Supporting Methods

Supporting Additional Information on the Subject. The subject from whom we recorded had been diagnosed with simple partial seizures. During a seizure, his head turned to the left with left arm tonic posturing, lasting from 5 s to 5 min. There was no report of any postictal neurological deficit such as motor weakness, emotional change, or personality change. There was also no evidence of nonepileptic seizures or migraine, and he had no history of psychiatric illness. Structural MRIs of his brain were normal. Resting PET scans (F-18-fluorodeoxyglucose) of his brain obtained during the interictal period showed a normal pattern of glucose uptake with no evidence of focal hypometabolism. Noninvasive EEG monitoring showed inter- and intraictal right central parietal spikes.

Prior to surgery, the patient was taking antiepileptic medication as follows: phenytoin (Dilantin) at 400 mg/day, tiagabine (Gabitril) at 64 mg/day, and lamotrigine (Lamictal) at 500 mg/day. During our recordings, these had been tapered to 300 mg/day, 48 mg/day, and 300 mg/day, respectively. Given that these drugs do not act on dopaminergic transmission (and given the normal performance of our patient on the gambling task), it is unlikely that the medication substantially influenced the responses we recorded.

Skin Conductance Recording. Electrodermal activity was continuously recorded in dc mode by using a MP-150 system (Biopac Systems, Goleta, CA.) with bipolar placement of Ag/AgCl electrodes (surface area = 0.79 cm^2) on the subject's thenar and hypothenar eminence of the nondominant hand, and sampled at 10 Hz. Raw waveforms were detrended by using exponential smoothing using the following calculation.

$$\mathrm{SCL}(t) = \gamma \mathrm{Sraw}(t) + (1 - \gamma)\mathrm{SCL}(t - 1) \qquad \text{with } \mathrm{SCL}(0) = \mathrm{Sraw}(1);$$

where Sraw is the raw value of recorded electrodermal activity and SCL is the low-frequency tonic trend component (skin conductance level). Skin conductance response (SCR) is defined as:

$$\mathrm{SCR(t)} = \mathrm{Sraw(t)} - \mathrm{SCL(t)}.$$

Here we used 0.05 as γ value. Note that SCR values can take negative values. Anticipatory SCR was determined from the average values of SCR(t) in the anticipatory time window, that is, 5 s before card choice.

Reinforcement Learning Modeling. We modeled the subject's behavioral data by means of a modified reinforcement learning algorithm (1, 2) .

First, we define the reward value on the *n*th trial, $r(n)$, as

$$r(n) = [(1-\sigma)r^{+}(n) + \sigma r^{-}(n)]/10$$

where $r^{+}(n)$ is the positive reward value obtained at the time of the *n*th card choice, and $r^{-}(n)$ is the subsequent negative reward value (punishment) obtained, both in \$. Here σ ($0 < \sigma < 1$) defines the subject's sensitivity to punishment (negative reward) (3). Second, we define the value at the *n*th trial, $V(n)$, as:

$$V(n) = \sum_{i} p(n,i) \cdot Q(n,i)$$

where, $p(n,i)$ is the action selection probability for action *i* on the *n*th trial, and $Q(n,i)$ is the action value of action *i* at *n*th trial. This is the average of all possible action values weighted by their action selection probability and is the general indicator of goodness or badness of the agent's state at the trial. We note that, in this formulation, *V* does not have to mean expected return (either discounted or not) or "critic" (1, 2, 4) because we modeled the task as a static action choice problem and used straight action learning mechanism. However, *V* is a good indicator to show how the subject learned the task and is similar in this respect to "subjective expected utility" (5).

Third, these action values are updated as:

for $i_{n+1} = i_n$

$$Q(n+1,i) = Q(n,i) + \beta[1 - p(n,i)]PE(n)$$

else

$$Q(n+1,i) = Q(n,i)$$

where β is a step-size parameter for the action value update ($0 < \beta < 1$) and $PE(n)$ represents the reward prediction error on the *n*th trial.

Action selection probability for taking action *i* on the *n*th trial, $p(n+1,i)$, (i.e., the subject's preference for choosing from a given deck on that trial) is determined by using the Softmax action selection rule using $Q(n,i)$:

$$p(n+1,i) = \frac{\exp(Q(n,i)/\tau)}{\sum_a \exp(Q(n,a)/\tau)}$$

where *a* represents an action (to choose a card from a given deck) and τ is a Boltzmann temperature parameter that represents the degree of exploitation and exploration (i.e., when τ goes toward 0, the selection becomes more and more deterministic, whereas if τ becomes large, the selections become more random).

Finally, the reward prediction error on the *n*th trial, $PE(n)$, can be defined as:

$$PE(n) = r(n) - Q(n,i)$$

Initial values of action selection probability and action values for each action (deck choice) were all set to 0.25 (= 1/4) and 0, respectively.

The model assumes that the subject maintains the magnitude of immediate monetary gain and the action value until the end of the trial, when the punishment or "wait" cue is administered, and at which point the prediction error is then calculated and the action value for choosing from that deck is updated. The reward prediction error and

ERP amplitude correlations that we report in our paper are thus based on data obtained at the same point in time: the end of each trial.

Maximum-likelihood estimates of the three free parameters were found by using a nonlinear parameter search algorithm with bounds for the parameters ($0 < \beta < 1$, $0 < \sigma < 1$); and yielded the following values that we used in the model: $\beta = 0.076$, $\tau = 1.319$, and $\sigma = 0.483$. To check that the estimated parameters specified a global minimum, log-likelihood values for various parameter sets were calculated and are presented in Fig. 11.

Multiresolution Analysis of ERP. Multiresolution decomposition of a signal with the discrete wavelet transform (DWT) has been successfully applied for the analysis of ERP (6, 7). This effectively decomposes the signal in a scale domain without redundancy and with fairly good time resolution, compared to the usual Fourier Transform method. Here we applied a six level multiresolution decomposition to ERP data using a nondecimated DWT (8) (the Over-Complete Discrete Wavelet Transform) to reduce the effect of shift-variance, as proposed by Bradley *et al.* (9). The decomposition yielded the following frequency subbands: D-1: 125-250 Hz; D-2: 62.5-125 Hz; D-3:31.3-62.5 Hz; D-4:15.6-31.3 Hz; D-5:7.8-15.6 Hz; D-6: 3.9-7.8 Hz; and A-6: 0-3.9 Hz (D refers to detail and A refers to approximate). This decomposition avoids the influence of choosing a particular starting point on the reconstructed waveform. We chose a biorthogonal quadratic spline wavelet (10) with a filter length of 4 and 20 as a mother wavelet function, due to its symmetry (linear phase filtering property), smoothness, nearly optimal time-frequency resolution, and compact support properties. Our analysis focused on the D-5 (corresponds to alpha) and D-6 (corresponds to theta) subband components. These were reconstructed by using the inverse transform, with only the coefficients of the corresponding detail level. Instantaneous amplitude and phase were calculated from the Hilbert transform of the reconstructed signal, and phase-locking values were calculated and evaluated by Rayleigh's test (11, 12). We used the usual pyramid algorithm implemented in MATLAB

(Mathworks, Natick, MA).

Statistical Evaluation. For statistical analysis, we calculated root-mean-square (rms) values of the reconstructed signal in a pre (200 to 0 msec before feedback for D-5 and 400 to 0 msec before feedback for D-6) and a postfeedback time window (200 to 400 ms after feedback for D-5 and 200 to 600 msec after feedback for D-6). These time windows were chosen based on the time course of ERP amplitude shown in Fig. 2c. For the correlation analyses performed in this study, we excluded trials with PE values less than -\$20 (2 trials were rejected with this criterion), since these PE values were due to very large punishments (more than \$200) that occurred infrequently and were accompanied by a special long feedback sound different from any of the other trials. The number of remaining valid trials used for analyses was 91. Pearson's correlation coefficients and their *P* values (raw values: not corrected for multiple comparisons) are reported throughout. All analyses were two-tailed, and *P* values remaining at $P < 0.05$ after adjustment for multiple comparisons were considered statistically significant.

Fig.6. Histogram showing the distribution of reward prediction error (PE) values. The two outliers (PE values less than -$ 20) were excluded from analyses presented in the text.

Fig. 7. Time course of reward PE. When the logarithm of the absolute value of PE is plotted against trial number, a significant regression was found (r = –0.217, P = 0.03, n = 100), indicating learning by the patient over time.

Fig. 8. Amplitudes and phase-locking values recorded at channel 1. (*a*) Mean instantaneous amplitude of field potentials on channel 1 (n=91). D1-D6 correspond decomposition levels of the discrete wavelet transform. 0 on the x-axis represents the time of punishment delivery (or "wait" cue delivery in non-punished trials). Roughly, D-1 to D-3 corresponds to gamma, D-4 corresponds to beta, D-5 corresponds to alpha, and D-6 corresponds to theta frequency bands. Note the clear increase of amplitude in alpha and theta bands at 200-600 msec. (*b*) Phase locking values (PLV) of field potentials on channel 1 (N=91). Significant phase concentration occurred in the alpha and theta frequency range from 200-500 msec. For this calculation, zero phase-shift windowed FIR (finite impulse response) band-pass filters of order 100 with center frequency 1-20 Hz (1-Hz step) were applied to the ERP data.

Fig. 9. Relationship between prediction error (PE) and alpha-band ERP amplitude. (*a*) For trials corresponding to choices from risky decks, there was a correlation between PE and ERP amplitude only for that subset in which no punishment was actually obtained (red), but not that subset in which punishment was given (blue). (*b*) PE versus alpha-band ERP amplitude for choices from safe decks, for that subset in which no punishment was given (blue) and that subset in which punishment was given (black); neither showed a significant correlation.

Fig. 10. Analyses of action values (Q). (*a*) Action values were significantly large for choices from the risky decks than for choices from the safe decks [$t(89) = 6.2$, $P < 0.001$, n =91]. Error bars are ± 1 SEM. Relationship between PE and alpha-band ERP amplitude (alpha rms value) for those trials that showed large Q values (*b*) did not

correlate, whereas PE and ERP amplitude were correlated for those trials that showed negative *Q* values (*c*).

Fig. 11. Log-likelihood maps for various parameter combinations, verifying that a global minimum was achieved. We held either τ or σ fixed and plot the values of the other two parameters. τ, temperature parameter; β, step-size parameter; and σ, sensitivity to punishment. Color encodes log-likelihood of the model.

References for supporting information:

1. Sutton, R. S. & Barto, A. G. (1998) *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA.).
2. Dayan, P. & Abbott, L. F. (2001) *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems* (MIT Press, Cambridge, MA.).
3. Busemeyer, J. R. & Stout, J. C. (2002) *Psychol. Assess.* **14,** 253-262.
4. Watkins, C. J. C. H. (1989) Ph.D. thesis (Cambridge Univ., Cambridge, U.K.).
5. Camerer, C. (1995) in *The Handbook of Experimental Economics*, eds. Kagel, J. H. & Roth, A. E. (Princeton Univ. Press, Princeton).
6. Demiralp, T. & Ademoglu, A. (2001) *Clin. Electroencephal.* **32,** 122-138.
7. Quian Quiroga, R., Sakowitz, O. W., Basar, E. & Schurmann, M. (2001) *Brain Res. Brain Res. Protoc.* **8,** 16-24.
8. Percival, D. B. & Walden, A. T. (2000) *Wavelet Methods for Time Series Analysis* (Cambridge Univ. Press, New York).
9. Bradley, A. P. & Wilson, W. J. (2004) *Clin. Neurophysiol.* **115,** 1114-1128.
10. Daubechies, I. (1992) *Ten Lectures on Wavelets* (Society for Industrial and Applied Mathematics, Philadelphia, PA.).
11. Fisher, N. I. (1993) *Statistical Analysis of Circular Data* (Cambridge Univ. Press, New York).
12. Mardia, K. V. (1972) *Statistics of Directional Data* (Academic, London).