PNAS

Electrophysiological correlates of reward prediction error recorded in the human prefrontal cortex

Hiroyuki Oya*, Ralph Adolphs*†‡§, Hiroto Kawasaki*, Antoine Bechara†, Antonio Damasio†, and Matthew A. Howard III*

Departments of *Neurosurgery and †Neurology, University of Iowa College of Medicine, Iowa City, IA 52242; and ‡Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125

Lesion and functional imaging studies have shown that the ventromedial prefrontal cortex is critically involved in the avoidance of risky choices. However, detailed descriptions of the mechanisms that underlie the establishment of such behaviors remain elusive, due in part to the spatial and temporal limitations of available research techniques. We investigated this issue by recording directly from prefrontal depth electrodes in a rare neurosurgical patient while he performed the Iowa Gambling Task, and we concurrently measured behavioral, autonomic, and electrophysiological responses. We found a robust alpha-band component of event-related potentials that reflected the mismatch between expected outcomes and actual outcomes in the task, correlating closely with the reward-related error obtained from a reinforcement learning model of the patient's choice behavior. The finding implicates this brain region in the acquisition of choice bias by means of a continuous updating of expectations about reward and punishment.

decision-making | emotion

Ventral and medial sectors of the prefrontal cortex (PFC) have been implicated in guiding behavior on the basis of the motivational value of the choices available (1–3). Damage to this region in monkeys impairs reversal learning and attenuates reinforcer devaluation effects (4, 5). Analogous lesions in humans result in behavior that is guided by the immediate rewarding or punishing properties of stimuli rather than by their prospective future contingencies (6). These findings may also be broadly related to functions of the adjacent anterior cingulate cortex in error (conflict) monitoring and response selection (7–9).

Neural activity within the human ventral and medial PFC is widely thought to track the incentive value of stimuli rather than their sensory properties (10–15). Recent studies have found activation both in expectation of monetary gain or loss (16, 17) and to prediction errors in appetitive conditioning (18). Whereas the human data have come from lesion and imaging studies, studies in monkeys have focused on electrophysiology, with similar findings (19, 20). Here we combined the rare opportunity to record directly from the PFC of a neurosurgical patient with the administration of a widely used, computerized task that probes decision-making under risk: the Iowa Gambling Task (21). In addition to the benefit of more direct comparisons between human and monkey studies, such an approach circumvents some of the limitations inherent in noninvasive studies. Lesion studies are limited with respect to their spatial and temporal resolution; functional MRI studies using blood oxygenation level-dependent imaging have limited temporal resolution and suffer from paramagnetic signal drop-out within ventral PFC; and scalp event-related potentials (ERP) studies suffer from poor localization and attenuated signal from more medial regions.

Patients with damage to the ventral and medial PFC fail to learn from negative monetary feedback in the Iowa Gambling Task (22), as they fail in real life to adjust their behavior on the basis of the mismatch between expectations and choice outcomes. We thus hypothesized that evoked potentials within the medial PFC should carry information about the reinforcement learning that occurs during this task. To obtain a detailed picture of the relationships between reward contingency, behavior, emotional response, and electrophysiology, we concurrently recorded behavioral choice, autonomic response [as measured by skin conductance response (SCR)], and field potentials from depth electrodes in a neurosurgical patient undergoing monitoring for medically intractable epilepsy.

Methods

Subject. Our participant was a 48-year-old left-handed man with a diagnosis of medically intractable epilepsy (simple partial seizures) who had depth electrodes implanted for invasive monitoring of his epilepsy. There was no evidence on magnetic resonance scans of any structural abnormality in the PFC, and his seizure focus was later localized distant from the regions in which we recorded (in the right premotor cortex). The subject performed normally on background neuropsychological tests of intelligence quotient (IQ) (total IQ was 110 accessed with the Wechsler Adult Intelligence Scale III), language (left language dominance was accessed with Wada testing, and he performed normally on aphasia and naming batteries), verbal and nonverbal memory (normal performances on the Auditory Verbal Learning Task, the Benton Visual Recognition Task, the Rey–Osterrieth Complex Figure Recall, and the Wechsler Memory Scale), executive function (assessed with the Trailmaking Test), and visual perception (assessed with the Benton Facial Recognition test). He had no history of psychiatric disease or any neurological illness other than his epilepsy (see *Supporting Methods*, which is published as supporting information on the PNAS web site).

Iowa Gambling Task. In the standard administration of the Iowa Gambling Task, we showed our patient four decks of cards (A, B, C, and D) on a computer monitor and gave him a $2,000 credit of play money to start. He used a computer mouse to select cards from the decks in 100 trials. The objective in the task is to make as much money in the long run as possible. After each choice, the computer provided two types of feedback before proceeding to the next trial: first, every deck produced a variable reward; second, an occasional variable punishment was obtained. Neither the reward/punishment contingencies of the cards nor the total number of trials was known to the subject. The reward was produced immediately at the time the subject chose a card, whereas the punishment was produced 3 sec later (Fig. 1); on trials where no punishment was obtained, a text screen with the

NEUROSCIENCE
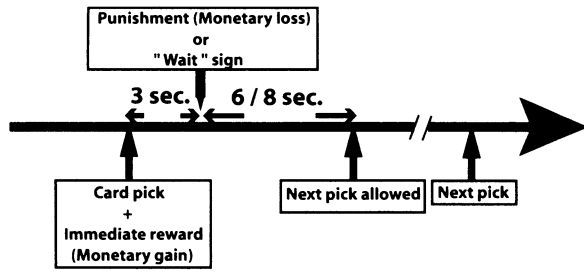
Fig. 1. Time course of the Iowa Gambling Task. The reward was produced immediately at the time the subject chose a card, whereas delayed feedback (either monetary loss or a "Wait" sign indicating an intertrial interval) was produced 3 sec after the choice. Whereas a minimum intertrial interval of 6–8 sec was interposed, the subject could take as long as he wished to deliberate the next choice; the resulting median intertrial-interval time was 11.3 sec. The immediate reward was produced in all trials, whereas the delayed feedback (punishment) occurred randomly on only some of the trials.

words "please wait" appeared after 3 sec, thus ensuring that the period after reward while the subject was waiting for a possible punishment was the same duration in all trials. A mandatory intertrial interval (pause) of 6–8 sec was interposed between all trials (so that it was impossible to pick cards faster than dictated by this interval constraint).

Two of the decks (the "safe" decks: C and D) feature relatively low monetary gains and occasional low losses with a net gain over time; the other two decks (the "risky" decks: A and B) feature larger monetary gains with occasional very large losses and a net loss over time. The mean initial reward per trial was $107.30 for deck A, $116.90 for deck B, $54.90 for deck C, and $52.50 for deck D; and the mean subsequent monetary loss was $149.50 for deck A, $163.30 for deck B, $26.30 for deck C, and $21.20 for deck D. The probability of obtaining a punishment (the subsequent monetary loss) on a given trial was 0.58 for deck A, 0.10 for deck B, 0.59 for deck C, and 0.08 for deck D. The total time taken by the subject to complete the task was 18.5 min. The experiments were carried out 5 days after implantation, when medications had been tapered (see *Supporting Methods*), and the subject had recovered from his surgery and was awake and alert. The subject did not have any seizure at least 12 h preceding the experiments.

The protocol was approved by the University of Iowa Institutional Review Board (Iowa City) and the subject's written consent was obtained.
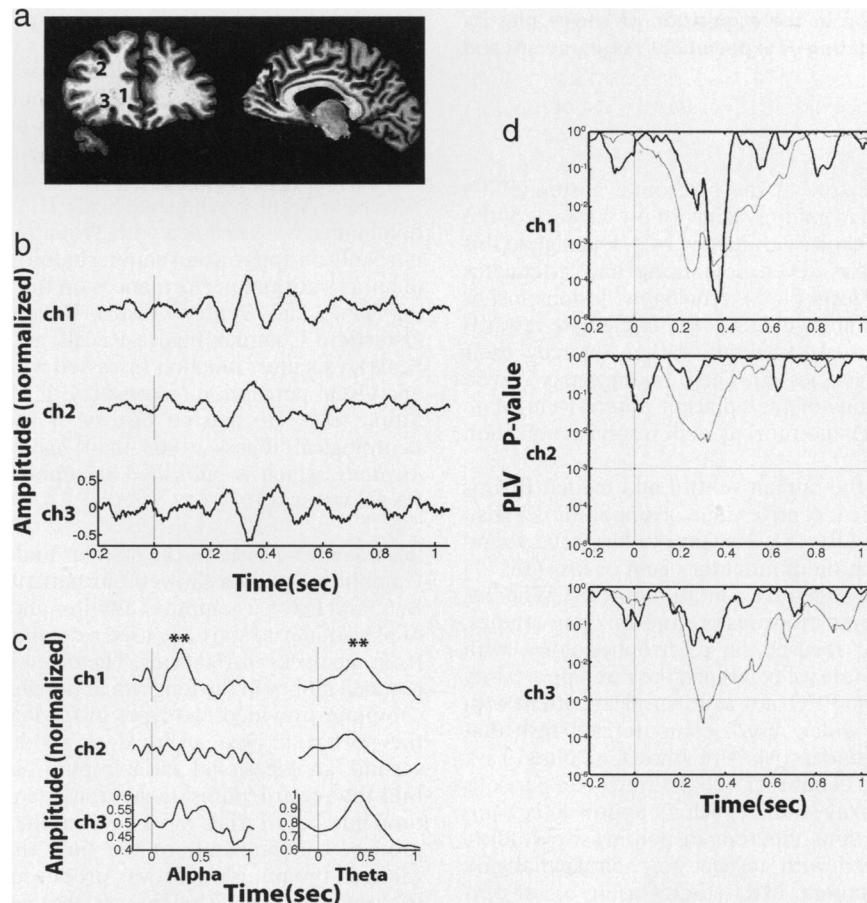
Fig. 2. Recording location and ERP response. (*a*) (*Left*) Recording sites (indicated by numbers on the figure) were mapped from postimplantation scans onto the corresponding locations in preimplantation scans (arrows) by using BRAINVOX (44). (*Right*) The sagittal MRI shows the projection of the most medial recording site (contact 1) onto the sagittal plane. (*b*) Averaged ERPs across all trials (*n* = 91 trials after artifact rejection) for each of the recording contacts shown in *a* (ch1, channel 1, etc.). Amplitude was linearly rescaled (see *Methods*). Vertical lines at 0 represent the onset of the feedback given 3 sec after card pick. (*c*) Averaged instantaneous amplitude at each recording site. Instantaneous amplitude was calculated from the Hilbert transform of the reconstructed waveform of D5 (alpha) and D6 (theta) decomposition levels (see *Methods*). **, $P < 0.01$ increase in amplitude. (*d*) Phase-locking value (PLV) for each decomposition level at the three recording sites. Black line, the alpha band; gray line, the theta band; and horizontal thin black lines, $P < 0.001$, assessed with Rayleigh's test. PLV represents the degree of phase concentration at particular time points in certain frequency bands across trials.

Electrode Implantation, Localization, and Electroencephalogram Recording. Two hybrid clinical-research depth electrodes (23) were inserted while the subject was under general anesthesia by using a CRW stereotactic system (Radionics, Burlington, MA). T1-weighted structural magnetic resonance scans of the whole brain were obtained both pre- and post-implantation and permitted mapping of the location of recording sites seen in the postimplantation scan onto the corresponding location of the preimplantation scan. Recording site 1 was located in area 10r, the granular paracingulate cortex, corresponding to the rostral part of monkey area 10m (24). Recording sites 2 and 3 were located in the right middle frontal gyrus and area 11l, respectively (Fig. 2a).

Continuous electroencephalogram data were acquired with bipolar contacts (separation: 200 μm, impedance 90–200 kΩ at 1 kHz). Filtered signals (2–6 kHz) were amplified ($\times$5,000) and stored for offline processing. Data were originally sampled at 20 kHz and decimated to 500 Hz, thresholded to discard trials containing artifacts (nine rejected trials), and normalized so that their rms values were set to 1.

SCR Recording. Electrodermal activity was continuously recorded in dc mode from the subject's nondominant hand by using a MP-150 system (Biopac Systems, Goleta, CA) and sampled at 10 Hz. Raw waveforms were detrended by using exponential smoothing, and SCR was calculated as the difference between the raw signal and the low-frequency tonic trend component (skin conductance level). Anticipatory SCR and postfeedback SCR values for each trial were determined, respectively, as average values of SCR(t) during the 5 sec immediately before card choice and the 5 sec between 1 and 6 sec after the punishment or "wait" cue.

Reinforcement Learning Modeling. We modeled the subject's card choices with a modified reinforcement learning algorithm (25–27) that predicts the best choice to be made on the basis of the statistical distribution of the winning and losing contingencies experienced on the four card decks. Our modeling of the subject's choices incorporated a trial-by-trial update of the action selection probability (p, the probability of selecting from a particular card deck), from estimations of parameters that specify an update value of action values for each possible action (Q, the value associated with choosing from a particular card deck), and the degree of exploration or exploitation (26, 27). This model assumes that the subject maintains expectations for each of the action values obtained after choosing from one of the four decks and updates the probability of choosing from a particular deck on the next trial by comparing action values of all possible actions at the trial. An indicator of the subject's learning in the task is the average value of all of the possible actions (expected value, V; the mean action values of all four decks weighted by the probability of choosing from them). We note that this does not have the same meaning as "state value" in actor–critic or temporal difference learning schemes as in ref. 25 because we modeled the task as a static action choice. Rather, it reflects the overall utility of the subject's prospect, as formulated in prospect theory (28, 29). As with the well-known Delta and Rescorla–Wagner rules, it is the error in reward (or punishment) prediction that plays a crucial role in updating the old estimate (the weights in the model). This error yielded the subject's reward prediction error (PE), corresponding to the difference between expected and obtained reward.

Multiresolution ERP Analysis. Frequency decomposition of the ERP waveform was used to extract components that overlap in
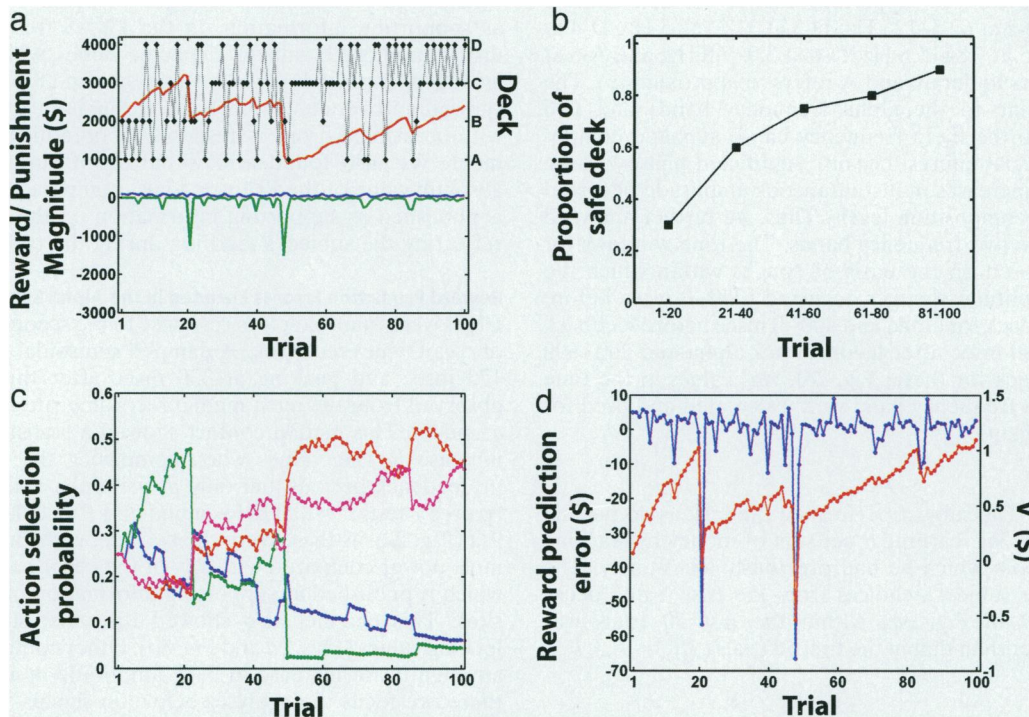
Fig. 3. Behavioral performance on the Iowa Gambling Task. (a) Card choice from one of the four decks and monetary gain/loss as a function of serial order of the 100 trials (x axis). Red curves, the total amount of money possessed at each trial point; blue and green curves, the amount won and lost, respectively (left y axis); black diamonds indicate the subject's choice from one of the decks (right y axis label). (b) The proportion of choices made from the two safe decks (decks C and D) is plotted over time, in successive 20-trial blocks. The subject learned to avoid risky decks during the task, as is typically observed in healthy subjects (43). (c) Probabilities for selecting each deck at the trial are plotted as a function of trial number (x axis). Blue, deck A; green, deck B; red, deck C; and magenta, deck D. The subject correctly learned that decks C and D are advantageous, as represented by larger action selection probabilities, especially in later trials. (d) Blue line, PE values (left y axis); and red line, expected values (V) (right y axis) as a function of trial number. The trend of expected values increased in addition to occasional large drops in response to getting large punishments.
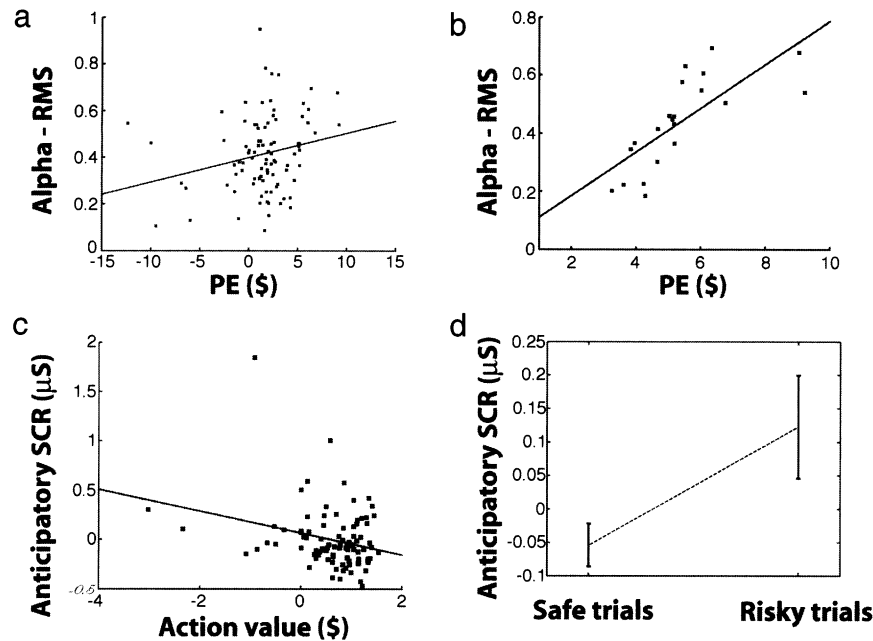
Fig. 4. Correlations with reinforcement learning parameters. (a) Scatter plot showing correlation between reward prediction error (PE, x axis) and alpha band rms values (y axis). For all trials n = 91, after trial rejection. (b) Scatter plot showing the same correlation as in A, but restricted to trials from risky decks (decks A and B) on which no punishment occurred (n = 21). (c) Scatter plot showing correlation between anticipatory SCR level (5–0 sec before card pick) and action values for the choices made (x axis) (n = 91). Low action values represent risky choices. (d) Difference of average anticipatory SCR level between risky and safe trials. Error bar represents mean and ±1 SEM.

time and might otherwise be obscured by averaging. We used a discrete wavelet transform that yielded six commonly used frequency-band levels (see *Supporting Methods* for details): D-1 at 125–250 Hz; D-2 at 62.5–125 Hz; D-3 at 31.3–62.5 Hz; D-4 at 15.6–31.3 Hz; D-5 at 7.8–15.6 Hz; D-6 at 3.9–7.8 Hz; and A-6 at 0–3.9 Hz (D refers to detail, and A refers to approximate). The D-5 (corresponding to the alpha frequency band) and D-6 (corresponding to the theta frequency band) subband components were of special interest because significant phase-locking across trials and increases in instantaneous amplitude occurred solely in these decomposition levels. Thus, we further analyzed responses in these two frequency bands. The time windows for analysis were chosen on the basis of epochs within which the largest mean amplitude changes occurred (200–0 msec before punishment feedback for alpha and 400–0 msec before feedback for theta; 200–400 msec after feedback for alpha and 200–600 msec after feedback for theta; Fig. 2c). rms values in the time windows for each frequency band were calculated and used for statistical assessment.

Results

Behavioral Choice. The subject performed the task as do normal subjects (21) (Fig. 3a), earning a net sum of money by learning to avoid decks from which he had previously lost money. The proportion of the subject's choices from the two "safe" decks (decks C and D in Fig. 3a) within the last 50 trials was significantly higher than that in the first 50 trials ($\chi^2_{(1)}$ = 16.3, $P <$ 0.001) (Fig. 3b).

Reinforcement Learning Model. Estimates of the parameters we used in our model were obtained from the subject's task performance and showed that the subject performed the task with a sufficiently large sensitivity to punishment and long memory for action value updates (see *Supporting Methods* for details). The time course of action selection probabilities showed that the subject learned to discriminate risky from safe decks (Fig. 3c). The estimated reward prediction error (PE) values and

expected values (V) are plotted in Fig. 3d as a function of trial number. Large negative PE values were observed in the trials on which large punishments were given (Fig. 6, which is published as supporting information on the PNAS web site, shows the distribution of PE values). Expected values, which represent the averaged expected reward value obtained on a given trial, showed an overall slow increase throughout the task, together with occasional drops in response to obtaining a large punishment. We also found a slow but significant decrease in the absolute value of the PE over trials (compare with Fig. 7, which is published as supporting information on the PNAS web site) reflecting the subject's learning during the task.

Reward Prediction Error Is Encoded in the Alpha Subband Component. ERPs were found solely in response to the second (monetary loss or "wait" cue) feedback. A damped sinusoidal ERP, starting at 170 msec and peaking at 330 msec after this feedback, was observed from the most medial recording site (contact 1; Fig. 2 a and b). This medial contact showed a statistically significant increase in rms values when comparing the postfeedback to prefeedback periods and only in the alpha and theta subbands (paired t test: P = 0.002 for alpha and P = 0.003 for theta, n = 91) (Fig. 2c), as these were the two subbands that contained the most power compared with any of the other bands (see Fig. 8a, which is published as supporting information on the PNAS web site). This channel also showed the most significant phase-locking values (Figs. 2d and Fig. 8b). Other contacts did not show any significant increase in ERP amplitude in any subband. We therefore focus the analyses below on signals in the alpha and theta bands recorded at this location.

We next examined the relationship between the physiological responses we recorded at contact 1 and the reward PE and other variables defined by the reinforcement learning model. There was a weak but significant linear correlation between ERP amplitude and PE only in the alpha frequency band (Pearson correlation: r = 0.226; P = 0.031; n = 91) (Fig. 4a). We carried out a further analysis on only those trials in which the subject had
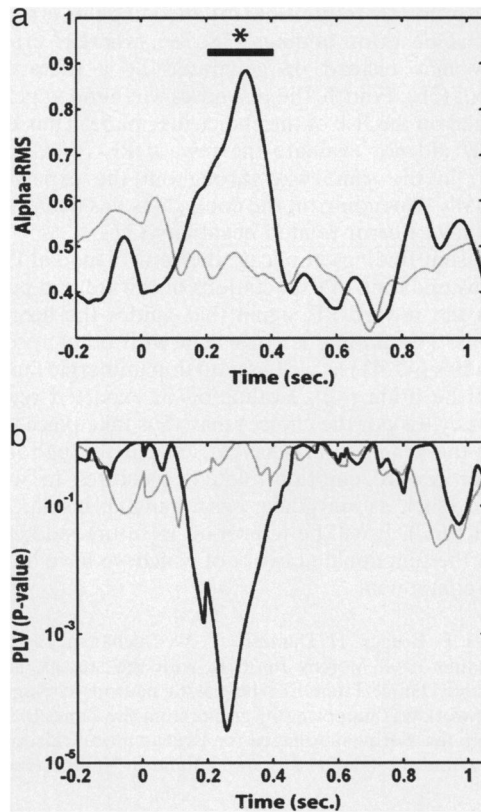
Fig. 5. Alpha-band ERPs sorted by PE. (a) Averaged alpha-band amplitude for trials from risky decks on which no punishment was actually obtained, divided into two groups according to PE values. Black line, those 50% of trials with the highest PE values in risky trials followed by no punishment ($n = 10$); gray line, those 50% of trials with the lowest PE values ($n = 10$); 0 on the x axis represents onset of the second feedback in the task. *, $P = 0.001$ when comparing high-PE trials to low-PE trials (mean rms values in the time-window between 200 and 400 msec after onset of second feedback, Mann–Whitney test, $U = 7.00$) (b) Phase of alpha-band responses. Trials were divided into high-PE trials and low-PE trials as above. y axis represents P value of phase-locking value in log scale. Low P values indicate that phases were concentrated in a specific direction.

made a choice from one of the risky decks (decks A or B), but had in fact not been given any punishment (i.e., trials from risky decks that did not result in any monetary loss; see Fig. 9a, which is published as supporting information on the PNAS web site, for data from all trials on the risky decks). For these trials, alpha-band ERP amplitude was highly correlated with PE values ($r = 0.747$; $P < 0.001$; $n = 21$) (Fig. 4b), demonstrating that the correlation is not driven solely by the actual administration of punishment, but rather by the prediction error associated with it (expecting such punishment, but not receiving it). Similarly, both ERP amplitude and phase-locking in the alpha band showed a striking difference between trials with large PE and trials with small PE (Fig. 5). By contrast, there was no significant correlation between alpha-band ERP amplitude and punishment magnitude itself (Spearman's $\rho = -0.118$; $P = 0.26$; $n = 91$), or reward magnitude itself ($\rho = 0.076$; $P = 0.47$; $n = 91$).

We did not find a correlation between alpha-band ERP amplitude and PE on trials in which the subject chose from the safe decks ($r = 0.035$; $P > 0.78$; $n = 64$, see Fig. 9b), showing that the physiological responses we recorded were not simply driven by monetary gain in the absence of punishment. Similarly, there was no significant relationship between monetary gain and alpha-band ERP amplitude across all trials on which no punishment occurred ($\rho = 0.212$; $P > 0.11$; $n = 57$). Rather, the

alpha-band ERP appears to encode the subject's reward PE, conditional on the risk of the choice taken at the trial.

We further examined the relationship between ERP amplitude and action-values (Q values). Low Q values correspond to choices that the model discourages, whereas high Q values correspond to choices that the model would encourage. Action-values were significantly higher when the subject chose a card from one of the safe decks than from one of the risky decks [t test: $t(89) = 6.20$; $P < 0.001$; $n = 91$; see Fig. 10a, which is published as supporting information on the PNAS web site], thus discriminating the differential risk associated with the decks. We carried out correlation analyses on those trials in which action values were ≤ 0 (no expected monetary gain), as well as on the trials in which the action values were >1 (an expected monetary gain). ERP amplitude and PE were significantly correlated for low action-value trials ($r = 0.648$; $P = 0.017$; $n = 13$), but not for large action-value trials ($r = -0.141$; $P > 0.4$, $n = 31$; see Fig. 10 b and c). The majority of low action-value trials consisted of choices from risky decks ($11/13 = 84.6\%$), whereas large action-value trials were from safe decks ($27/31 = 87\%$). For more details on the model see Fig. 11, which is published as supporting information on the PNAS web site.

Because SCR in the anticipatory period (5–0 sec before card pick) predicts whether the choice was from the safe or risky decks in normal individuals (6), we examined this variable as well. As expected, we found a significant negative correlation between anticipatory SCR and the action values of the deck that the subject was going to pick at the trial ($r = -0.267$; $P = 0.01$; $n = 91$) (Fig. 4c). There was also a significant difference in anticipatory SCR between trials of choices from risky and safe decks (Mann–Whitney: $U = 555.0$; $P = 0.007$; $n = 91$; Fig. 4d), indicating that SCR in the anticipatory period was larger when the subject was about to make a risky choice.

Discussion

Judgment under risk and uncertainty is ubiquitous and unavoidable in our daily life and has been of considerable interest in economics and psychology. A number of important aspects of our decision-making have been revealed, in terms of how we evaluate gains and losses and how we can be risk-averse or risk-seeking (28, 29). In real life, relative comparisons are often more important than absolute values of choice outcomes to assess the values of an action and to guide advantageous decision-making. Similarly, prospect theory (30) states that our decisions are strongly influenced by evaluating changes of values from a subjective "reference point" rather than their absolute values, rendering choices susceptible to how they are "framed" (explained) and leading to preference-reversal when the reference point is shifted. Importantly, we do not weight gains and losses (seen as changes from some reference point), or low and high probabilities, in the same way.

We must often predict the outcomes of our choices from our prior experience. Reinforcement-learning works in this setting, without a "teaching signal," to predict the outcome of an action. The resulting discrepancy between predicted and actual outcome is coded as the PE, and this biases our future choices. The present investigation of the underlying neural mechanisms assigned an important component of reinforcement learning to neurons in the medial prefrontal cortex.

Specifically, we found that: (i) reward PE correlated with the ERP's alpha band component recorded from medial prefrontal cortex; (ii) the association between reward PE and alpha-band ERP was strongly driven by choices that were anticipated to be risky but violated the expectation of punishment; and (iii) emotional response (anticipatory SCR) was negatively correlated with action values.

It is important to note that the risky trials in which punishment was omitted were perceptually indistinguishable from safe trials

with no punishment. In both cases, the point in time at which a punishment could have occurred showed the text "please wait" instructing the subject to wait until the next trial. Yet we found a significant association between PE and alpha-band ERP in the risky, but not in the safe, case. Similarly, although triggered by the expectation of possible punishment (i.e., the temporal onset of the punishment feedback epoch), the ERPs we recorded could not have encoded actual punishment magnitude because none was administered on those trials we chose for analysis. We thus believe that the differences in electrophysiological activity we found must have been driven by the patient's expectations rather than the sensory stimuli presented: namely, the PE arising from the discrepancy between the two.

In studies that have recorded electroencephalograms at the scalp, medial PFC including anterior cingulate gyrus has been thought to be a main generator of "error-related" or "feedback-related" negativity (31–33) elicited by an incorrect response or by stimulus feedback that indicates error, respectively. Several features of our results provide further details to these prior electrophysiological studies. First, our recording sites were more ventral than those of the estimated generators of error-related negativity (31, 32, 34–36). The signals we recorded from closely spaced, bipolar, high-impedance contacts within the brain provide better localization because they avoid potential contamination by potentials generated at distant sources. Second, the ERP alpha-band component we found was sensitive to positive reward PE, in contrast to some other reports that found greater sensitivity of scalp-ERP signals for negative PE (37). Third, a subband component corresponding to the alpha-frequency

range was primarily responsible for encoding this error signal, in line with some prior findings (38, 39), whereas error-related negativity may instead be generated by a theta frequency component (36). Fourth, the responses we found appeared to be conditioned on the risk of the choice just made. Thus the subject apparently did not evaluate the reward PE (or values of the feedback) in the same way throughout the experiment, but differentially depending on the context, as has been reported in a prior study of error-related negativity (37).

The present findings implicate the ventro–medial PFC in the continuous updating of expectations of reward and punishment based on the reward PE signal that guides the acquisition of choice bias. The finding is consistent with other recently published studies (40, 41) as well as with dopaminergic inputs to this region of the brain (42). Evaluation of expected reward level (equivalently, risk of the choice) may thus take place within this region of the brain, or may be carried out through interaction with other reward/emotion-related structures to which it is connected (such as amygdala, basal ganglia, insula, and other sectors of PFC). It will be important in future studies to begin to dissect the functional network of which we have here studied only one component.

1. Fuster, J. M. (1989) The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe (Raven, New York), 2nd Ed.
2. Damasio, A. R. (1994) Descartes' Error: Emotion, Reason, and the Human Brain (Putnam, New York).
3. Rolls, E. T. (1999) The Brain and Emotion (Oxford Univ. Press, New York).
4. Jones, B. & Mishkin, M. (1972) Exp. Neurol. 36, 362–377.
5. Izquierdo, A., Suda, R. K. & Murray, E. A. (2004) J. Neurosci. 24, 7540–7548.
6. Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. (1997) Science 275, 1293–1295.
7. Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D. & Cohen, J. D. (1998) Science 280, 747–749.
8. Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S. & Cohen, J. D. (1999) Nature 402, 179–181.
9. Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, V. A. & Carter, C. S. (2004) Science 303, 1023–1026.
10. Schultz, W., Tremblay, L. & Hollerman, J. R. (1998) Neuropharmacology 37, 421–429.
11. Schultz, W., Tremblay, L. & Hollerman, J. R. (2000) Cereb. Cortex 10, 272–284.
12. Hikosaka, K. & Watanabe, M. (2000) Cereb. Cortex 10, 263–271.
13. Schoenbaum, G. & Eichenbaum, H. (1995) J. Neurophysiol. 74, 751–762.
14. Gottfried, J. A., O'Doherty, J. & Dolan, R. J. (2002) J. Neurosci. 22, 10829–10837.
15. Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M. & Parrish, T. (2003) Neuron 39, 701–711.
16. Breiter, H. C., Aharon, I., Kahneman, D., Dale, A. & Shizgal, P. (2001) Neuron 30, 619–639.
17. O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J. & Andrews, C. (2001) Nat. Neurosci. 4, 95–102.
18. O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H. & Dolan, R. J. (2003) Neuron 38, 329–337.
19. Tremblay, L. & Schultz, W. (2000) J. Neurophysiol. 83, 1864–1876.
20. Roesch, M. R. & Olson, C. R. (2004) Science 304, 307–310.
21. Bechara, A., Damasio, A. R., Damasio, H. & Anderson, S. W. (1994) Cognition 50, 7–15.
22. Bechara, A., Damasio, H., Tranel, D. & Anderson, S. W. (1998) J. Neurosci. 18, 428–437.

23. Howard, M. A., Volkov, I. O., Granner, M. A., Damasio, H. M., Ollendieck, M. C. & Bakken, H. E. (1996) J. Neurosurg. 84, 129–132.
24. Ongur, D., Ferry, A. T. & Price, J. L. (2003) J. Comp. Neurol. 460, 425–449.
25. O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K. & Dolan, R. J. (2004) Science 304, 452–454.
26. Sutton, R. S. & Barto, A. G. (1998) Reinforcement Learning: An Introduction (MIT Press, Cambridge, MA).
27. Dayan, P. & Abbott, L. F. (2001) Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems (MIT Press, Cambridge, MA).
28. Kahneman, D. & Tversky, A. (2000) Choices, Values, and Frames (Cambridge Univ. Press, New York, and Russell Sage Foundation, New York).
29. Camerer, C. (1995) in The Handbook of Experimental Economics, eds. Kagel, J. H. & Roth, A. E. (Princeton Univ. Press, Princeton), pp. 587–703.
30. Kahneman, D. & Tversky, A. (1979) Econometrica 47, 263–292.
31. Gehring, W. J. & Willoughby, A. R. (2002) Science 295, 2279–2282.
32. Miltner, W. H., Lemke, U., Weiss, T., Holroyd, C., Scheffers, M. K. & Coles, M. G. (2003) Biol. Psychol. 64, 157–166.
33. Yeung, N., Holroyd, C. B. & Cohen, J. D. (2004) Cereb. Cortex 15, 535–544.
34. Herrmann, M. J., Rommler, J., Ehlis, A. C., Heidrich, A. & Fallgatter, A. J. (2004) Brain Res. Cogn. Brain Res. 20, 294–299.
35. Nieuwenhuis, S., Yeung, N., Holroyd, C. B., Schurger, A. & Cohen, J. D. (2004) Cereb. Cortex 14, 741–747.
36. Luu, P., Tucker, D. M. & Makeig, S. (2004) Clin. Neurophysiol. 115, 1821–1835.
37. Holroyd, C. B., Larsen, J. T. & Cohen, J. D. (2004) Psychophysiology 41, 245–253.
38. Schutter, D. J., de Haan, E. H. & van Honk, J. (2004) Neuropsychologia 42, 939–943.
39. Jensen, O., Gelfand, J., Kounios, J. & Lisman, J. E. (2002) Cereb. Cortex 12, 877–882.
40. Ito, S., Stuphorn, V., Brown, J. W. & Schall, J. D. (2003) Science 302, 120–122.
41. Fletcher, P. C., Anderson, J. M., Shanks, D. R., Honey, R., Carpenter, T. A., Donovan, T., Papadakis, N. & Bullmore, E. T. (2001) Nat. Neurosci. 4, 1043–1048.
42. Fiorillo, C. D., Tobler, P. N. & Schultz, W. (2003) Science 299, 1898–1902.
43. Bechara, A. & Damasio, H. (2002) Neuropsychologia 40, 1675–1689.