + a_S \text{Sacc}_t + a_{\text{RT}} \text{RT}_t, \end{aligned} \quad (4)$$

where a_{C2}, $\mathbf{A_{G2}}$, and $\mathbf{A_{L2}}$ are the regression coefficients for the two-way interactions, and a_{G3} and a_{L3} are the regression coefficients for the three-way interactions that involve the gain or loss outcome in trial $t - 1$ as well as the animal's choices in trials $t - 1$ and t. It should be noted that, for the dummy variables used in the three-way interactions, their main effects and two-way interactions were also included in this model. Therefore, the regression coefficients for the three-way interactions estimated the conjunctive effects of three variables that were not accounted for by their main effects or two-way interactions (Aiken and West, 1991).

In the reinforcement learning model described above (Eq. 2), the probability of choosing a particular target is determined by the difference in the value functions for the two alternative targets. This suggests that neural activity related to the difference in the value functions, namely, $V_t(\text{right}) - V_t(\text{left})$, might influence the animal's choice. To investigate whether and how individual neurons contribute to the animal's upcoming choice according to the difference in the value functions, the following regression model was applied to the activity during the delay period.

$$y_t = a_0 + a_C C_t + a_A \text{Asset}_t + a_D\{V_t(\text{right}) - V_t(\text{left})\} + a_S\{V_t(\text{right}) + V_t(\text{left})\}, \quad (5)$$

where $V_t(x)$ is the value function for target x in trial t, and a_0, a_C, a_A, a_D, and a_S are the regression coefficients. We included the animal's choice, the asset, and the sum of the value functions in this model to control for any changes in neural activity that might be related to these additional variables (Seo and Lee, 2008). The value functions of successive trials are correlated, because they are updated iteratively, and this violates the independence assumption in the regression model. Therefore, the statistical significance for the regression coefficients in this model was determined by a permutation test (Seo and Lee, 2008). For this, we shuffled the trials separately for the blocks in which the left target was the safe target and those in which the right target was the safe target. We then estimated the value functions from these shuffled trials using the same parameters of the reinforcement learning model obtained for the unshuffled trials. This shuffling procedure was repeated 1000 times, and the p value for a given independent variable was determined by the fraction of the shuffles in which the magnitude of the regression coefficient from the shuffled trials exceeded that of the original regression coefficient.

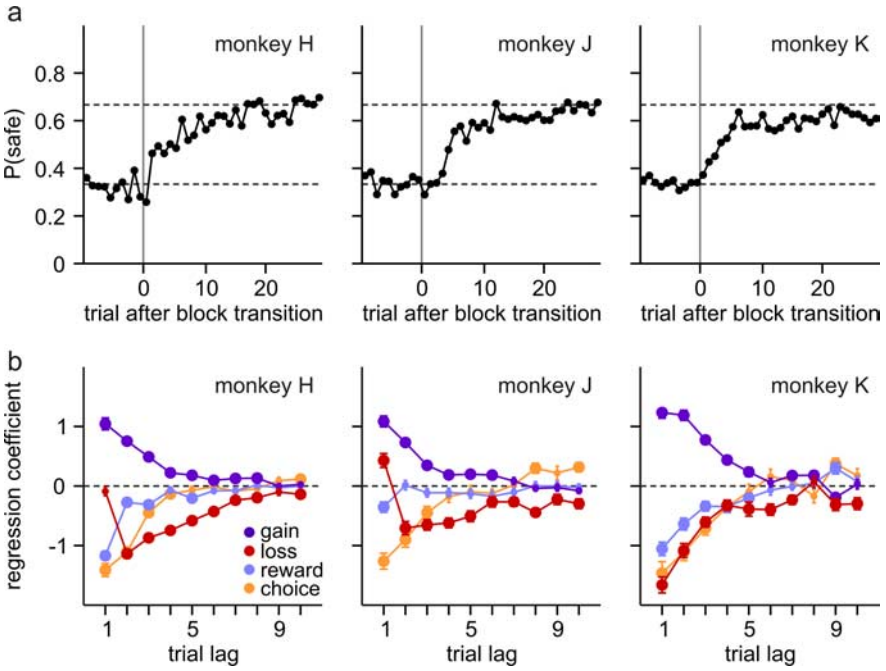

Figure 2. Behavioral performance in a token-based binary choice task. ***a***, Changes in the animal's choice probability after block transitions. P(safe) refers to the probability that the animal would choose the safe target in the new block (the risky target in the previous block). The dotted lines indicate the choice probabilities for safe (2/3) and risky (1/3) targets that correspond to the Nash equilibrium. ***b***, Average regression coefficients associated with gains, losses, choices, and juice rewards in 10 previous trials. The large circles indicate that they are significantly different from 0 (t test, $p < 0.05$). Error bars indicate SEM.

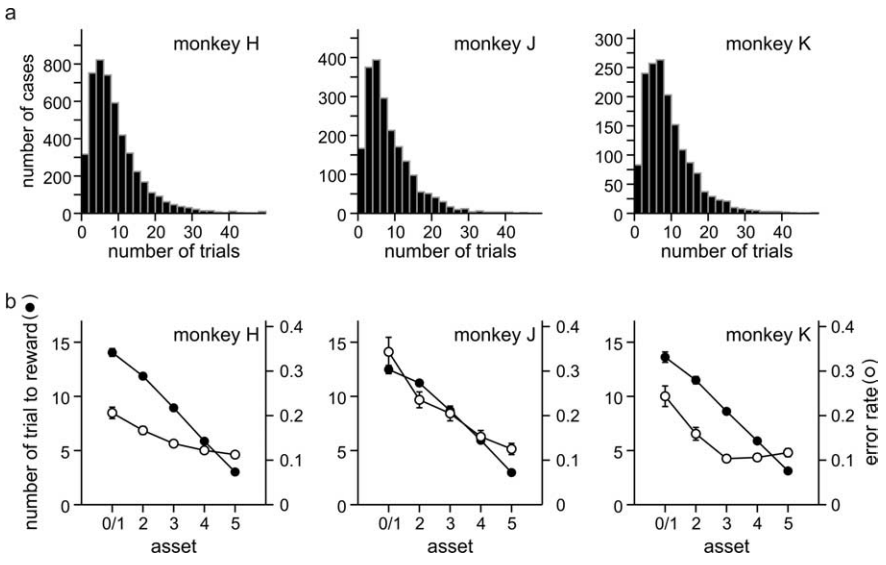

Figure 3. Frequency of rewards. ***a***, Frequency histograms for the number of trials between two successive reward delivery periods. ***b***, Average number of trials necessary to acquire juice rewards (filled symbols) and error rates (empty circles) plotted as a function of the number of tokens (assets) possessed by the animal. Error bars (SEM) are often smaller than the symbols.

Results

Effects of tokens on choices

Given that the computer opponent simulated a rational player in a zero-sum game, the optimal strategy for the animal during the token-based binary choice used in the present study was to choose the safe and risky target with 2/3 and 1/3 probabilities, respectively. Indeed, the probability that the animal would choose the safe or risky target approached the value predicted for the optimal strategy of the game within ~10 trials after the blocks switched (Fig. 2*a*). Nevertheless, the animal tended to choose the risky target significantly more frequently than predicted for the optimal strategy. Overall, the average fraction of trials in which the animal chose the risky target was 0.399, 0.403, 0.382 for monkeys H, J, and K, respectively, and this was significantly larger than 1/3 in all animals (t test, $p < 0.001$). The outcome of the animal's choice was a loss in 18.2% of the trials (Fig. 1*b*), which was also significantly higher than the value expected for the optimal strategy (1/9). The overall percentages of trials in which the animal's choice resulted in the neutral and gain outcomes were 31.1 and 50.8%, respectively (Fig. 1*b*).

Because the animal was rewarded with juice only when it accumulated six tokens and the outcomes in individual trials were probabilistic, the number of trials between the two successive juice deliveries varied substantially (Fig. 3*a*). Averaged across all animals, the number of trials between the two successive juice deliveries was 9.5. As expected, the average number of trials before the next juice delivery decreased monotonically with the

Table 1. Summary of model comparisons

	Monkey		
Models	H	J	K
Regression	864.0 (2)	794.7 (0)	693.1 (0)
Regression − CR	795.7 (1)	710.7 (0)	620.8 (0)
RL	730.9	643.4	554.7
RL + $r_{neutral}$ + r_{loss}	733.9 (11)	646.9 (5)	558.0 (4)
N sessions	82	37	35
N trials/session	574.7	509.7	457.2

Average values of BIC for different models used to analyze the choice behavior of each animal. The numbers in the parentheses indicate the number of sessions in which the corresponding model performed better than the simple reinforcement learning model (RL). Regression refers to the logistic regression model, whereas Regression − CR refers to the model without the terms related to choice and reward.

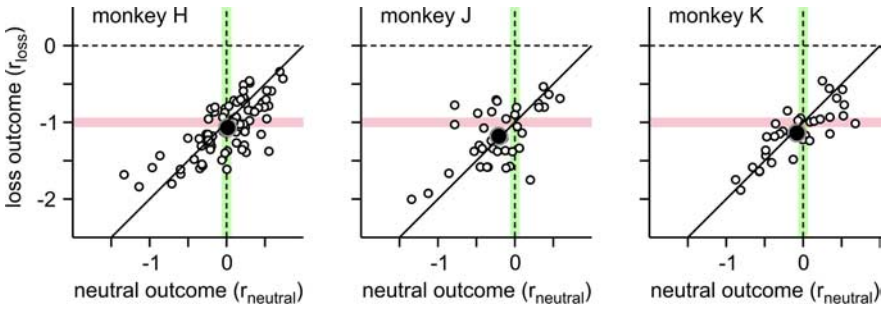

Figure 4. Reward values for the loss (r_{loss}) and neutral ($r_{neutral}$) outcomes estimated as free parameters in a reinforcement learning model. The pink and green lines correspond to the default values used for the loss (−1) and neutral (0) outcomes in the original reinforcement learning model. The solid black line corresponds to $r_{loss} = r_{neutral} - 1$. The empty circles correspond to the values obtained for individual sessions, whereas the large filled symbols correspond to their means for individual animals.

number of tokens (or assets) (Fig. 3*b*). Similarly, the frequency of error trials in which the animal broke their fixations prematurely decreased with the number of tokens possessed by the animal, presumably reflecting an increased level of motivation (Fig. 3*b*).

To test how the animal's choice was influenced by the outcomes of its previous choices, we applied a logistic regression model to the behavioral data (Eq. 1). The results showed that, in all three animals, the gain of a token from a particular choice increased the likelihood that the animal would choose the same target in the next trial, and this reinforcing effect decayed gradually over several trials (Fig. 2*b*; gain, blue). In contrast, the loss of a token from a particular choice decreased the likelihood that the same target would be chosen in subsequent trials (Fig. 2*b*; loss, red). The immediate effect of a token loss on the choice in the next trial varied across different animals, and this might be attributable to the fact that a strict "lose-switch" strategy could be exploited by the computer opponent and therefore was suboptimal. In addition, there was a significant tendency for the animals to avoid repeating the same choices in successive trials, as reflected in the negative coefficients associated with the animal's previous choices (Fig. 2*b*; choice, orange). Compared with the gains and losses of tokens, primary reinforcers had weaker effects on the animal's choice behavior and tended to counteract the reinforcing effects of tokens (Fig. 2*b*; reward, light blue).

In the logistic regression analysis, the effect of gaining or losing a conditioned reinforcer is modeled separately for each trial lag. Nevertheless, their weights decreased gradually with the number of intervening trials, suggesting that it might be more parsimonious to model the reinforcing and punishing effect of gains and losses, respectively, using a reinforcement learning model (Eq. 2). Indeed, according to the BIC, a simple reinforcement learning model performed better than the logistic regression model in almost all sessions (152 of 154 sessions) (Table 1). We also calculated the BIC for the logistic regression model after excluding the variables related to the previous choices and rewards. This reduced logistic regression model still performed more poorly than the reinforcement learning model, except for one session (Table 1).

In the above reinforcement learning model, the loss, neutral, and gain outcomes were respectively coded as −1, 0, and 1, respectively. To test directly whether the neutral outcome was less reinforcing than the gain outcome and whether the loss outcome was more punishing than the neutral outcome, we tested a modified reinforcement learning model in which the loss and neutral outcomes were coded by two separate free parameters (r_{loss} and $r_{neutral}$). This analysis showed that the influence of the loss outcome on the animal's choices was approximately equal and opposite to that of the gain outcome, whereas the effect of the neutral outcome was intermediate compared with the effects of gain and loss outcomes. First, the average value of the parameter r_{loss} was −1.07, −1.18, and −1.14 for monkeys H, J, and K, and therefore relatively close to −1, although for monkeys J and K, this was significantly more negative than −1 (*t* test, $p < 0.05$) (Fig. 4). Second, the average value of the parameter $r_{neutral}$ was significantly smaller than +1, and this was true for all animals (*t* test, $p < 10^{-28}$) (Fig. 4). Third, the average value of $r_{neutral}$ was 0.01, −0.21, and −0.08 for monkeys H, J, and K, respectively, and therefore relatively close to 0 (Fig. 4). The null hypothesis that $r_{neutral} = 0$ was rejected only for monkey J ($p < 0.05$). Finally, the average value of the parameter $r_{neutral}$ was significantly more positive than the average value of r_{loss}, which was again true for all animals (*t* test, $p < 10^{-16}$) (Fig. 4). Because the estimated values of loss and neutral outcomes were not substantially different from those used in the original reinforcement learning model, their use as free parameters might have resulted in the overfitting. Indeed, the BIC for this new reinforcement learning model was larger than the original reinforcement learning model in the majority of sessions (134 of 154 sessions) (Table 1). Therefore, it is parsimonious to assume that the effects of loss and gain outcomes are equal and opposite with the neutral outcomes having intermediate effects.

Neural activity related to gains and losses

Single-neuron activity was recorded from the DMFC (76 neurons), the ACCd (75 neurons), and the DLPFC (76 neurons) (Fig. 5). During the 0.5 s time window immediately after feedback onset, many neurons in these three cortical areas changed their activity significantly when the animal acquired or lost a token in the same trial, compared with when the outcome of their choice was neutral. For example, two of the three DMFC neurons illustrated in Figure 6, *a* and *b*, changed their activity significantly during the feedback period according to the outcome of the animal's choice in the same trial. The neuron illustrated in Figure 6*a* significantly increased and decreased its activity during the feedback period of loss and gain trials, respectively, compared with the activity in the trials with neutral outcomes. In contrast, the neuron in Figure 6*b* increased its activity during the feedback

period of gain trials compared with the activity in neutral trials, whereas it showed a modest but statistically significant decrease in its activity in loss trials (t test, $p < 0.005$). Neurons in the ACCd and DLPFC also commonly changed their activity during the feedback period according to the outcome of the animal's choice. For example, the ACCd neuron shown in Figure 7*a* significantly increased its activity during the feedback period of gain trials compared with the activity in neutral trials (Fig. 7*a*, top left). Similarly, the DLPFC neuron shown in Figure 7*b* significantly increased its activity during the feedback period of gain and loss trials compared with the activity in neutral trials (Fig. 7*b*, top left). Overall, the percentage of neurons showing significant activity changes during the feedback period in gain trials relative to the activity in neutral trials was 65.8, 62.7, and 54.0% for the DMFC, ACCd, and DLPFC, respectively, whereas the corresponding percentages for the loss trials was 39.5, 28.0, and 19.7% (Fig. 8*a*, left). To test whether the activity during the feedback period seemingly related to gains and losses resulted from the color selectivity of the neurons, a subset of neurons in each cortical area were tested with a different set of colors for the feedback rings. The results from this control experiment showed that the neural activity related to gains and losses were relatively unaffected by the colors of the feedback rings (Fig. 7; supplemental Fig. 1, available at www.jneurosci.org as supplemental material).

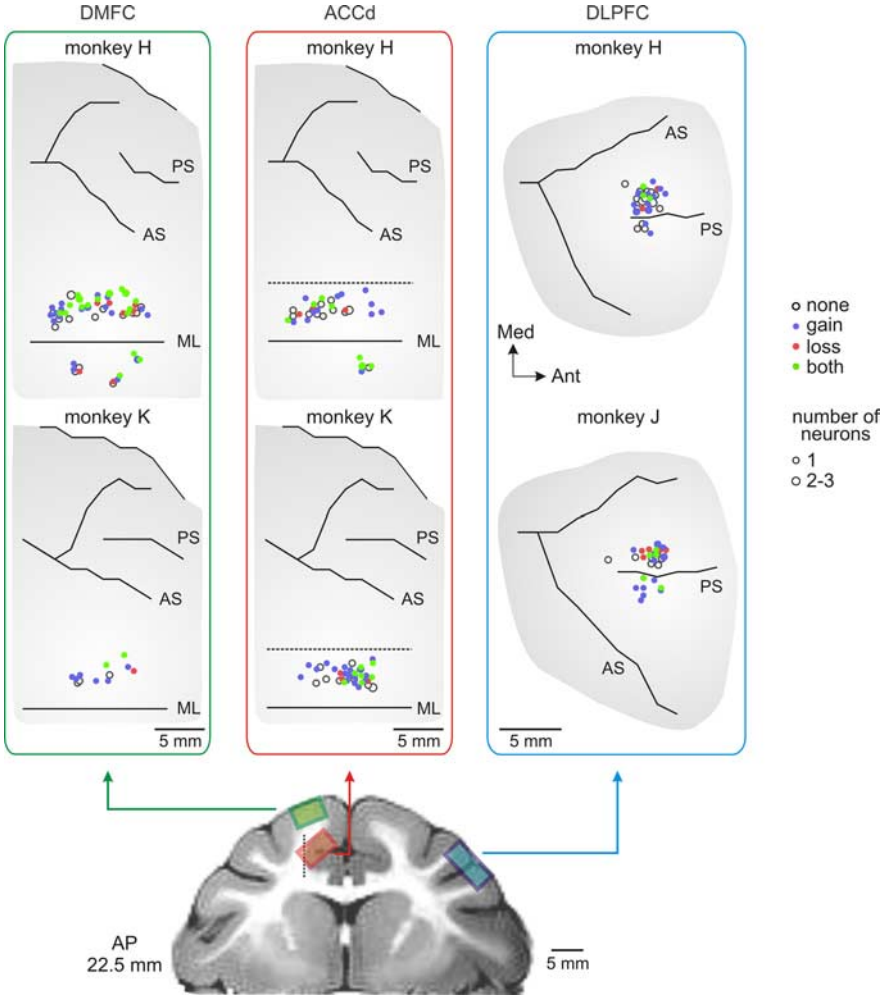

Figure 5. Anatomical distributions of the neurons in the DMFC, ACCd, and DLPFC. Neurons that showed significant effects for gain and/or loss are indicated in different colors. AS, Arcuate sulcus; ML, midline; PS, principal sulcus.

Whether the activity of individual neurons increased or decreased during the feedback period of gain and loss trials compared with the activity in neutral trials varied across neurons in each cortical area. In addition, whether the activity of a given neuron increased or decreased during the feedback period of gain trials compared with the activity in neutral trials was not systematically related to the activity related to the loss outcome. For example, compared with the activity in neutral trials, some neurons changed their activity during the feedback period oppositely for gain and loss outcomes (Fig. 6*a*,*b*). In contrast, some neurons changed their activity in response to either gains or losses compared with their activity in neutral trials (Fig. 7*a*), whereas others changed their activity similarly to both gains and losses (Fig. 7*b*). Accordingly, the regression coefficients related to gains were not significantly correlated with those related to losses in any of the cortical areas tested in this study (Fig. 9). Similarly, within each cortical area, whether a given neuron increased or decreased its activity significantly in gain and loss trials was not systematically related (χ^2 test, $p > 0.2$) (Table 2).

We also found that the gain or loss of a token in a given trial influenced the activity of many neurons during the delay period of the next trial. For example, the DMFC neuron shown in Figure 6*a* increased its activity significantly during the delay period when the outcome in the previous trial was a loss compared with when the previous outcome was neutral. Overall, during the delay period, 31.6, 37.3, and 21.1% of the neurons in the DMFC, ACCd, and DLPFC, displayed significant modulations in their activity related to the gain of a token in the previous trial, whereas the corresponding percentages for the loss outcome in the previous trial were 43.4, 21.3, and 17.1%, respectively (Fig. 8*a*, right). Therefore, signals related to the gains and losses in a given trial were reflected in the activity of neurons in the frontal cortex during the delay period of the next trial.

Conjunctive coding of choices and outcomes

A particular outcome influences the animal's subsequent behaviors differently depending on the action responsible for that outcome. For example, the gain of a token after choosing the leftward and rightward target tends to increase and decrease, respectively, the probability of choosing the leftward target in subsequent trials (Fig. 2*b*). This suggests that neurons involved in updating the animal's decision-making strategy might encode the animal's previous choice and its resulting gain or loss conjunctively. Indeed, during the feedback period, activity of neurons in the DMFC often encoded the animal's choice in the same trial and its outcome conjunctively, whereas during the delay period, neural activity tended to reflect the conjunction of the animal's choice in the previous trial and its outcome. In addition, neurons encoded specific conjunctions of the animal's choice and its outcomes more frequently in the DMFC than in the DLPFC or ACCd (Fig.

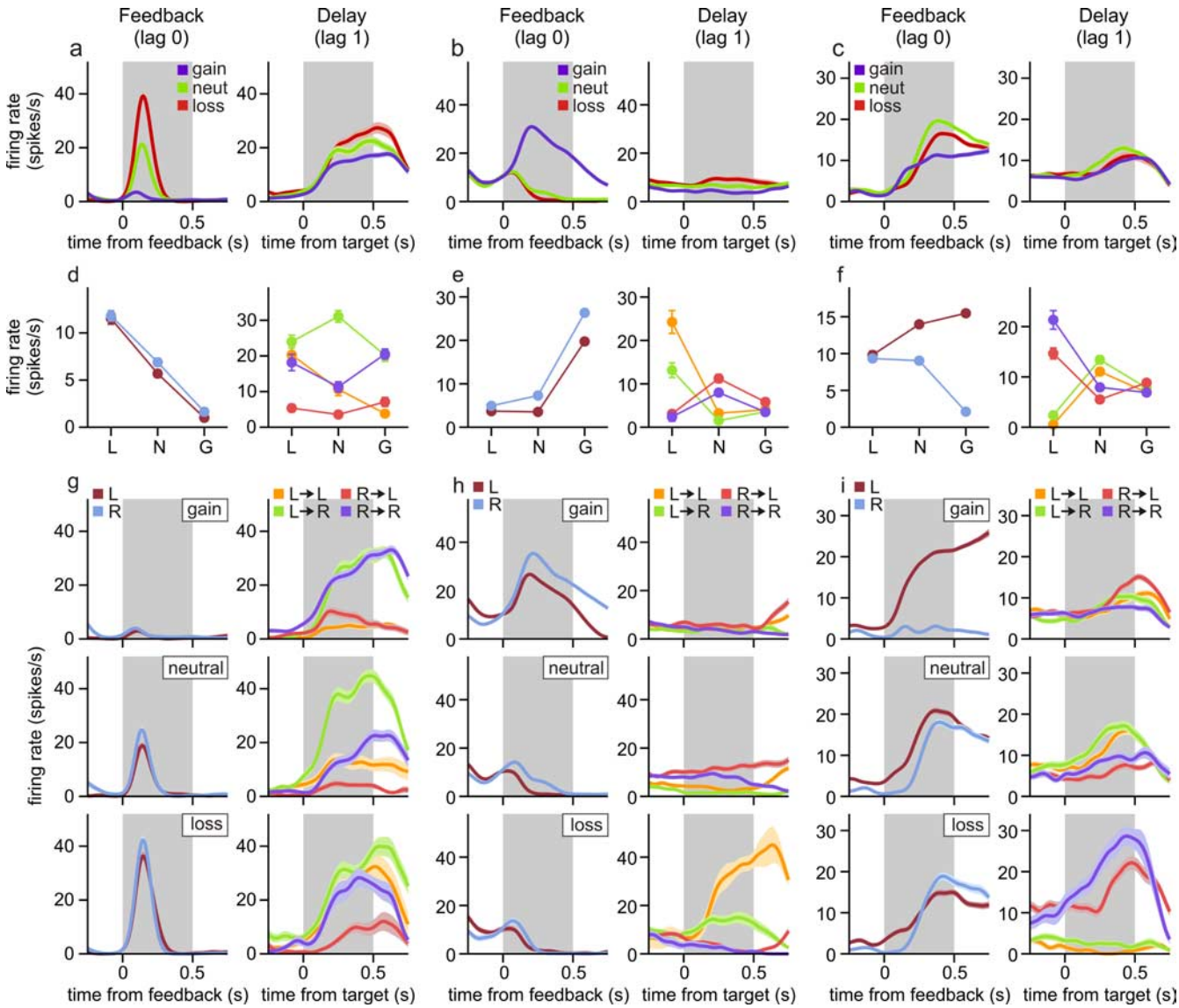

Figure 6. Neural activity related to gains and losses in the DMFC. ***a–c***, Spike density functions (SDFs) of three DMFC neurons during the feedback period (gray background, left) or delay period (gray background, right) in trial *t* sorted by the outcome of the current trial (*t*, left) or previous trial ($t-1$, right). ***d–f***, Left, Average spike rates during the feedback period of trial *t* sorted by the animal's choice and its outcome in the same trial (L, loss; N, neutral; G, gain). Right, Average spike rates during the delay period of trial *t* sorted by the animal's choice in the same trial in addition to the choice and its outcome in the previous trial ($t-1$). The animal's choices in two successive trials are color coded. For example, the orange line indicates that the animal chose the leftward target in both trials (L⟶L). ***g–i***, SDFs aligned to feedback onset (left) or target onset (right) sorted as in ***d–f***. Error bars in ***d–f*** indicate SEM; the shaded areas in the SDF plots indicate mean ± SEM.

8*b*), indicating that DMFC might play a key role in adaptively adjusting the animal's choice behaviors based on the gains and losses of conditioned reinforcers.

An example neuron in DMFC showing such conjunctive coding is illustrated in Figure 6*c*. The activity of this neuron during the feedback period of loss, neutral, and gain trials changed differently depending on the animal's choice in the same trial (Fig. 6*f*,*i*, left). In particular, when the animal chose the leftward target (Fig. 6*f*, purple), the activity of this neuron was higher during the feedback period of gain trials than in neutral or loss trials. In contrast, when the animal chose the rightward target (Fig. 6*f*, blue), the activity of the same neuron in the loss trials was higher than that in the gain or neutral trials. The results from the regression analysis with interaction terms (Eq. 4) confirmed that for the activity of this neuron during the feedback period, two-way interaction was significant both for the animal's choice and loss outcome and for the animal's choice and gain outcome in the same trial (*t* test, $p < 0.0001$). During the feedback period, 61.8 and 25.0% of the neurons in the DMFC showed significant interactions between the animal's choice and gain and between the choice and loss, respectively.

Similar to the activity during the feedback period, more than one-third of the neurons in the DMFC displayed significant interaction for the animal's choice in the previous trial and its outcome during the delay period (36.8 and 43.4% for gains and losses, respectively). Significant two-way interactions for choice and gain in the previous trial were found during the delay period for all three neurons shown in Figure 6, *a–c*, right, and two of these neurons also showed significant two-way interactions for choice and loss (Fig. 6*b*,*c*, right). Therefore, signals related to the conjunctions of the animal's choice and its outcome were still available in the DMFC, when the animal was ready to make its choice in the next trial (Fig. 8*b*, middle, choice × out). Compared with the neurons in the DMFC, the proportion of the neurons

showing such conjunctive activity related to the animal's choice and its outcome in the previous trial was significantly lower in the ACCd and DLPFC (χ^2 test, $p < 0.05$). For the ACCd, the percentages of neurons showing significant interactions for the animal's previous choice and gain outcome and for the previous choice and loss outcome was both 13.3%. For the DLPFC, these corresponding percentages were 9.2 and 18.4%, respectively.

Neurons mediating the effects of gains and losses on the animal's behavior must change their activity related to the animal's upcoming choice differently depending on the animal's previous choice and its outcome. For example, the neuron illustrated in Figure 6*a* showed a robust change in its activity during the delay period related to the animal's upcoming choice, but this differential activity was greatly reduced when the animal's leftward choice in the previous trial resulted in a loss (Fig. 6*d*,*g*, right; orange and green lines). For the neuron illustrated in Figure 6*b*, its activity reflected the animal's upcoming choice most robustly after the outcome of the leftward choice in the previous trial was a loss (Fig. 6*e*,*h*, right; orange and green lines). For both of these neurons, the activity during the delay period showed a significant three-way interaction between the animal's choice in the previous trial, the loss resulting from that choice, and the animal's choice in the current trial (*t* test, $p < 0.05$). The percentage of neurons in the DMFC with such significant three-way interactions was 21.1 and 23.7% for gains and losses (Fig. 8*b*, bottom), respectively, suggesting that some DMFC neurons encoded the animal's upcoming choice differently depending on its previous choice and outcome. As in the two-way interaction for the animal's choice and its outcome, the three-way interaction for the animal's choices in two successive trials and the previous outcome was observed more frequently in the DMFC than in the DLPFC or ACCd (Fig. 8*b*, bottom).

Coding of value functions

In the reinforcement learning models used to analyze the animal's choice behavior, the probability of choosing a particular target was a function of the difference in the value functions for the two alternative targets (Eq. 2). Therefore, neurons encoding the difference in the value functions might be involved in the selection of a particular target. Indeed, many neurons in the DMFC modulated their activity during the delay according to the difference in the value functions. For example, the activity

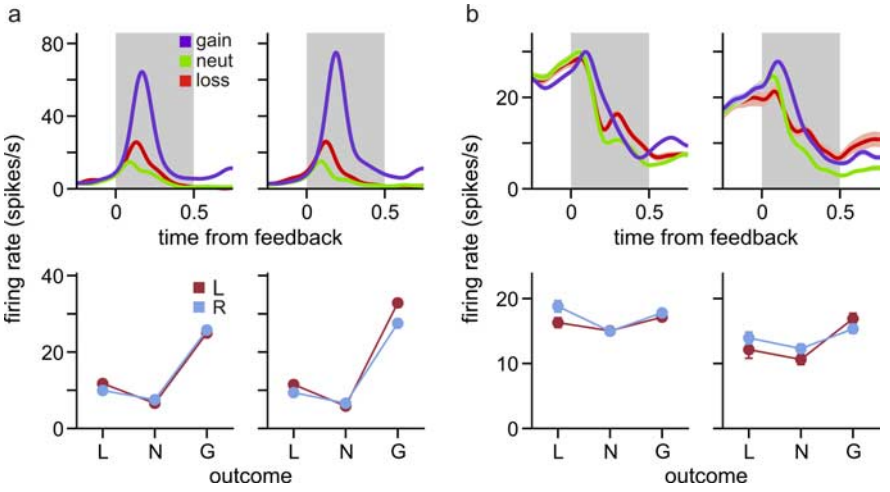

Figure 7. Activity of example neurons in the ACCd (***a***) and DLPFC (***b***). Top, Spike density function aligned to feedback onset sorted by the outcome of the animal's choice. The shaded areas indicate mean ± SEM. Bottom, Average spike rates during the feedback period sorted by the animal's choice and its outcome in the same trials (L, loss; N, neutral; G, gain). The left and right columns in each panel correspond to the activity during the trials tested with two different sets of feedback ring colors. Error bars indicate SEM.

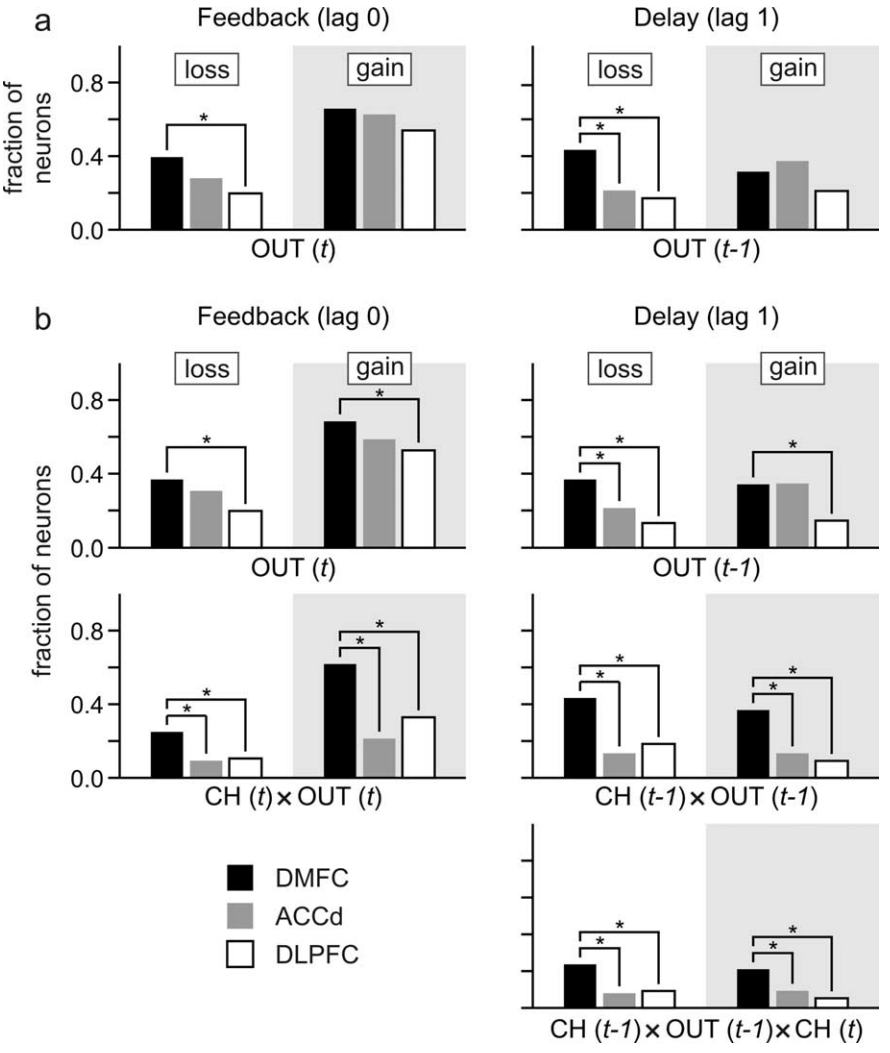

Figure 8. Population summary of neural activity related to gains and losses during feedback (left) and delay (right) periods. ***a***, Fraction of neurons showing significant effects of gains and losses, estimated with a regression model without interaction terms. ***b***, Fraction of neurons showing significant effects of gains and losses estimated along with their interactions with the animal's choices. CH(*t*) refers to the animal's choice in trial *t*, and OUT(*t*) the gain or loss outcome in trial *t*. *$p < 0.05$, χ^2 test.

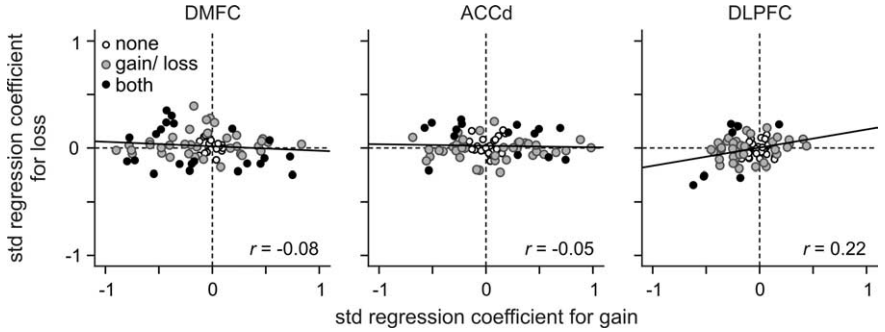

Figure 9. Population summary of activity related to gains and losses in different cortical areas. Standardized regression coefficients associated with gain outcomes are plotted against those associated with loss outcomes. The black symbols indicate the neurons that showed significant effects for both gain and loss, whereas the gray symbols indicate the neurons that showed significant effects for either gain or loss. The values within each plot show Spearman's rank correlation coefficient (r), and the solid line is the best-fitting regression line. The correlation was not significant for any of the cortical areas ($p > 0.05$).

Table 2. Contingency tables for the number of neurons that displayed significant modulations in their activity related to gains and losses during the feedback period

	Gain −	Gain +	No gain	All
DMFC ($\chi^2 = 5.062, p = 0.2810$)				
Loss −	7	6	2	15
Loss +	7	2	6	15
No loss	16	12	18	46
All	30	20	26	76
ACCd ($\chi^2 = 4.5890, p = 0.3321$)				
Loss −	1	3	4	8
Loss +	6	5	2	13
No loss	14	18	22	54
All	21	26	28	75
DLPFC ($\chi^2 = 1.4254, p = 0.8398$)				
Loss −	4	0	4	8
Loss +	3	1	3	7
No loss	24	9	28	61
All	31	10	35	76

Gain −/+ (Loss −/+), The number of neurons significantly decreasing/increasing their activity in gain (loss) trials (t test, $p < 0.05$). No gain (loss), The number of neurons without significant effect of gain (loss).

of the neuron illustrated in Figure 6*b* increased its activity during the delay period as the value function for the rightward target increased relative to that for the leftward target (Fig. 10*a*). This was true regardless of whether the animal chose the leftward or rightward target. In contrast, the neuron shown in Figure 6*c* increased its activity as the value function for the leftward target increased relative to that for the rightward target (Fig. 10*b*). Similarly, some neurons in the ACCd changed their activity according to the difference in the value functions for the two targets (Fig. 10*c*). To evaluate the activity related to the value functions statistically, we applied a regression model that included the difference in the value functions as an independent variable in addition to the animal's choice, asset, and the sum of the value functions (Eq. 5). The results showed that 47.4, 22.7, and 7.9% of the neurons in the DMFC, ACCd, and DLPFC, respectively, modulated their activity significantly according to the difference in the value functions (permutation test, $p < 0.05$) (Fig. 11*a*). Whereas these percentages were significantly higher than the 5% significance level (binomial test, $p < 0.05$) for the DMFC and ACCd, this was not the case for the DLPFC.

We hypothesized that the neurons encoding the conjunction of a particular choice and its outcome would signal the difference in the value functions. To test this, we compared the percentage of neurons that significantly changed their activity according to the difference in the value functions, separately for the neurons showing significant interaction effects for the animal's choice in the previous trial and gain outcome and the neurons without such interaction effects (Eq. 4). The same analysis was repeated separately for the loss outcome. The results mostly confirmed the hypothesis, especially for the DMFC. For example, among the DMFC neurons that showed significant interactions between the animal's choice and loss outcome, 72.7% of them modulated their activity significantly according to the difference in the value functions, whereas this percentage was reduced to 27.9% for the neurons that did not have such conjunctive coding (Fig. 11*a*). Neurons in the ACCd were also more likely to change their activity according to the difference in the value functions, when they showed significant conjunctive coding for choice and loss outcome compared with when they did not (Fig. 11*a*). Similarly, the neurons in the DMFC were more likely to encode the difference in the value functions when they showed significant conjunctive coding for choice and gain outcome (Fig. 11*a*). However, this was not true for the ACCd. The fact that the percentage of neurons encoding the difference in the value functions was affected consistently by the interactions between choice and outcome only for the DMFC is not surprising, because the overall percentage of neurons encoding the difference of value functions (Fig. 11*a*) or interactions between the animal's choice and outcome (Fig. 8*b*, middle) was relatively low for ACCd and DLPFC.

Conjunctive coding of the animal's choice and loss outcome, namely, the interaction between these two variables, was evaluated by comparing the activity during the delay period after the trials with loss and neutral outcomes. Therefore, for the neurons in the DMFC, we hypothesized that the neurons with such conjunctive coding might be more likely to encode the difference in the value functions when the outcome of the animal's choice in the previous trial was loss or neutral, compared with the previous outcome was gain. Similarly, because conjunctive coding of choice and gain outcome was evaluated by examining the activity related to the gain and neutral outcomes in the previous trials, we predicted that the DMFC neurons showing significant conjunction for choice and gain outcome in the previous trial would be more likely to encode the difference in the value functions when the previous outcome was gain or neutral compared with when the previous outcome was loss. To test these predictions, we computed the percentage of neurons that showed changes in their activity related to the difference in the value functions separately according to the outcome of the animal's choice in the previous trial. The results from this analysis mostly confirmed the predictions (Fig. 11*b*). For example, DMFC neurons that showed conjunctive coding of choice and loss outcome encoded the difference in the value functions more frequently than the neurons without such significant interactions, when the outcome in the previous trial was loss or neutral, but not when the previous outcome was gain. Similarly, DMFC neurons with

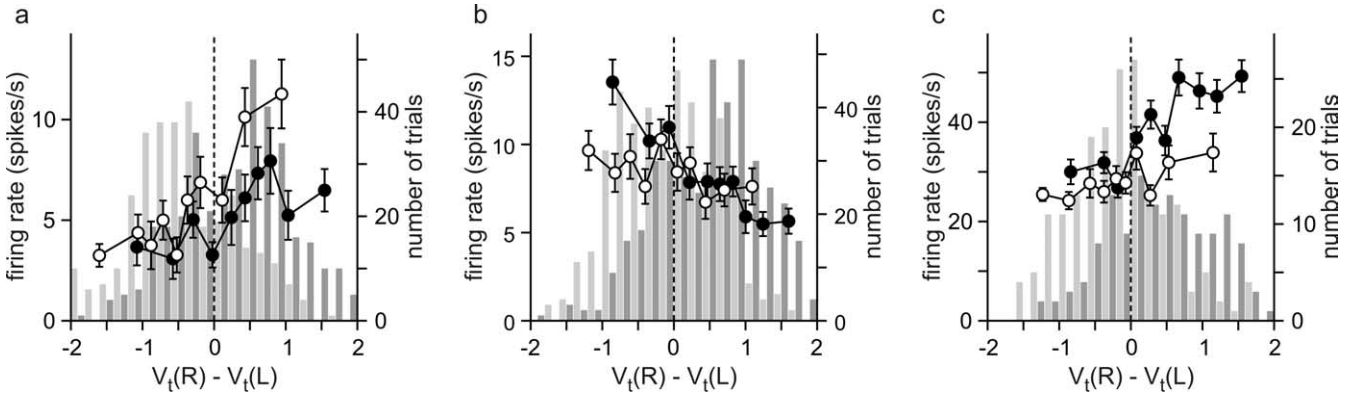

Figure 10. Example neurons in the DMFC (***a***, ***b***) and ACCd (***c***) modulating their activity according to the difference in the value functions for the two targets. The empty and filled circles represent the average spike rates during the delay period of the trials in which the animal chose the leftward and rightward targets, respectively. Each symbol corresponds to a decile of trials sorted by the difference in the value functions. Error bars indicate SEM. The light and dark histograms show the distribution of the difference in the value functions for the trials in which the animal chose the leftward and rightward targets, respectively.

significant conjunctive coding of choice and gain outcome changed their activity according to the difference in the value functions more frequently when the previous outcome was gain or neutral (Fig. 11*b*).

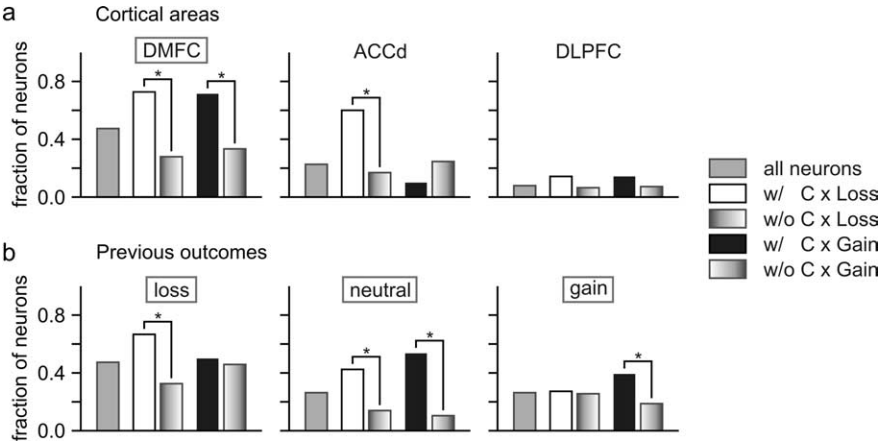

Figure 11. Population summary for neural activity related to value functions. ***a***, Fraction of neurons that significantly modulated their activity during the delay period according to the difference in the value functions. This is shown for all neurons in each cortical area (gray) and separately for a subset of neurons that showed (or did not show) significant interaction effects for the animal's choice in the previous trial and its outcome (gain or loss). ***b***, Fraction of neurons in the DMFC that changed their activity significantly according to the difference in the value functions when the animal's choice in the previous trial led to the loss, neutral, or gain outcome. As in ***a***, this was computed separately according to whether the neurons showed significant interaction between the animal's choice in the previous trial and its outcome (gain or loss). *χ^2 test, $p < 0.05$.

Neural activity related to assets and reward

The performance of the animals improved with the number of tokens owned by the animal, suggesting that the animal's motivation was systematically influenced by the number of tokens (Fig. 3*b*). Many neurons examined in this study also modulated their activity during the delay period according to the number of tokens. Neurons with such significant asset-related activity were found in all three cortical areas tested in this study, although they were more common in the DMFC (64.5%) and ACCd (57.3%) than in the DLPFC (23.7%). For example, ACCd neuron illustrated in Figure 12*a* increased its activity during the delay period gradually as the number of tokens increased, whereas the DMFC neuron in Figure 12*b* decreased its activity as the number of tokens increased. The estimates of this asset-related activity was relatively unaffected when a separate regressor was included to factor out the activity changes that might occur when the animal owned five tokens and therefore might have expected immediate reward. The regression coefficient for this new dummy variable was significant in 51.3, 45.3, and 11.0% of the neurons in the DMFC, ACCd, and DLPFC, respectively. Nevertheless, the percentage of neurons with significant asset-related activity in each of these areas was 44.7, 45.3, and 21.1%, with the addition of this dummy variable, and therefore relatively unaffected. This indicates that many neurons in the frontal cortex changed their activity gradually with the number of tokens.

This asset-related activity was evaluated, using a multiple linear regression analysis that controlled for the effect of gains and losses in the previous trial (Eq. 3). Nevertheless, for the neurons in the DMFC and ACCd, there was a significant positive correlation between the asset-related activity during the delay period and the gain-related activity during the feedback period (Fig. 12*c*). This suggests that the signals related to the number of tokens might be conveyed, in a consistent manner, by the same population of neurons that initially registered the positive outcomes of the animal's actions. In contrast, the asset-related signals during the delay period were not significantly correlated with the loss-related activity during the feedback period in any of the cortical areas tested in the present study (Fig. 12*c*).

Conjunctive coding of assets and outcomes

For the majority of the neurons recorded in this study, the sign and magnitude of activity related to gain and loss outcomes did not change with the assets significantly. For example, for the DMFC neuron shown in Figure 13*a*, its activity related to the gain, neutral, and loss outcomes was mostly unaffected by the assets during the feedback period. Similarly, although the activity of the ACCd neuron shown in Figure 13*b* gradually increased its activity during the feedback period according to the assets, the difference in the activity among the gain, neutral, and loss trials did not change significantly. In contrast, some neurons in each of

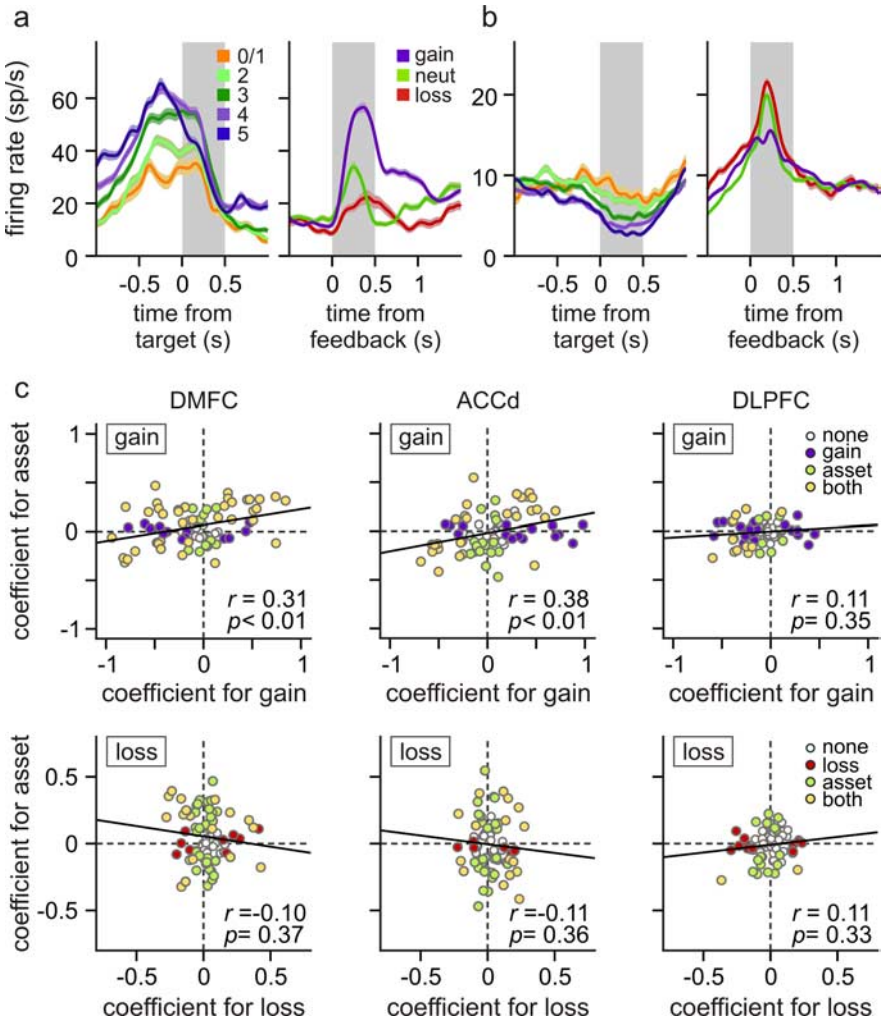

Figure 12. Example neurons in the ACCd (***a***) and DMFC (***b***) that significantly increase and decrease their activity with the number of tokens (assets) owned by the animal, respectively. The left panels show the activity aligned at the time of target onset separately for different levels of assets, whereas the right panels show the activity during the feedback period sorted by the outcomes. ***c***, Standardized regression coefficients related to assets for activity during the delay period (gray background in the left panels in ***a*** and ***b***) are plotted against the standardized regression coefficients related to the gains (top) and losses (bottom) for activity during the feedback period (gray background in the right panels in ***a*** and ***b***). Neurons in which the effects of assets and outcome were statistically significant are indicated by different colors. The values within each plot show Spearman's rank correlation coefficient (*r*), and the solid line is the best-fitting regression line.

the cortical areas tested in this study changed the magnitude of the outcome-related activity according to the assets. For example, for the DMFC neuron shown in Figure 13*c* and the ACCd neuron in Figure 13*d*, the difference in their activity during the feedback period of gain and neutral trials increased with the number of tokens. For the DMFC neuron shown in Figure 14*a*, the difference in its activity during the feedback period of loss and neutral trials gradually decreased with the number of tokens, whereas for the ACCd neuron shown in Figure 14*b*, this difference increased with the number of tokens. The statistical significance of such asset-dependent changes in outcome-related activity was evaluated by including the interaction terms for assets and outcomes in a regression model (see Materials and Methods). Overall, among the neurons that showed significant difference in their activity during the gain and neutral trials, the percentages of neurons that showed significant asset by gain interaction were 38.0, 42.6, and 22.0% for the DMPFC, ACCd, and DLPFC, respectively (Table 3). The corresponding percentages for asset by loss interaction were 36.7, 19.1, and 46.7% (Table 3).

To investigate further whether there was a significant tendency for the magnitude of activity changes related to gain outcomes to increase or decrease with the assets, we analyzed the standardized regression coefficients associated with the interaction term for assets and gain against the standardized regression coefficients for the gain outcome estimated from the regression model without the interaction terms (Eq. 3). These two coefficients would be positively correlated if the magnitude of gain-related activity increased with the assets. For example, these two coefficients would have the same signs for a neuron that increases the difference in its activity for gain and neutral trials as the number of tokens increases. In contrast, the signs of these two coefficients would be opposite if the magnitude of activity associated with gain outcomes decreases with assets. We also analyzed how the loss-related activity changes with assets.

For the DMFC and ACCd, we found that the coefficients related to the interaction term for gain and assets did not show any significant correlation with the coefficients for the gain outcome (supplemental Fig. 2, available at www.jneurosci.org as supplemental material). Spearman's correlation coefficient between these two variables was 0.108 ($p = 0.364$) and 0.077 ($p = 0.532$) for the DMFC and ACCd, respectively. Therefore, although there were a substantial number of neurons that significantly changed their gain-related activity according to the number of tokens, there was no consistent bias for gain-related activity to weaken or strengthen with the assets at the population level in the DMFC or ACCd. Accordingly, for these two cortical areas, difference in normalized activity between the gain and neutral trials did not change substantially, when averaged separately according to whether the neurons increased or decreased their activity during the gain trials relative to their activity in the neutral trials (Fig. 13*e*,*f*). Interestingly, for the DLPFC, the coefficients associated with asset by gain interaction were significantly and negatively correlated with those related with the gain outcome (Spearman's $r = -0.372$; $p = 0.046$). This was mostly attributable to the fact that all six neurons significantly decreasing their activity during gain trials compared with their activity during neutral trials showed significantly negative coefficients for asset by gain interactions (supplemental Fig. 2*c*, available at www.jneurosci.org as supplemental material). Consistent with this finding, the difference in the normalized activity for the trials with gain and neutral outcomes tended to decrease with assets for the DLPFC neurons that decreased their activity during the gain trials compared with the neutral trials (Fig. 13*g*, green vs blue lines). For the activity related to the loss outcomes, the coefficients related to asset by loss outcomes were not significantly correlated with the coefficients related to the loss outcome in any cortical areas. The corresponding values of the Spearman's

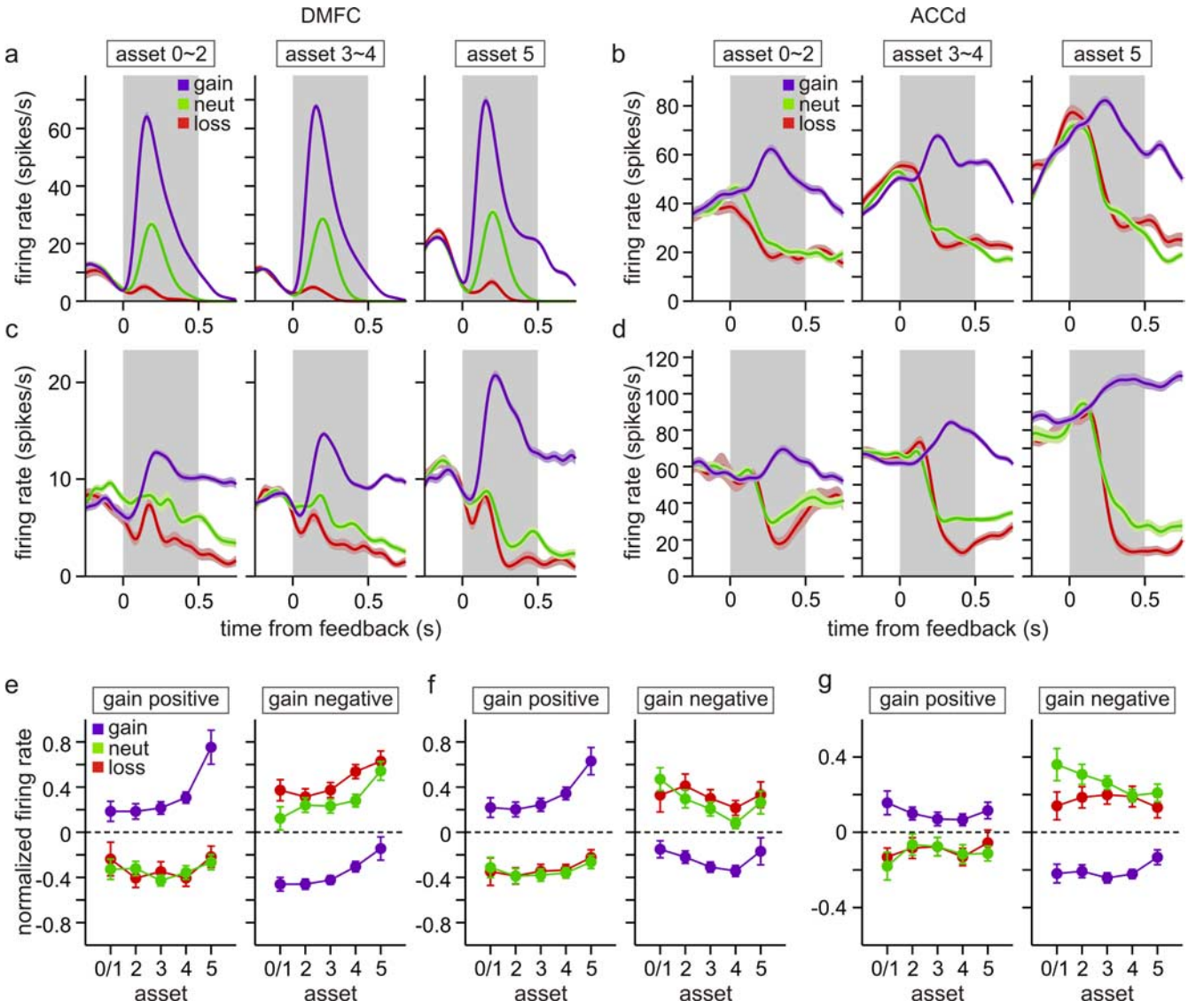

Figure 13. Effects of assets on gain-related activity. ***a***, ***b***, Two example neurons (***a***, DMFC; ***b***, ACCd) that did not change their gain-related activity with assets. The plots show spike density functions for each neuron during the feedback period (gray background) averaged separately according to the animal's choice outcome and the number of tokens (assets). ***c***, ***d***, Two example neurons (***c***, DMFC; ***d***, ACCd) that significantly changed their gain-related activity with assets. The shaded areas indicate mean ± SEM. ***e–g***, Normalized activity averaged across the neurons that increased (gain positive) or decreased (gain negative) their activity during the feedback period of gain trials relative to the activity in the neutral trials in each cortical area. For each neuron, the activity during the feedback period of individual trials was converted to *z*-scores, before they were averaged across neurons. Error bars indicate SEM.

correlation coefficient were −0.173, −0.089, and 0.046 for the DMFC, ACCd, and DLPFC, respectively ($p > 0.1$) (supplemental Fig. 2, available at www.jneurosci.org as supplemental material). Accordingly, the difference in the normalized activity averaged separately for loss and neutral outcomes did not change substantially with the number of tokens in any cortical areas, regardless of whether the neurons increased or decreased their activity during the loss trials compared with the activity in the neutral trials (Fig. 14*c–e*). In summary, there were a significant number of neurons in the frontal cortex that changed their activity related to gain or loss outcomes according to the number of tokens. However, at the population level, there were no systematic biases for such outcome-related activity to strengthen or weaken with the assets.

Discussion

Reinforcement learning in competitive games

The process of acquiring adaptive behavioral strategies can be described by reinforcement learning algorithms, in which the probability of taking each action is determined by a set of value functions (Sutton and Barto, 1998). These value functions are adjusted when reward or penalty received by the animal deviates from the predicted outcomes. Reinforcement learning algorithms have been successfully applied to many decision-making problems (Barraclough et al., 2004; Lee et al., 2004; Samejima et al., 2005; Daw and Doya, 2006; Matsumoto et al., 2007; Seo and Lee, 2007; Gold et al., 2008; Lau and Glimcher, 2008), including decision making during social interactions (Mookherjee and Sopher, 1994; Erev and Roth, 1998; Camerer, 2003; Lee, 2008). For example, animals in the present study approximated the optimal strategy during a competitive game according to a reinforcement learning algorithm. The payoff matrix during this game included a negative value, which was realized as the removal of a conditioned reinforcer. Using the analyses based on logistic regression models and reinforcement learning models, we demonstrated that the gains and losses of tokens resulting from the choice of a given target increased and decreased, respectively, the animal's

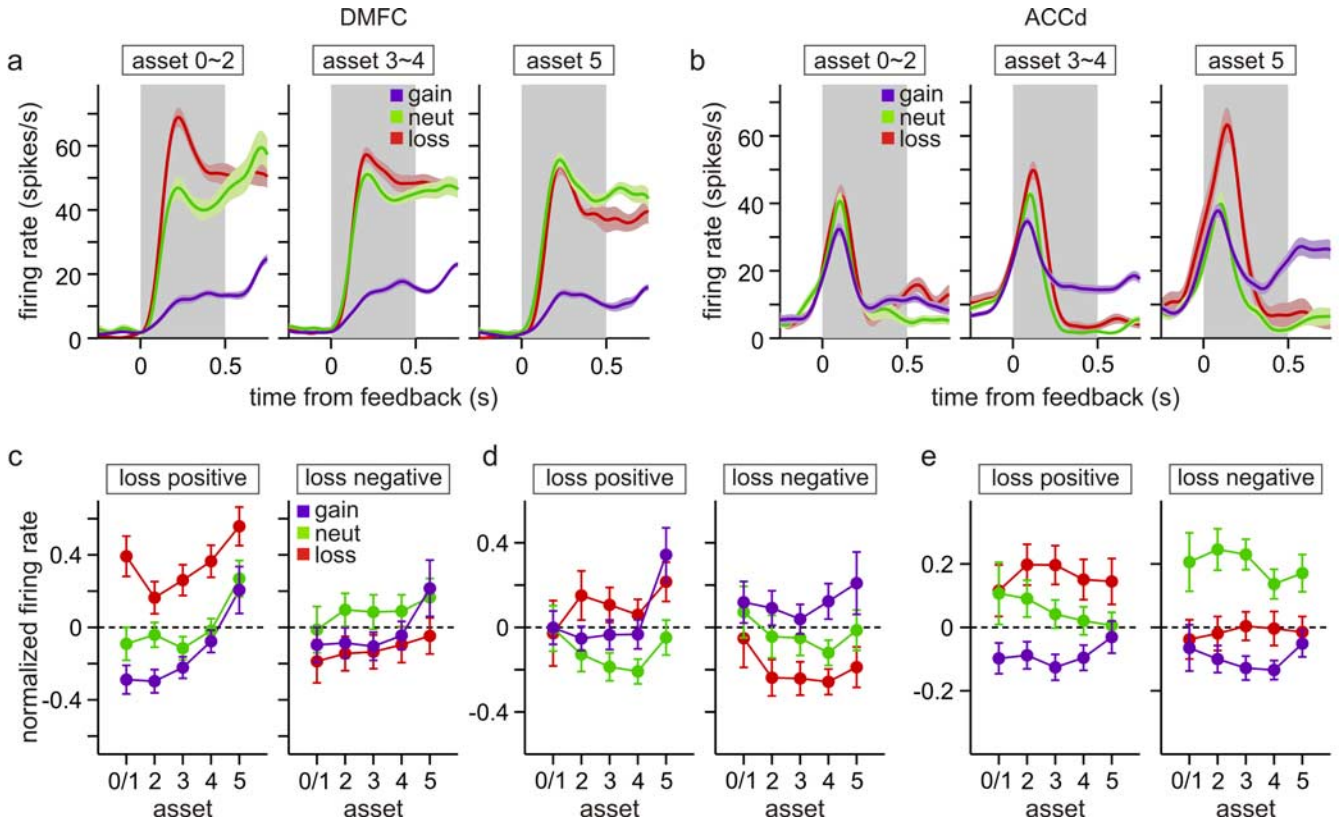

Figure 14. Effects of assets on loss-related activity. ***a***, ***b***, Two example neurons (***a***, DMFC; ***b***, ACCd) that significantly changed their loss-related activity with assets. The format is the same as in Figure 13*a*. ***c–e***, Normalized activity averaged across the neurons that increased (loss positive) or decreased (loss negative) their activity during the feedback period of loss trials relative to the activity in the neutral trials in each cortical area. The format is the same as in Figure 13*e*.

Table 3. Percentages and number of neurons showing significant interactions for asset and outcome

	Asset by gain		Asset by loss	
	Gain	No gain	Loss	No loss
DMFC	38.0 (19/50)	42.3 (11/26)	36.7 (11/30)	17.4 (8/46)
ACCd	42.6 (20/47)	28.6 (8/28)	19.1 (4/21)	22.2 (12/54)
DLPFC	22.0 (9/41)	17.1 (6/35)	46.7 (7/15)	4.9 (3/61)

Denominators inside the parentheses in the columns labeled Gain (No gain) and Loss (No loss) indicate the number of neurons with (without) significant gain and loss effects, respectively, whereas the numerators indicate the number of neurons with significant asset by gain or asset by loss interactions.

tendency to choose the same target in subsequent trials. The effects of gains and losses on the animal's behavior were approximately equal and opposite, whereas neutral outcomes tended to have intermediate effects. To our knowledge, this provides the first evidence that the choice behavior of nonhuman primates can be punished by the losses of conditioned reinforcers.

Neural signals related to gains and losses

Previous studies based on brain imaging and event-related potentials have implicated the medial frontal cortex in monitoring monetary gains and losses resulting from previous behaviors (Elliott et al., 1997; O'Doherty et al., 2001, 2003; Holroyd et al., 2004; Remijnse et al., 2005; Liu et al., 2007; Wrase et al., 2007). Some neurons in the primate anterior cingulate and dorsomedial frontal cortex also modulate their activity according to the outcome of the animal's action (Shima and Tanji, 1998; Matsumoto et al., 2007; Sallet et al., 2007; Seo and Lee, 2007; Uchida et al., 2007; Quilodran et al., 2008). In the present study, we found that the neurons in the DMFC, ACCd, and DLPFC also changed their activity when the animal lost a conditioned reinforcer, suggesting that such outcome-related activity did not simply signal the commission of error or omission of reward (Stuphorn et al., 2000; Ito et al., 2003; Holroyd et al., 2006), but also the abstract nature of decision outcomes. Interestingly, neurons encoding gain outcomes were found more frequently compared with loss-encoding neurons in all three cortical areas, although the magnitude of the behavioral effects were similar for gain and loss outcomes. Resolution of this apparent discrepancy awaits additional studies on the neural mechanisms responsible for translating the outcome-related activity to subsequent behavioral changes. In addition, although some neurons in each cortical area changed their activity related to gain or loss outcomes according to the number of tokens, there was no systematic bias for the activity related to gains or losses to weaken or strengthen as the animal accumulated more tokens. Therefore, at the population level, neurons in these cortical areas provided signals related to the outcomes of the animal's choices reliably throughout the task.

The results from this study also suggest that the DMFC might play a particularly important role in adjusting the animal's choice strategy based on conditioned reinforcers. First, compared with the ACCd and DLPFC, neurons in the DMFC modulated their activity more frequently according to gain and loss outcomes resulting from a particular action. For example, activity of neurons in the DMFC during the gain trials often differed depending on the action chosen by the animal. Activity of such neurons encoding the animal's choices and their outcomes conjunctively might represent whether a particular choice is likely to produce

better outcomes than other alternative choices. Indeed, in the DMFC and ACCd, neurons encoding choice and loss outcome conjunctively were more likely to change their activity according to the difference in the value functions than the neurons without such conjunctive coding. Similarly, DMFC neurons encoding choice and gain outcome conjunctively tended to show modulations in their activity related to the difference in the value functions frequently. Therefore, activity related to the conjunction of choice and outcome might provide a substrate for computing the difference in the value functions. Previous studies have shown that many neurons in the medial (Uchida et al., 2007) and lateral (Barraclough et al., 2004) prefrontal cortex also change their activity according to the presence or absence of reward differently according to the animal's action. Combined with the results from the present study, this suggests that signals related to the conjunctions of action and outcome might become more prevalent in the DMFC compared with the DLPFC, when the animal's choices lead to gains and losses of conditioned reinforcers without immediate rewards. Therefore, DMFC might play a particularly important role in adjusting the animal's behavioral strategies, when the feedback is provided in more abstract forms. Second, DMFC neurons often encoded the animal's upcoming choice in conjunction with the previous choice and its outcome. For example, some neurons might increase their activity before the animal chooses a particular target only when the outcome of the same choice in the previous trial was a loss. Therefore, DMFC might provide a neural substrate necessary for modifying the animal's behavioral strategy flexibly, even when the behaviors are not immediately rewarded or punished by primary reinforcers. Third, neurons in the DMFC and ACCd were more likely to change their activity systematically according to the number of tokens owned by the animals. These signals might reflect the level of reward expectancy (Shidara and Richmond, 2002; Satoh et al., 2003; Hikosaka and Watanabe, 2004; Nakahara et al., 2004; Roesch and Olson, 2004; Sohn and Lee, 2007), because the number of tokens was closely related to the number of trials before the juice reward. This also suggests that such asset-related signals might contribute to improving the animal's performance. During the task used in this study, tokens owned by the animal were constantly displayed on the computer screen, and therefore the animals were not required to store the number of tokens in its working memory. Nevertheless, the fact that gain-related signals during the feedback period were correlated with asset-related activity during the delay period suggests that the asset-related activity might at least in part arise from the gain-related activity that is temporally integrated.

Neural circuitry for conditioned reinforcement

By definition, conditioned reinforcers are conditioned stimuli predicting upcoming rewards or aversive stimuli, and can influence the animal's behavior, even when rewards or physically aversive stimuli are omitted. In addition, conditioned reinforcers can be delivered through different sensory modalities and linked to a large number of possible actions. Therefore, the neural mechanisms for processing various aspects of conditioned reinforcers and inducing the corresponding behavioral changes are likely to be implemented in multiple brain regions. Indeed, neurons that modulate their activity according to the reward or punishment expected from a particular sensory stimulus or motor response have been found in many different brain areas, including the prefrontal cortex (Watanabe, 1996; Rolls, 2000; Barraclough et al., 2004; Kim et al., 2008), posterior parietal cortex (Platt and Glimcher, 1999), basal ganglia (Schultz, 1998; Samejima et al., 2005), and amygdala (Nishijo et al., 1988; Paton et al., 2006; Tye et al., 2008). Some of these previous studies focused on identifying the neural processes involved in extracting the affective values of conditioned reinforcers. In particular, neurons in the orbitofrontal cortex and amygdala tend to track the value of reward or punishment expected from a particular stimulus more closely than its physical attributes (Nishijo et al., 1988; Rolls, 2000; Paton et al., 2006), suggesting that these two structures might play an important role in assigning the appropriate values to conditioned reinforcers (Parkinson et al., 2001; Pears et al., 2003; Burke et al., 2008). The results from the present study suggest that the DMFC might be a key structure in linking such outcome-predicting signals to appropriate actions. In addition, we found a substantial overlap between the populations of neurons encoding gains and losses of conditioned reinforcers. However, the source of signals related to gains and losses identified in the present study is presently unknown. The orbitofrontal cortex is connected to the medial and lateral prefrontal areas (Cavada et al., 2000), and the amygdala also projects to the medial frontal regions (Porrino et al., 1981). Consistent with this anatomical connectivity, we found that the signals related to gains and losses from a particular action were represented robustly in the DMFC. It would be important to test whether and how neurons in the orbitofrontal cortex and amygdala represent the losses of conditioned reinforcers in addition to their gains, and to determine whether the altered neural activity in orbitofrontal cortex and amygdala affects the signals related to gains and losses in the DMFC.

References

Aiken LS, West SG (1991) Multiple regression: testing and interpreting interactions. Thousand Oaks, CA: Sage.

Barraclough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci 7:404–410.

Bruce CJ, Goldberg ME, Bushnell MC, Stanton GB (1985) Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements. J Neurophysiol 54:714–734.

Burke KA, Franz TM, Miller DN, Schoenbaum G (2008) The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards. Nature 454:340–344.

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach. New York: Springer.

Camerer CF (2003) Behavioral game theory: experiments in strategic interaction. Princeton, NJ: Princeton UP.

Cavada C, Compañy T, Tejedor J, Cruz-Rizzolo RJ, Reinoso-Suárez F (2000) The anatomical connections of the macaque monkey orbitofrontal cortex. A review. Cereb Cortex 10:220–242.

Daw ND, Doya K (2006) The computational neurobiology of learning and reward. Curr Opin Neurobiol 16:199–204.

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000) Tracking the hemodynamic responses to reward and punishment in the striatum. J Neurophysiol 84:3072–3077.

Elliott R, Frith CD, Dolan RJ (1997) Differential neural response to positive and negative feedback in planning and guessing tasks. Neuropsychologia 35:1395–1404.

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.

Gehring WJ, Willoughby AR (2002) The medial frontal cortex and the rapid processing of monetary gains and losses. Science 295:2279–2282.

Gold JI, Law CT, Connolly P, Bennur S (2008) The relative influences of priors and sensory evidence on an oculomotor decision variable during perceptual learning. J Neurophysiol 100:2653–2668.

Hikosaka K, Watanabe M (2004) Long- and short-range reward expectancy in the primate orbitofrontal cortex. Eur J Neurosci 19:1046–1054.

Holroyd CB, Larsen JT, Cohen JD (2004) Context dependence of the event-related brain potential associated with reward and punishment. Psychophysiology 41:245–253.

Holroyd CB, Hajcak G, Larsen JT (2006) The good, the bad and the neural:

electrophysiological responses to feedback stimuli. Brain Res 1105:93–101.
Hosokawa T, Kato K, Inoue M, Mikami A (2007) Neurons in the macaque orbitofrontal cortex code relative preference of both rewarding and aversive outcomes. Neurosci Res 57:434–445.
Ito S, Stuphorn V, Brown JW, Schall JD (2003) Performance monitoring by the anterior cingulate cortex during saccade countermanding. Science 302:120–122.
Joshua M, Adler A, Mitelman R, Vaadia E, Bergman H (2008) Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. J Neurosci 28:11673–11684.
Kazdin AE (1972) Response cost: the removal of conditioned reinforcers for therapeutic change. Behav Ther 3:533–546.
Kazdin AE (1977) The token economy. New York: Plenum.
Kelleher RT, Gollub LR (1962) A review of positive conditioned reinforcement. J Exp Anal Behav 5:543–597.
Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol 4:e233.
Kim S, Hwang J, Lee D (2008) Prefrontal coding of temporally discounted values during intertemporal choice. Neuron 59:161–172.
Knutson B, Westdorp A, Kaiser E, Hommer D (2000) FMRI visualization of brain activity during a monetary incentive delay task. Neuroimage 12:20–27.
Kobayashi S, Nomoto K, Watanabe M, Hikosaka O, Schultz W, Sakagami M (2006) Influences of rewarding and aversive outcomes on activity in macaque lateral prefrontal cortex. Neuron 51:861–870.
Koyama T, Kato K, Tanaka YZ, Mikami A (2001) Anterior cingulate activity during pain-avoidance and reward tasks in monkeys. Neurosci Res 39:421–430.
Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451–463.
Lee D (2008) Game theory and neural basis of social decision making. Nat Neurosci 11:404–409.
Lee D, Seo H (2007) Mechanisms of reinforcement learning and decision making in the primate dorsolateral prefrontal cortex. Ann N Y Acad Sci 1104:108–122.
Lee D, Conroy ML, McGreevy BP, Barraclough DJ (2004) Reinforcement learning and decision making in monkeys during a competitive game. Cogn Brain Res 22:45–58.
Lee D, McGreevy BP, Barraclough DJ (2005) Learning and decision making in monkeys during a rock-paper-scissors game. Cogn Brain Res 25:416–430.
Liu X, Powell DK, Wang H, Gold BT, Corbly CR, Joseph JE (2007) Functional dissociation in frontal striatal areas for processing of positive and negative reward information. J Neurosci 27:4587–4597.
Matsumoto M, Hikosaka O (2009) Representation of negative motivational value in the primate lateral habenula. Nat Neurosci 12:77–84.
Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci 10:647–656.
Mookherjee D, Sopher B (1994) Learning behavior in an experimental matching pennies game. Games Econ Behav 7:62–91.
Nader MA, Morgan D (2001) Effects of negative punishment contingencies on cocaine self-administration by rhesus monkeys. Behav Pharmacol 12:91–99.
Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O (2004) Dopamine neurons can represent context-dependent prediction error. Neuron 41:269–280.
Nash JF (1950) Equilibrium points in *n*-person games. Proc Natl Acad Sci U S A 36:48–49.
Nishijo H, Ono T, Nishino H (1988) Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. J Neurosci 8:3570–3583.
O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci 4:95–102.
O'Doherty J, Critchley H, Deichmann R, Dolan RJ (2003) Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. J Neurosci 23:7931–7939.
Parkinson JA, Crofts HS, McGuigan M, Tomic DL, Everitt BJ, Roberts AC (2001) The role of the primate amygdala in conditioned reinforcement. J Neurosci 21:7770–7780.
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.
Pears A, Parkinson JA, Hopewell L, Everitt BJ, Roberts AC (2003) Lesions of the orbitofrontal but not medial prefrontal cortex disrupt conditioned reinforcement in primates. J Neurosci 23:11189–11201.
Pietras CJ, Hackenberg TD (2005) Response-cost punishment via token loss with pigeons. Behav Processes 69:343–356.
Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400:233–238.
Porrino LJ, Crane AM, Goldman-Rakic PS (1981) Direct and indirect pathways from the amygdala to the frontal lobe in rhesus monkeys. J Comp Neurol 198:121–136.
Quilodran R, Rothé M, Procyk E (2008) Behavioral shifts and action valuation in the anterior cingulate cortex. Neuron 57:314–325.
Remijnse PL, Nielen MM, Uylings HB, Veltman DJ (2005) Neural correlates of a reversal learning task with an affectively neutral baseline: an event-related fMRI study. Neuroimage 26:609–618.
Roesch MR, Olson CR (2004) Neuronal activity related to reward value and motivation in primate frontal cortex. Science 304:307–310.
Rolls ET (2000) The orbitofrontal cortex and reward. Cereb Cortex 10:284–294.
Sallet J, Quilodran R, Rothé M, Vezoli J, Joseph JP, Procyk E (2007) Expectations, gains, and losses in the anterior cingulate cortex. Cogn Affect Behav Neurosci 7:327–336.
Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340.
Satoh T, Nakai S, Sato T, Kimura M (2003) Correlated coding of motivation and outcome of decision by dopamine neurons. J Neurosci 23:9913–9923.
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, Barraclough DJ, Lee D (2007) Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. Cereb Cortex 17:i110–i117.
Seymour B, Daw N, Dayan P, Singer T, Dolan R (2007) Differential encoding of losses and gains in the human striatum. J Neurosci 27:4826–4831.
Shidara M, Richmond BJ (2002) Anterior cingulate: single neuronal signals related to degree of reward expectancy. Science 296:1709–1711.
Shima K, Tanji J (1998) Role for cingulate motor area cells in voluntary movement selection based on reward. Science 282:1335–1338.
Sohn JW, Lee D (2007) Order-dependent modulation of directional signals in the supplementary and supplementary motor areas. J Neurosci 27:13655–13666.
Stuphorn V, Taylor TL, Schall JD (2000) Performance monitoring by the supplementary eye field. Nature 408:857–860.
Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.
Tehovnik EJ, Sommer MA, Chou IH, Slocum WM, Schiller PH (2000) Eye fields in the frontal lobes of primates. Brain Res Brain Res Rev 32:413–448.
Tye KM, Stuber GD, de Ridder B, Bonci A, Janak PH (2008) Rapid strengthening of thalamo-amygdala synapses mediates cue-reward learning. Nature 453:1253–1257.
Uchida Y, Lu X, Ohmae S, Takahashi T, Kitazawa S (2007) Neuronal activity related to reward size and rewarded target position in primate supplementary eye field. J Neurosci 27:13750–13755.
Watanabe M (1996) Reward expectancy in primate prefrontal neurons. Nature 382:629–632.
Weiner H (1962) Some effects of response cost upon human operant behavior. J Exp Anal Behav 5:201–208.
Wolfe JB (1936) Effectiveness of token-rewards for chimpanzees. Comp Psychol Monogr 12:1–72.
Wrase J, Kahnt T, Schlagenhauf F, Beck A, Cohen MX, Knutson B, Heinz A (2007) Different neural systems adjust motor behavior in response to reward and punishment. Neuroimage 36:1253–1262.