Behavioral/Systems/Cognitive

Reinforcement Learning Signals Predict Future Decisions

Michael X Cohen,[1,2] **and Charan Ranganath**[2]

[1]Department of Epileptology and Center for Mind and Brain, University of Bonn, 53105 Bonn, Germany, and [2]Center for Neuroscience, University of California, Davis, Davis, California 95616

Optimal behavior in a competitive world requires the flexibility to adapt decision strategies based on recent outcomes. In the present study, we tested the hypothesis that this flexibility emerges through a reinforcement learning process, in which reward prediction errors are used dynamically to adjust representations of decision options. We recorded event-related brain potentials (ERPs) while subjects played a strategic economic game against a computer opponent to evaluate how neural responses to outcomes related to subsequent decision-making. Analyses of ERP data focused on the feedback-related negativity (FRN), an outcome-locked potential thought to reflect a neural prediction error signal. Consistent with predictions of a computational reinforcement learning model, we found that the magnitude of ERPs after losing to the computer opponent predicted whether subjects would change decision behavior on the subsequent trial. Furthermore, FRNs to decision outcomes were disproportionately larger over the motor cortex contralateral to the response hand that was used to make the decision. These findings provide novel evidence that humans engage a reinforcement learning process to adjust representations of competing decision options.

Key words: **reward prediction error; ERN; decision-making; reinforcement learning; dopamine; event-related potential**

Introduction

Recent research in neuroscience and computational modeling suggests that reinforcement learning theory provides a useful framework within which to study the neural mechanisms of reward-based learning and decision-making (Schultz et al., 1997; Sutton and Barto, 1998; Dayan and Balleine, 2002; Montague and Berns, 2002; Camerer, 2003). According to many reinforcement learning models, differences between expected and obtained reinforcements, or reward "prediction errors," can be used to form and adjust associations between actions or stimuli and their ensuing reinforcements (Sutton, 1992; Sutton and Barto, 1998; Montague and Berns, 2002). Critically, these models suggest that reward prediction errors can guide decision-making by signaling the need to adjust future behavior. In particular, larger prediction errors should be associated with adjustments in subsequent decisions, which occur because prediction errors strengthen or weaken representations of winning and losing actions, respectively.

Research using scalp-recorded event-related brain potentials (ERPs) in humans has revealed an ERP modulation called the "feedback-related negativity" (FRN) that might reflect a neural reward prediction error signal (Holroyd and Coles, 2002). The FRN is a relatively negative ERP deflection at frontocentral scalp sites ~200–400 ms after negative compared with positive feedback (Nieuwenhuis et al., 2002; Holroyd et al., 2003; Yasuda et al., 2004; Frank et al., 2005), and it reflects neural processes that share many characteristics with prediction errors (Schultz et al., 1997; Holroyd and Coles, 2002; Ruchsow et al., 2002; Nieuwenhuis et al., 2004; Yasuda et al., 2004; Frank et al., 2005). Holroyd and Coles (2002) suggested that the anterior cingulate cortex uses these prediction error signals to adapt reward-seeking behavior and demonstrated that a computational reinforcement learning model can emulate behavioral and neural responses during simple learning tasks. If neural prediction error signals are used to guide decision-making, as suggested by reinforcement learning models, we would expect that FRN magnitudes in response to decision outcomes should relate to subsequent decision behavior.

Accordingly, in the present study, we used ERPs to test how prediction errors might relate to adjustments in decision-making. In the experiment, subjects played a strategic game against a computer opponent and could maximize their winnings only by dynamically adjusting their decision strategies. We used prediction errors and decision option representations generated from a computational reinforcement learning model to generate novel hypotheses about human ERP and behavioral responses in this task. Based on the idea that prediction errors are used to adjust action representations, our analyses tested two critical predictions: (1) FRNs elicited by decision feedback should be related to how subjects adjusted their decision behavior on the subsequent trial, and (2) decision feedback should modulate the magnitude of FRNs recorded over motor cortex sites.

Materials and Methods

Subjects. Fifteen right-handed subjects (9 male, aged 21–28 years) were recruited from the University of Bonn community. Subjects had normal or corrected-to-normal vision and reported having no psychiatric conditions. The study was approved by the local ethics review committee,

Received March 9, 2006; revised Nov. 26, 2006; accepted Nov. 27, 2006.

This research was supported by Deutscher Akademischer Austauschdienst and the Institute for Research on Pathological Gambling and Related Disorders. M.X.C. is supported by a National Research Service Award from National Institute on Drug Abuse. We thank Beate Newport and Nicole David for technical assistance, Juergen Fell for insightful comments on previous drafts of this manuscript, and two anonymous reviewers for their useful suggestions and insights.

Correspondence should be addressed to Michael X Cohen, Center for Neuroscience, University of California, Davis, 1544 Newton Court, Davis, CA 95616. E-mail: mcohen@ucdavis.edu.

DOI:10.1523/JNEUROSCI.4421-06.2007

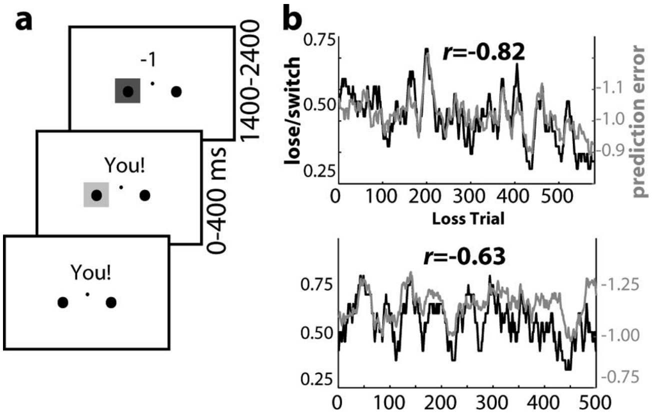

Figure 1. Trial events and correspondence between model outputs and behavior. ***a***, Depiction of trial events. Numbers indicate time of onset and offset of stimuli in milliseconds after the subjects' behavioral response. Displayed is a loss trial. ***b***, Outputs of the reinforcement learning model (gray lines) predicted subjects' trial-to-trial behavioral changes (black lines). Results are displayed for two subjects for whom the model closely fit the behavioral results. The calculated prediction error of the model on each loss trial closely matched the local fraction (calculated by smoothing behavioral choices, coded as 0 or 1, with a 10 trial kernel running-average filter) of subjects' loss/opposite versus loss/same choices on each of those trials.

and subjects signed informed consent documents before the start of the experiment.

Behavioral procedure. Subjects played a competitive, zero-sum game called "matching pennies" against a computer opponent. In the game, the subject and a computer opponent each selected one of two targets. If the subject and the computer opponent chose the same target, the subject lost one point, and if the subject and the computer opponent chose opposite targets, the subject won one point. On each of 1020 trials, subjects first saw the two targets, a fixation dot and "you!" on the screen, and pressed the left or right button with the left or right index finger on a response box to indicate their decision, which they were instructed to make as quickly possible. A green box surrounded their chosen target for 400 ms, followed by a 1000 ms delay, followed by the computer opponent's choice highlighted in violet and "−1" or "+1" displayed above the targets for 1000 ms. An intertrial interval of 1500 ms separated each trial (Fig. 1*a*).

The computer opponent was programmed to search for and exploit patterns in the subject's recent response history in attempt to beat the subject. Specifically, it kept in memory the subject's selections from the previous six trials and searched for patterns in these selections [left–right–left–right–left (and, by extension, right–left–right–left–right); left–left–right–right; left–left–left–right–right; win/stay–lose/switch]. The strategy search process was rote and preprogrammed (i.e., not a neural network or intelligent pattern classifier). If the computer opponent found a strategy, it chose the decision option that completed the pattern. For example, if the subject responded left–left–right–right, the computer opponent chose left on the following trial. This way, if the subject continued with this pattern, the computer opponent would win. In addition to searching for such patterns, the computer opponent also searched for an overall bias (one target is selected on at least four of the six previous trials). When no pattern was detected, the computer opponent chose a target randomly. An additional condition limited the number of successive wins to four.

The matching pennies game is often used to study economic decision-making and reinforcement learning (Mookherjee and Sopher, 1994; Sutton and Barto, 1998), and it has been used recently in neuroscience to study activity in the prefrontal cortex (Barraclough et al., 2004). The optimal decision policy in this game is to choose each target equally often and with no easily identifiable pattern. Thus, this game is useful for studying how reinforcements are used to adjust behavior on the trial-by-trial level rather than examining learning of optimal response patterns over a longer timescale. Furthermore, the competitive nature of this game helped ensure that subjects were constantly evaluating reinforcements and adjusting their behavior accordingly. Indeed, if subject's behavior was patterned and detectable by the computer opponent, it was possible, and in fact easy, to lose to the computer on 100% of trials. Thus, the results might be different if subjects selected both targets equally often but were selecting randomly rather than guided by the need to constantly adapt decision behavior.

ERP recording and analysis. EEG data were recorded at 1000 Hz (with an anti-aliasing low-pass filter set at 300 Hz) from 23 scalp electrodes spread out across the scalp and four ocular (two horizontal electrooculograms and two vertical electrooculograms) electrodes. All EEG channels were referenced to the left mastoid and were re-referenced on-line to the average of the left and right mastoids by the acquisition software. Scalp channels were Fpz, AFz, Fz, FCz, Cz, Pz, Oz, FC1, AF7, F3, C3, P3, CP5, FC5, T7, FC2, AF8, F4, C4, P4, CP6, FC6, and T8. Data were resampled to 250 Hz and band-pass filtered from 0.1 to 40 Hz off-line. Trials containing blink or other artifacts, identified as having voltage amplitudes greater than ±90 μV, were removed before averaging (mean ± SD, 4 ± 3%). Although the topography of the FRN is fairly anterior in our study and in other studies, these effects are not likely contaminated by eyeblinks, because eyeblink artifacts are more anterior and are spread out across the *x*-axis of the scalp (i.e., extending from eye to eye) (Jung et al., 2000; Li et al., 2006) rather than being small and focused in central electrodes.

Statistical analyses were performed by entering average ERP voltage potentials from a 240–260 ms post-feedback window into a 2 (feedback: positive or negative) × 2 (decision on following trial: same or opposite) repeated-measures ANOVA. We chose this time window based on the peak of the FRN (the loss–win difference) from electrode FCz, which occurred at 250 ms. ERPs were averaged across time windows because average amplitude measures are more robust than peak amplitude measures with respect to noise fluctuations in ERP waveforms. We selected FCz for analyses based on the loss–win difference topography, which demonstrates that the FRN was maximal at this site. Electrode sites C3 and C4 were used in analyses involving motor potentials (for spatial position of electrodes FCz, C3, and C4, see red circles in Figs. 3, 6) For illustration purposes, we *z*-transformed the ERP data in Figure 4 so they are more easily visually compared with results from the model. We calculated and plotted the predictive FRN (pFRN) (see Results) as the difference between loss/opposite and loss/same trials.

Reinforcement learning model. To test whether ERP and behavioral responses reflected a reinforcement learning process, we examined responses of a computational reinforcement learning model. The model used a reward prediction error to update weights associated with each target and probabilistically chose the target with the stronger weight (Schultz et al., 1997; Egelman et al., 1998; Holroyd and Coles, 2002; Montague and Berns, 2002; Schultz, 2004). Thus, after receiving negative feedback, the model generates a negative prediction error, which is used to decrease the strength of the weight of the chosen decision option (e.g., the right-hand target), making the model less likely to choose that decision option on the following trial. Specifically, the probability (p) of choosing the right-hand target on trial t is the logit transform of the difference in the weights on each trial (w_t) associated with each target, passed through a biasing sigmoid function (Egelman et al., 1998; Montague et al., 2004):

$$p(\text{right})_t = \frac{\exp(w(\text{right})_t)}{\exp(w(\text{right})_t) + \exp(w(\text{left})_t)}.$$

After each trial, a prediction error (δ) is calculated as the difference between the outcome received (−1 or 1 for losses and wins) and the weight for the chosen target [e.g., $\delta = -1 - w(\text{right})_t$ in the case in which the model lost after choosing the right-hand target]. Weights are then updated according to $w_{t+1} = \alpha \times w_t + \pi \times \eta \times \delta$, where α is a discount parameter, π is 1 for the chosen target and 0 for the nonchosen target, and η is the learning rate, which scales the effect of the prediction error on future weights. Note that, in this model, there is no temporal discounting that occurs between the response and the outcome. In the current study, it would not make sense to discount action values from the time of the response until the time of outcome, because the interval between the response and the outcome was fixed at 1500 ms and because the outcomes unequivocally resulted from the preceding response. Instead, the model discounts weights from previous trials, as in other studies (Barraclough et al., 2004; Cohen, 2006), rather than discounting the value of the action before the receipt of the outcome.

Many computational learning models exist, some of which might perform as well or better on this task as the model we used. We chose our model because (1) it has a proposed neurobiological basis (for review, see Montague et al., 2004) and thus makes testable predictions appropriate for neuroscience data, and (2) similar models have been used previously to study ERP correlates of reinforcement learning (Holroyd and Coles, 2002; Nieuwenhuis et al., 2002, 2004). Nonetheless, other prediction-error-driven learning models could be used that would generate similar predictions (for an example, see supplemental information, available at www.jneurosci.org as supplemental material).

We used the model in two different ways. First, we had the model play the matching pennies game against the computer opponent. This was done to assess the performance of the model and examine its prediction errors and decision option weights as it played the game. For these analyses, we set α to 0.8 and η to 1 for both wins and losses. Second, we wanted to examine whether the subjects' behavioral responses reflected a reinforcement learning process. To do this, we gave the model the unique history of decisions and reinforcements from each subject and compared subjects' behavioral and ERP responses to the reward prediction error and weights for the two decision options that the model calculated for each trial for each subject. In these analyses, we mathematically estimated α and η for each subject using a maximum likelihood minimization procedure in Matlab 6.5 (MathWorks, Natick, MA) (Luce, 1999; Barraclough et al., 2004; Cohen and Ranganath, 2005; Cohen, 2006). The procedure uses the nonlinear, unconstrained Nelder–Mead simplex method (Lagarias et al., 1998) to find values of the learning parameters that maximize the sum of $p(\text{right})_t$ or $p(\text{left})_t$ across the experiment (depending on the target selected on trial t). Finally, to examine the correspondence between behavioral responses and model outputs, we coded "loss/stay" (i.e., the subject lost on trial t and chose the same target on trial $t + 1$) and "loss/switch" trials as 0 and 1, respectively, and smoothed the resulting vector with a running average filter with a 10 trial kernel, an analysis often used to examine correspondence between model predictions and behavioral selections (Sugrue et al., 2004; Bayer and Glimcher, 2005; Samejima et al., 2005).

Results

Behavioral results

Subjects won an average ± SE of 51.4 ± 1.1% of trials. To examine whether behavioral responses reflected a reinforcement learning process, we compared subjects' behavioral choices to outputs of the model. The model makes two predictions about subjects' behavioral choices during the task: (1) larger negative prediction errors make the subject more likely to choose the opposite target on the following trial, and (2) the stronger the weight of a decision option, the more likely the subject is to choose that decision option. To test the first hypothesis, we used the model to calculate prediction errors on each trial for each subject and compared these prediction errors to the local fraction of "loss/same" (i.e., when the subject lost and chose the same target on the following trial as on the current one) versus "loss/opposite" (i.e., when the subject lost and chose the opposite target on the following trial as on the current one) decisions. As seen in Figure 1*b*, the predictions of the model correlated with the subjects' behavior. Specifically, larger negative prediction errors calculated by the model were associated with increased likelihood of subjects choosing the opposite target on the following trial. The correlation between these variables was significant across subjects (average $r = -0.29$; $p = 0.007$).

We tested the second prediction, that weights calculated by the model would correspond to left-hand versus right-hand decisions chosen by the subjects, in a similar manner: the model calculated weights of the two targets for each trial, based on each subject's unique history of decisions, and the difference between the two weights at each trial was compared with the local fraction of left-have versus right-hand target selections (coded as 0 or 1). Again, we found a correspondence between what the model estimated the weights should be and what the subjects actually chose (supplemental Fig. S4, available at www.jneurosci.org as supplemental material). Specifically, greater relative weights of the right-hand versus left-hand target were associated with increased likelihood that the subjects would choose the right-hand target. This correlation was significant across subjects (average $r = 0.35$; $p < 0.001$).

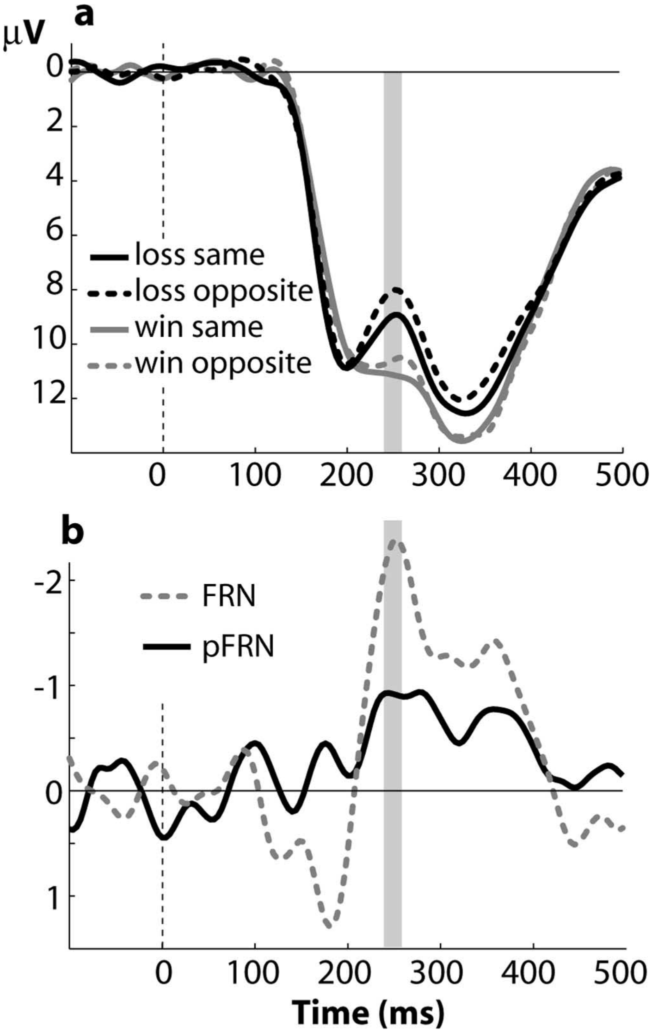

Figure 2. Feedback-locked ERPs sorted according to current outcome and future decision. ***a***, Grand-average ERPs after losses (black) and wins (gray) separated according to whether subjects chose the opposite (dashed lines) or the same (solid lines) target on the following trial as on the current trial. Light gray bar indicates time window used for analyses. ***b***, Grand-averaged FRN (loss–win effect; dotted gray line) and pFRN (loss/opposite − loss/same trials; solid black line) plotted over time.

ERP results

The correspondence between the behavioral data and the model suggests that humans engage a reinforcement learning-like process during the task. We next sought to investigate the neural mechanisms that might underlie this process by examining ERPs recorded while subjects played the game. Consistent with previous studies of the FRN, feedback-locked ERPs recorded at fron-

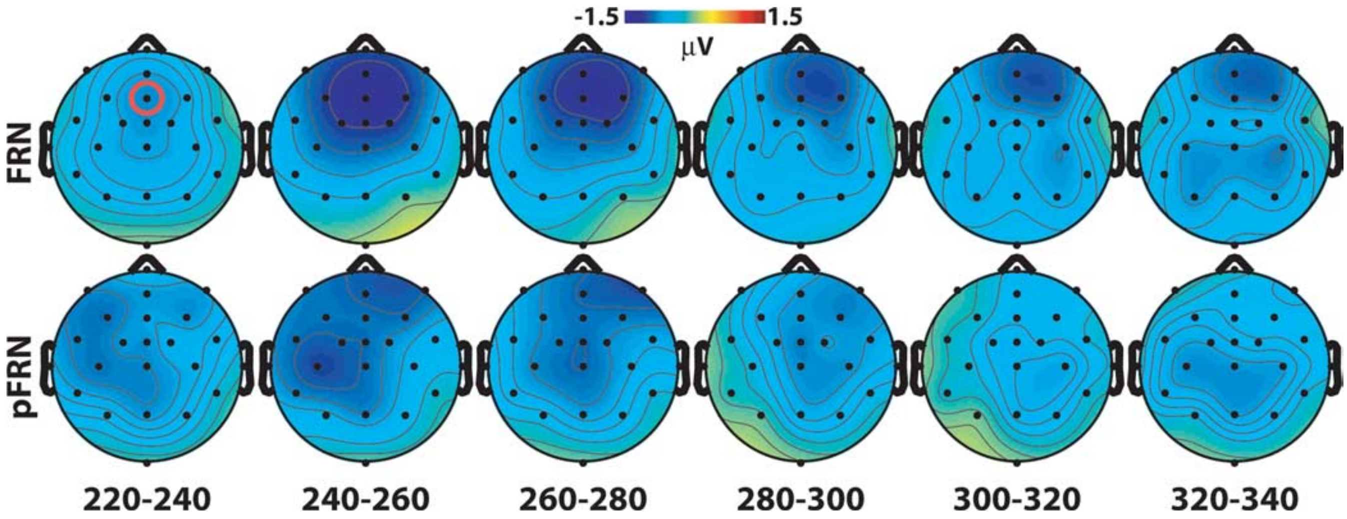

Figure 3. Topographical distributions of FRN (loss–win; top row) and pFRN (loss/opposite − loss/same trials; bottom row) over time. Red circle in top left map shows location of electrode FCz, used for ERP plots in Figure 2 and statistical analyses.

tocentral scalp sites were relatively negative after losses compared with ERPs after wins [repeated-measures one-way (feedback: win or loss) ANOVA at FCz, $F_{(1,14)} = 15.3; p = 0.002$] (Fig. 2*a*,*b*). This ERP difference is similar in timing and topography to the FRN [similar effects are also called the "feedback error-related negativity" (fERN) or "medial frontal negativity"], which may reflect a neural computation of a reward prediction error (Holroyd and Coles, 2002; Yasuda et al., 2004; Frank et al., 2005). If prediction errors signal the need to adjust future behavior, as the model predicts and the behavioral data confirm, feedback-locked ERPs at medial frontal sites should predict adjustments in decision-making on the subsequent trial. To test this hypothesis, we separately averaged ERPs during wins and losses according to the decision that was made in the following trial. As shown in Figure 2, *a* and *b*, ERPs after losses were significantly more negative on trials when subjects chose the opposite target on the following trial (loss/opposite trials) compared with ERPs during losses when subjects chose the same target on the following trial (loss/same trials). This was confirmed by a 2 (outcome: win or loss) × 2 (next trial: same or opposite) repeated-measures ANOVA (main effect of next trial decision, $F_{(1,14)} = 4.75; p = 0.04$). This effect was significant for losses ($F_{(1,14)} = 5.49; p = 0.03$) but not for wins ($F_{(1,14)} = 1.17; p = 0.29$). In other words, loss/opposite trials elicited a larger FRN than did loss/same trials. We refer to the difference between ERPs after loss/opposite and loss/same trials as the pFRN effect (Fig. 3). A LORETA source localization procedure (Pascual-Marqui et al., 1994) estimated overlapping generators of the FRN and pFRN in the dorsal and posterior cingulate cortex (supplemental Fig. S2, available at www.jneurosci.org as supplemental material), consistent with other source estimations of the FRN, as well as with the response-related ERN (Ruchsow et al., 2002; Herrmann et al., 2004; Debener et al., 2005).

We next compared the pFRN with prediction errors generated by the reinforcement learning model. The model played the matching pennies game against the computer opponent in 15 separate sessions and won an average ± SE of 48.8 ± 0.27%. This is near the Nash equilibrium of winning in 50% of trials, which is the optimal behavior given two competitive opponents, and is comparable with our subjects' performance. The prediction errors of the model were negative after losses and positive after wins (thus, it exhibited an FRN). More importantly, prediction errors were larger during loss/opposite trials compared with those during loss/same trials (a pFRN; $t_{(14)} = 26; p < 0.001$) (Fig. 4*a*). These parallels between the prediction errors of the model and human ERPs to decision outcomes (Fig. 4) are consistent with the idea that the pFRN effect reflects the computation of negative prediction errors and that it signals the need to adjust behavior.

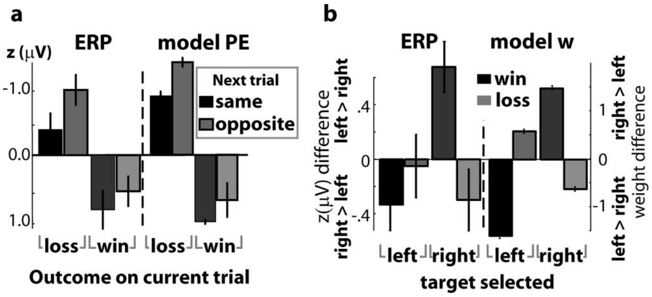

Figure 4. Human ERP responses closely resemble model outputs. ***a***, Prediction errors generated by the reinforcement model parallel human ERP responses. The prediction errors of the model (right) paralleled the feedback-locked ERP responses of humans (left). ERP data are *z*-transformed for ease in comparison with model output. PE, Prediction error. ***b***, Motor cortex ERP amplitudes and changes in model weights are modulated by both the decision made and the feedback received. The *y*-axis depicts *z*-transformed activity differences between electrodes C3 and C4 (left) or the differences in the adjustments of the model in the weights for left and right target (right), separated according to whether the subjects or model chose the left- or right-hand target and whether they lost or won. Negative values indicate relatively more activity over right versus left motor cortex (human ERP) or left versus right weights (model). The parallels between the model weights and motor ERPs suggest that representations of winning responses are strengthened and representations of losing responses are weakened. W, Weight.

How might a neural prediction error signal be used to guide future decisions? In the model, prediction errors are used to strengthen or weaken the weights of the two decision options. Indeed, analysis of changes in the weights in the model after losses and wins showed that weights were strengthened or weakened depending on the type of response that led to the outcome. More specifically, if the right-hand target was selected and led to a loss, the weight for right-hand target was weakened relative to that of the left-hand target, whereas the opposite pattern was apparent for wins. Likewise, if the left-hand target was selected and led to a loss, the weight for left-hand target was weakened relative to that

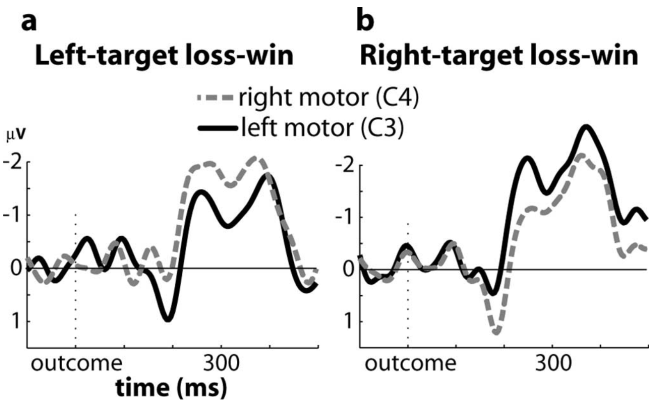

Figure 5. ERP evidence that feedback processing involves adjustments of representations of competing responses. Grand-averaged LFRN effects according to whether subjects chose the left-hand target (***a***) or the right-hand target (***b***). ERPs are shown from motor cortex electrode sites (C3 and C4). The LFRN difference is larger over right motor cortex after left target selections and is larger over left motor cortex after right target selections.

of the right-hand target [2 (target selected: left or right) × 2 (feedback: loss or win) factorial ANOVA; $F_{(1,14)} = 979$; $p < .001$] (Fig. 4*b*, right).

In humans, weights for decision options might correspond to neural representations of actions used to indicate decisions (Schall, 1995; Gold and Shadlen, 2000; Samejima et al., 2005; Schall, 2005), and the strength of these representations might correspond to the relative magnitudes of ERPs measured over right versus left motor cortex (for left-hand vs right-hand responses). Based on this reasoning, we hypothesized that lateral scalp sites over motor cortex should exhibit a sensitivity to loss versus win feedback when that motor cortex was used to indicate the decision. Two complementary analyses support this hypothesis. First, we examined the FRN (i.e., the loss–win ERP difference) at motor cortex sites C3 and C4 as a function of the target selected on each trial. This analysis revealed a significant hand × hemisphere interaction ($F_{(1,14)} = 4.63$; $p = 0.04$), such that the FRN was enhanced over motor cortex electrode sites contralateral to the hand used to make the preceding response (Fig. 5). We refer to this modulation as a lateralized FRN (LFRN) effect. Follow-up analyses showed that this LFRN effect was significantly larger on the left hemisphere than on the right hemisphere on trials with right-hand responses ($F_{(1,14)} = 5.56$; $p = 0.03$). On trials with left-hand responses, this effect was in the expected direction, although not significant ($F_{(1,14)} = 1.12$; $p = 0.28$). Figure 3 illustrates the time course of the enhanced FRN effect at motor cortex electrode sites contralateral to the response hand used. Current source density maps confirmed that these effects were maximal over C3 and C4 (supplemental Fig. S5, available at www.jneurosci.org as supplemental material). The time course of the topographical distribution is displayed in Figure 6.

The previous analysis demonstrated that feedback about decision outcomes modulates ERPs recorded at lateral motor sites. In an additional analysis, we investigated how these feedback signals might modulate activation of motor representations by examining lateralized ERP potentials (e.g., the difference between the potentials recorded at electrode C3 and those at C4) after feedback. In this analysis, positive values indicate that the left motor cortex has a relatively positive potential than that of the right, and negative values indicate that the left motor cortex has a relatively negative potential compared with that of the right. An ANOVA on these motor cortex difference values, with response hand (left or right) and feedback (loss or win) as factors, revealed a significant interaction ($F_{(1,14)} = 5.28$; $p = 0.03$), and inspection of this interaction suggests that representations of winning responses are strengthened whereas representations of losing responses are weakened (Fig. 4*b*, left). As shown in Figure 4*b*, the pattern of feedback-locked ERP effects over motor cortex was similar to the weight changes produced by the computational model after wins and losses.

Discussion

In the present study, we examined ERPs after decision outcomes to test the idea that reinforcement learning signals guide dynamic changes in decision behavior. Our results were consistent with predictions of a computational model, suggesting that neural prediction error signals guide future decisions through the adjustment of competing action representations.

Neural responses to feedback predict adjustments in future behavior

According to a recent theory, the FRN reflects a reward prediction error signal sent from the midbrain dopamine system to the anterior cingulate cortex, in which it is used to adapt behavior

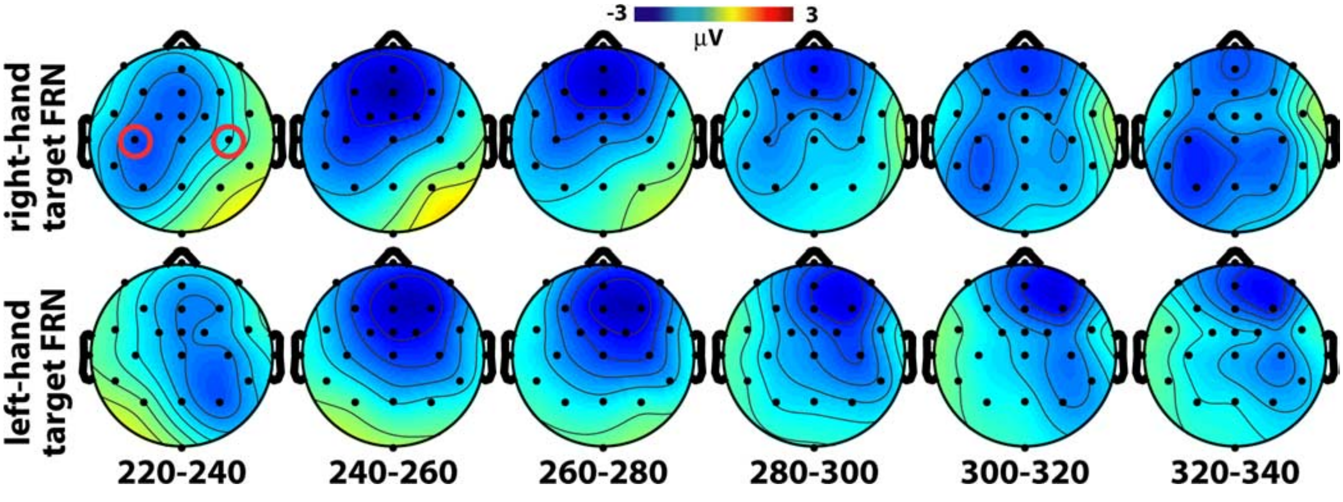

Figure 6. Topographical maps of LFRN effect over time, separated for right-hand (top row) and left-hand (bottom row) responses. Red circles on the top left map show locations of electrodes C3 (left hemisphere) and C4 (right hemisphere), which were used for ERPs in Figure 5 and for statistical analyses.

according to principles of reinforcement learning (Holroyd and Coles, 2002). Given that prediction errors might signal the need to adjust behavior (Ridderinkhof et al., 2004a,b), the FRN should reflect not only whether the current feedback is good or bad but also how behavior is adjusted in the future. Consistent with this idea, we found that ERPs elicited by loss feedback were more negative when subjects chose the opposite versus the same target on the subsequent trial. Several considerations suggest that this pFRN effect is driven by the same neural mechanisms as the FRN: both the pFRN and the FRN have similar topographical distributions, time courses, and estimated neural generators (Figs. 2, 3) (supplemental Fig. S2, available at www.jneurosci.org as supplemental material). Furthermore, the reinforcement learning model uses a single mechanism to produce effects that strongly resembled both the FRN and the pFRN (Fig. 4*a*).

Other studies using different paradigms have reported feedback-locked potentials that have been referred to as the FRN, the fERN, or the medial frontal negativity (Gehring and Willoughby, 2002; Holroyd et al., 2002, 2003). Although there may be functional differences between these effects (Holroyd et al., 2002), they share many functional characteristics and can be distinguished from later-occurring modulations of ERP components such as the P300 complex or error positivity (Nieuwenhuis et al., 2001; Hajcak et al., 2005). Several factors suggest that the pFRN effect may be a modulation of the FRN rather than the P300. First, the peaks of the FRN and pFRN effects overlap in time and occur well before the expected time range of the P300 peak (250 ms in our analyses vs 300–400 ms for a typical P300 peak) (Fig. 2*a*). Second, both the FRN and the pFRN effects have an anterior scalp topography, whereas the P300 typically has a more posterior topography with a spatial peak at Pz.

Other ERP studies have related the magnitude of the FRN to overall learning or decision-making strategies, although these were not on the trial-by-trial level (Yeung and Sanfey, 2004; Frank, 2005; Hewig et al., 2006). Additionally, some studies have shown that, during speeded reaction time tasks (when errors are common), ERPs after the response predict increases in reaction time on the subsequent trial (Gehring et al., 1993; Garavan et al., 2002; Ridderinkhof et al., 2003; Debener et al., 2005). However, ERPs are not correlated with subsequent reaction time adjustments in every study (Gehring and Fencsik, 2001), and, in the present study, the pFRN was unrelated to reaction times on subsequent trials (one-way ANOVA, $F_{(1,14)} < 1$). This is probably because our study did not require speeded responses, and so reaction times were not relevant to task performance. It is likely that the FRN/ERN signals prediction errors, and the impact of prediction errors on future behavior will vary across studies, depending on which specific behaviors are task relevant (Fiehler et al., 2005).

Although the neural generator(s) of the FRN remain somewhat debated, accumulating evidence from ERP source localization studies (including our own) (supplemental Fig. S2, available at www.jneurosci.org as supplemental material) suggests that the anterior cingulate or surrounding medial frontal cortex is a likely generator (Ruchsow et al., 2002; Herrmann et al., 2004; Debener et al., 2005; Wang et al., 2005; Taylor et al., 2006). Consistent with these source estimations, and with our finding that the pFRN predicts adjustments in decision-making, a recent study showed that cingulate lesions in monkeys impair the ability to use previous reinforcements to guide decision-making behavior (Kennerley et al., 2006).

Neural responses to feedback reflect adjustment of competing action representations

According to many reinforcement learning models, prediction error signals can guide decision-making by modulating the strength of weights for competing decision options (Barto, 1995; Egelman et al., 1998; Braver and Brown, 2003). Consistent with these models, single-unit recording studies of monkeys have shown that activity in specific response-related neurons is modulated by expected reward (Schall, 1995, 2005; Gold and Shadlen, 2000; Sugrue et al., 2004). The present results provide the first evidence, to our knowledge, to suggest that humans might engage a similar mechanism. We found that FRN-like responses over motor cortex electrode sites were lateralized according to the response hand that was used to make the decision (the LFRN). This finding suggests that feedback information may be processed in motor cortical regions, such that representations of winning responses are strengthened, whereas representations of losing responses are weakened. Scalp-recorded ERPs lack the spatial precision to determine whether the LFRN was driven by activity in the hand areas of motor cortex. However, C3 and C4 are commonly used to study motor responses (Mordkoff and Gianaros, 2000; Galdo-Alvarez and Carrillo-de-la-Pena, 2004; Carrillo-de-la-Pena et al., 2006), and current source density maps confirmed that these effects are maximal over C3 and C4 (supplemental Fig. S5, available at www.jneurosci.org as supplemental material). Future research using fMRI could be used to more precisely localize the generators of this effect.

Importantly, the LFRN effect was observed after feedback, which was 1400 ms after the response made on that trial and ~2000 ms before the response made on the following trial. Consequently, it is highly unlikely that the LFRN was driven by overt motor responses. Furthermore, these results could not be attributable to differences in horizontal eye movements because the horizontal electrooculogram channels showed no "FRN-like" effect (supplemental information, available at www.jneurosci.org as supplemental material). It is possible that subjects maintained a memory trace of the previous motor response throughout the delay from the decision to the feedback (i.e., they kept active a representation of the right motor response during trials in which they selected the right-hand target), but such a difference in baseline activity levels could not explain the LFRN, because it was impossible for subjects to predict the feedback before receiving it. Indeed, whether motor cortex was active before feedback per se was not of interest but rather the change in motor cortex activity as a function of the decision outcome. We suggest that the LFRN reflects a process by which prediction error signals are used to adjust representations of competing actions associated with different decision options. This interpretation is supported by the parallel pattern of results obtained from the reinforcement learning model when it was subjected to the same analyses (Fig. 4).

Finally, we note that the response-lateralized FRN observed here need not accompany the FRN under all circumstances. Indeed, it is possible to obtain an FRN when no responses are required (Donkers et al., 2005; Yeung et al., 2005). It is likely that the LFRN would be observed only when lateralized responses are required.

Neural mechanisms of reinforcement-guided decision-making

Several lines of evidence suggest that calculations of prediction errors are expressed through dynamic changes in midbrain dopamine neuron activity (Schultz et al., 1997; Waelti et al., 2001). Interestingly, larger or longer dopamine dips follow larger violations of expected reward (Fiorillo et al., 2003) and thus might indicate a larger prediction error. Midbrain dopamine neurons can directly modulate activity of pyramidal cells in the cingulate cortex (Gariano and Groves, 1988; Williams and Goldman-Rakic, 1998; Onn and Wang, 2005) and may transmit prediction error signals through this connection. Cingulate neurons can also modulate activity in the striatum and midbrain (Eblen and Graybiel, 1995; Joel and Weiner, 2000), so it is possible that prediction error signals might be calculated in the cingulate and transmitted to midbrain dopamine regions. Future research using simultaneous recordings from the medial frontal cortex and midbrain may shed light into whether prediction errors are first signaled by cortical or subcortical areas.

Dopaminergic prediction error signals might guide adjustments in action representations through modulation of the basal ganglia–thalamic–cortical motor loop (Alexander and Crutcher, 1990; Orieux et al., 2002; Frank, 2005). Specifically, the globus pallidus may gate activations of motor commands in the thalamus (Frank, 2005). Phasic bursts or dips of dopamine modulate the gating mechanism of this system over the thalamus and thus may allow cortical representations of actions (e.g., left- or right-hand responses) to be strengthened or weakened (Alexander et al., 1986; Gurney et al., 2001; Frank, 2005). Thus, the LFRN effect observed here might reflect adjustments of motor response representations induced by phasic modulations of the thalamic–pallidal–cortical motor system.

References

Alexander GE, Crutcher MD (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci 13:266–271.

Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci 9:357–381.

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

Barto AG (1995) Reinforcement learning. In: Handbook of brain theory and neural networks (Arbib MA, ed), pp 804–809. Cambridge, MA: MIT.

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129–141.

Braver TS, Brown JW (2003) Principles of pleasure prediction: specifying the neural dynamics of human reward learning. Neuron 38:150–152.

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

Carrillo-de-la-Pena MT, Lastra-Barreira C, Galdo-Alvarez S (2006) Limb (hand vs. foot) and response conflict have similar effects on event-related potentials (ERPs) recorded during motor imagery and overt execution. Eur J Neurosci 24:635–643.

Cohen MX (2006) Individual differences and the neural representations of reward expectation and reward prediction error. Soc Cogn Affect Neurosci, in press.

Cohen MX, Ranganath C (2005) Behavioral and neural predictors of upcoming decisions. Cogn Affect Behav Neurosci 5:117–126.

Dayan P, Balleine BW (2002) Reward, motivation, and reinforcement learning. Neuron 36:285–298.

Debener S, Ullsperger M, Siegel M, Fiehler K, von Cramon DY, Engel AK (2005) Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J Neurosci 25:11730–11737.

Donkers FC, Nieuwenhuis S, van Boxtel GJ (2005) Mediofrontal negativities in the absence of responding. Brain Res Cogn Brain Res 25:777–787.

Eblen F, Graybiel AM (1995) Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. J Neurosci 15:5999–6013.

Egelman DM, Person C, Montague PR (1998) A computational role for dopamine delivery in human decision-making. J Cogn Neurosci 10:623–630.

Fiehler K, Ullsperger M, von Cramon DY (2005) Electrophysiological correlates of error correction. Psychophysiology 42:72–82.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898–1902.

Frank MJ (2005) Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cogn Neurosci 17:51–72.

Frank MJ, Woroch BS, Curran T (2005) Error-related negativity predicts reinforcement learning and conflict biases. Neuron 47:495–501.

Galdo-Alvarez S, Carrillo-de-la-Pena MT (2004) ERP evidence of MI activation without motor response execution. NeuroReport 15:2067–2070.

Garavan H, Ross TJ, Murphy K, Roche RA, Stein EA (2002) Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. NeuroImage 17:1820–1829.

Gariano RF, Groves PM (1988) Burst firing induced in midbrain dopamine neurons by stimulation of the medial prefrontal and anterior cingulate cortices. Brain Res 462:194–198.

Gehring WJ, Fencsik DE (2001) Functions of the medial frontal cortex in the processing of conflict and errors. J Neurosci 21:9430–9437.

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

Gehring WJ, Goss B, Coles MG, Meyer DE, Donchin E (1993) A neural system for error detection and compensation. Psychol Sci 4:385–390.

Gold JI, Shadlen MN (2000) Representation of a perceptual decision in developing oculomotor commands. Nature 404:390–394.

Gurney K, Prescott TJ, Redgrave P (2001) A computational model of action selection in the basal ganglia. I. A new functional anatomy. Biol Cybern 84:401–410.

Hajcak G, Holroyd CB, Moser JS, Simons RF (2005) Brain potentials associated with expected and unexpected good and bad outcomes. Psychophysiology 42:161–170.

Herrmann MJ, Rommler J, Ehlis AC, Heidrich A, Fallgatter AJ (2004) Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). Brain Res Cogn Brain Res 20:294–299.

Hewig J, Trippe R, Hecht H, Coles MG, Holroyd CB, Miltner WH (2006) Decision-making in blackjack: an electrophysiological analysis. Cereb Cortex, in press.

Holroyd CB, Coles MG (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol Rev 109:679–709.

Holroyd CB, Coles MG, Nieuwenhuis S (2002) Medial prefrontal cortex and error potentials. Science 296:1610–1611; author reply 1610–1611.

Holroyd CB, Nieuwenhuis S, Yeung N, Cohen JD (2003) Errors in reward prediction are reflected in the event-related brain potential. NeuroReport 14:2481–2484.

Joel D, Weiner I (2000) The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. Neuroscience 96:451–474.

Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ (2000) Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37:163–178.

Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9:940–947.

Lagarias JC, Reeds JA, Wright MH, Wright PE (1998) Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J Optim 9:112–147.

Li Y, Ma Z, Lu W (2006) Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach. Physiol Meas 27:425–436.

Luce DP (1999) Individual choice behavior. New York: Wiley.

Montague PR, Berns GS (2002) Neural economics and the biological substrates of valuation. Neuron 36:265–284.

Montague PR, Hyman SE, Cohen JD (2004) Computational roles for dopamine in behavioural control. Nature 431:760–767.

Mookherjee D, Sopher B (1994) Learning behavior in an experimental matching pennies game. Games Econ Behav 7:62–91.

Mordkoff JT, Gianaros PJ (2000) Detecting the onset of the lateralized readiness potential: a comparison of available methods and procedures. Psychophysiology 37:347–360.

Nieuwenhuis S, Ridderinkhof KR, Blom J, Band GP, Kok A (2001) Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. Psychophysiology 38:752–760.

Nieuwenhuis S, Ridderinkhof KR, Talsma D, Coles MG, Holroyd CB, Kok A, van der Molen MW (2002) A computational account of altered error processing in older age: dopamine and the error-related negativity. Cogn Affect Behav Neurosci 2:19–36.

Nieuwenhuis S, Holroyd CB, Mol N, Coles MG (2004) Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. Neurosci Biobehav Rev 28:441–448.

Onn SP, Wang XB (2005) Differential modulation of anterior cingulate cortical activity by afferents from ventral tegmental area and mediodorsal thalamus. Eur J Neurosci 21:2975–2992.

Orieux G, Francois C, Feger J, Hirsch EC (2002) Consequences of dopaminergic denervation on the metabolic activity of the cortical neurons projecting to the subthalamic nucleus in the rat. J Neurosci 22:8762–8770.

Pascual-Marqui RD, Michel CM, Lehmann D (1994) Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int J Psychophysiol 18:49–65.

Ridderinkhof KR, Nieuwenhuis S, Bashore TR (2003) Errors are foreshadowed in brain potentials associated with action monitoring in cingulate cortex in humans. Neurosci Lett 348:1–4.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004a) The role of the medial frontal cortex in cognitive control. Science 306:443–447.

Ridderinkhof KR, van den Wildenberg WP, Segalowitz SJ, Carter CS (2004b) Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. Brain Cogn 56:129–140.

Ruchsow M, Grothe J, Spitzer M, Kiefer M (2002) Human anterior cingulate cortex is activated by negative feedback: evidence from event-related potentials in a guessing task. Neurosci Lett 325:203–206.

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340.

Schall JD (1995) Neural basis of saccade target selection. Rev Neurosci 6:63–85.

Schall JD (2005) Decision making. Curr Biol 15:R9–R11.

Schultz W (2004) Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. Curr Opin Neurobiol 14:139–147.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599.

Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. Science 304:1782–1787.

Sutton RS (1992) Introduction: the challenge of reinforcement learning. Mach Learn 8:225–227.

Sutton RS, Barto AG (1998) Reinforcement learning. Cambridge, MA: MIT.

Taylor SF, Martis B, Fitzgerald KD, Welsh RC, Abelson JL, Liberzon I, Himle JA, Gehring WJ (2006) Medial frontal cortex activity and loss-related responses to errors. J Neurosci 26:4063–4070.

Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412:43–48.

Wang C, Ulbert I, Schomer DL, Marinkovic K, Halgren E (2005) Responses of human anterior cingulate cortex microdomains to error detection, conflict monitoring, stimulus-response mapping, familiarity, and orienting. J Neurosci 25:604–613.

Williams SM, Goldman-Rakic PS (1998) Widespread origin of the primate mesofrontal dopamine system. Cereb Cortex 8:321–345.

Yasuda A, Sato A, Miyawaki K, Kumano H, Kuboki T (2004) Error-related negativity reflects detection of negative reward prediction error. NeuroReport 15:2561–2565.

Yeung N, Sanfey AG (2004) Independent coding of reward magnitude and valence in the human brain. J Neurosci 24:6258–6264.

Yeung N, Holroyd CB, Cohen JD (2005) ERP correlates of feedback and reward processing in the presence and absence of response choice. Cereb Cortex 15:535–544.