SI Appendix

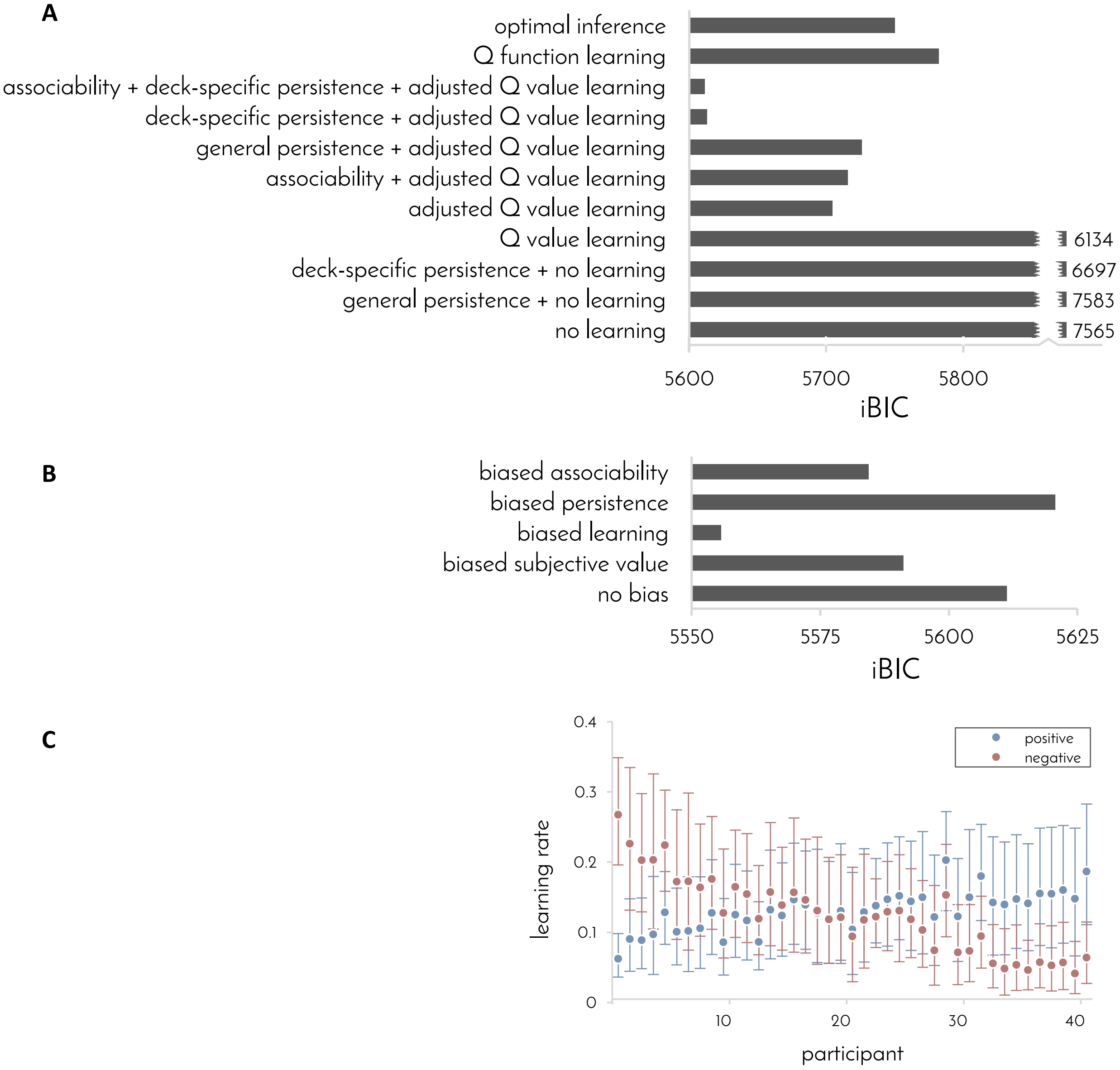

Figure S1. Model comparison and parameter fitting. **(A)** Eleven different learning algorithms were fitted to participants' behavior. Goodness of fit was computed using the integrated Bayesian Information Criterion (iBIC[15]). A difference larger than 10 constitutes very strong evidence in favor of the model with lower iBIC value. The best-fitting model ('adjusted Q value learning + associability + deck-specific persistence') learns a value for taking a gamble with each of the three decks. Learning in the model is driven by associability-weighted prediction errors (i.e., the difference between actual and expected outcomes), where outcome expectations factor in previous experience with the deck and the computer's number. Associability was modeled as in previous work[11,12]. In addition, the model tends to repeat actions recently taken with each deck (Deck-specific persistence, modeled as in previous work[10]). Because the same model without associability explained the data almost equally well (iBIC difference = 2), we proceeded to evaluate learning/persistence biases both with, and without, associability. **(B)** The best fitting model from Figure S1A was compared as is ('no bias') with four variants of the model, each including a different type of learning/persistence bias. Note that all models already include a baseline decision bias parameter. Of the four variants, the best fitting model involved a bias in learning, implemented by allowing two different learning rates for negative and positive prediction errors. We also tested the same biases on the model without associability, but these did not fit the data as well (iBIC difference between best variants of each model = 16.7 in favor of model with associability). **(C)** Individual learning rates fitted to each participants' behavior using the best-fitting model from **(B)**. Learning rates for negative prediction errors (red) and for positive prediction errors (blue) were widely distributed anti-correlated (r_s = –0.57, p = 10^{-4}, permutation test). Error bars: 95% CI.

A

simulated model	best-fitting model(s) (10 trials)									
no learning: 1	1	1	1	1	1	1	1	1	1	1
general persistence + no learning: 2	1	1	1	1	1	1	1	1	1	1
deck-specific persistence + no learning: 3	3	3	3	3	3	3	3	3	3	3
Q value learning: 4	4	4	4	4	4	4	4	4	4	4
adjusted Q value learning: 5	5	5	5	5	5	5	5	5	5	5
associability + adjusted Q value learning: 6	6	6	6	6	6	6	6	6	6,5	6
general persistence + adjusted Q value learning: 7	5	5	5	5	5	5	5	5	5	5
deck-specific persistence + adjusted Q value learning: 8	8	8	8	8	8	8	8	8	8	8
associability + deck-specific persistence + adjusted Q value learning: 9	9	9	9	9,8	9,8	8,9	8,9	9	9	8
Q function learning: 10	10	10	10	10	10	10	10	10	10	10

B

simulated model	best-fitting model(s) (10 trials)									
no bias: 1	1,4	1	1	1	1	1	1	1	1,4	1
biased subjective value: 2	2,3	2	2	2	3,2	3,2	3,2	3,2	2	2,3
biased learning: 3	3	3	3	3	3	3	3	3	3	2,3
biased persistence: 4	1	1	1	1	1	1	1	1	1	1
biased associability: 5	1	1	5	5	5	5	1	1	5	5

Figure S2. Validation of the model comparison procedure. We used each of the models to generate 10 full experimental data sets (each data set comprised 41 participants, 180 trials per participant) by having each model perform the experiment with each of the parametrizations that best-fitted individual participants. The signal-to-noise ratio in these simulations was determined by setting the β parameters as those which fitted participants' behavior the best. We then applied the model-comparison procedure to each simulated data set. The best-fitting models were defined as the models with the lowest BIC score or within 6 of the lowest BIC, since a BIC difference of 6 indicates strong evidence[14]. **(A)** Validation of model comparison shown in **Figure S1**A. The models that best-fitted the real experimental data (models 8 and 9) best-fitted only datasets generated by these same models (20/20) and none of the data sets generated by other models (0/80). Note that, as expected, models, in which some parameters were poorly justified by the experimental data, were sometimes confused with simpler models. Algorithm 11 ('optimal inference') was omitted from the validation due to its prohibitive computational complexity, as it involves a nested slice-sampling procedure on each simulated trial. **(B)** Validation of model comparison shown in **Figure S1**B. The model that best-fitted the real experimental data (model 3) only best-fitted (as a sole winner) datasets generated by that same model (9/10) and none of the datasets generated by other models (0/40).

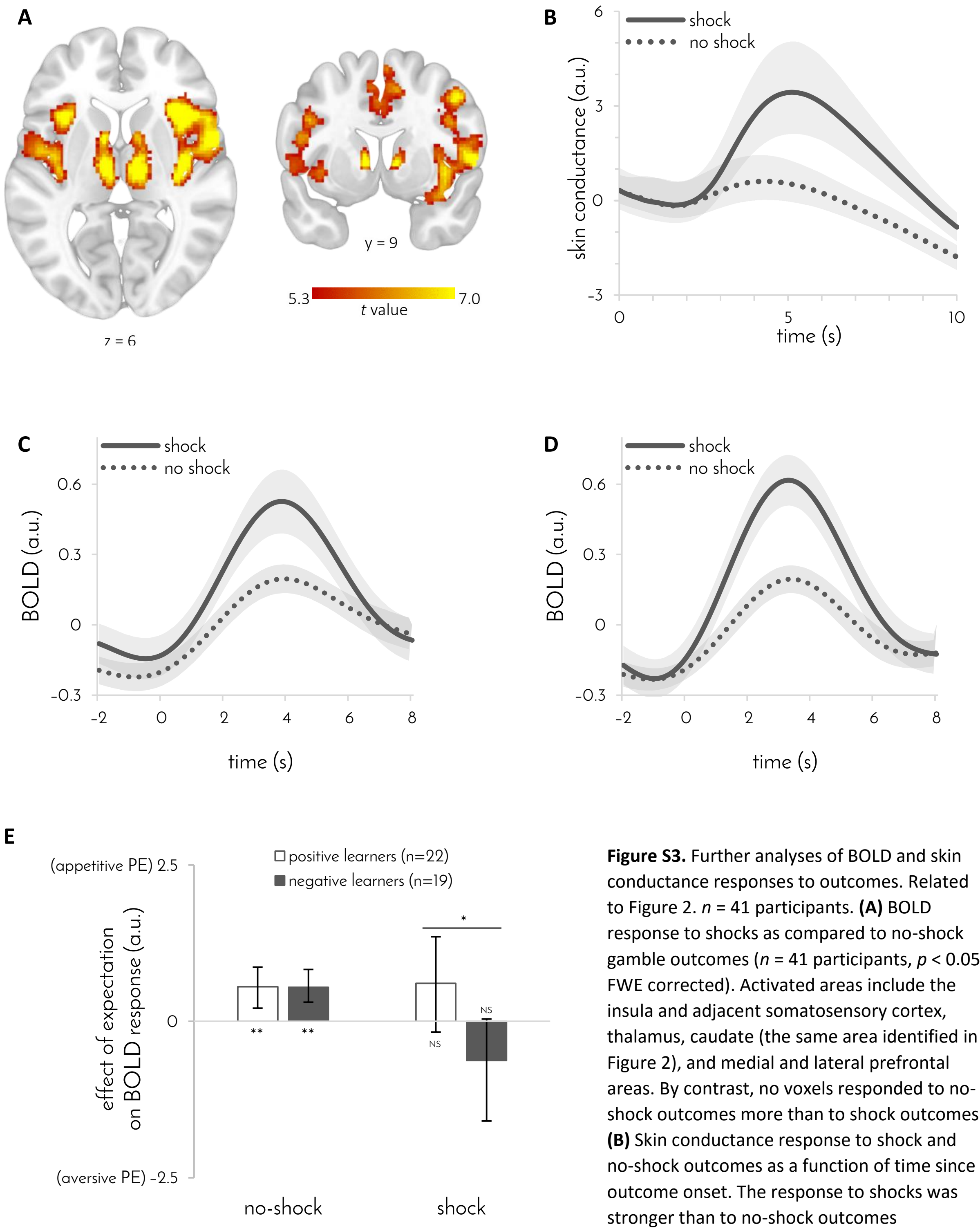

Figure S3. Further analyses of BOLD and skin conductance responses to outcomes. Related to Figure 2. $n = 41$ participants. **(A)** BOLD response to shocks as compared to no-shock gamble outcomes ($n = 41$ participants, $p < 0.05$ FWE corrected). Activated areas include the insula and adjacent somatosensory cortex, thalamus, caudate (the same area identified in Figure 2), and medial and lateral prefrontal areas. By contrast, no voxels responded to no-shock outcomes more than to shock outcomes. **(B)** Skin conductance response to shock and no-shock outcomes as a function of time since outcome onset. The response to shocks was stronger than to no-shock outcomes (difference between outcomes 3.7, CI 1.0 to 7.4, GLM, $p = 0.007$, bootstrap test) and this effect was similar in positive and negative learners (difference between groups 3.8, CI –4.4 to 9.6, GLM, $p = 0.34$, bootstrap test). Skin conductance responses were baseline-corrected by the average level at the first two seconds. Shaded area: 95% bootstrap CI. a.u.: arbitrary units. **(C)** BOLD response to shock and no-shock outcomes in negative learners ($n = 19$ participants). **(D)** BOLD response to shock and no-shock outcomes in positive learners ($n = 22$ participants). In **(C)** and **(D)**, shaded area denotes s.e.m, and time 0 indicates outcome onset. **(E)** Effect of expectations on BOLD response to outcomes in periaqueductal gray (PAG) as a function of learning bias and outcome type. The pattern of activity resembles that found in the striatum (**see Figure 2**C). Following Linnman et al. (2012), GLM coefficients were taken from MNI coordinates [±4 –29 –12]. Error bars: 95% bootstrap CI, **: $p < 0.002$, *: $p < 0.02$, NS: $p > 0.05$.

SI Material and Methods

Participants. 43 human volunteers (age range = 18–42 years, 30 female, 12 male, recruited from a participant pool at University College London) participated in the experiment. Inclusion criteria were based on age (minimum = 18 years, maximum = 50 years) and right-handedness. Exclusion criteria included color blindness, neurological or psychiatric illness, and psychoactive drug use. Before the experiment, participants completed an 80-item questionnaire composed of several measures of different mood and anxiety traits[1-5]. Age, sex and mood and anxiety traits did not differ between participants later classified as positive and negative learners (all $p > 0.1$, bootstrap test). To allow sufficient statistical power for comparisons between two groups of participants, the sample size was set as roughly double the sample sizes that are recommended in the literature and that have been used in recent functional Magnetic Resonance Imaging (fMRI) studies of decision-making. Two participants failed to complete the experiment due to anxiety or discomfort and were excluded, leaving 41 participants in all subsequent behavioral and neural analysis. Participants received monetary compensation for their time (between £25 and £30). The experimental protocol was approved by the University of College London local research ethics committee, and informed consent was obtained from all participants.

Experimental task. To test for individual differences in learning from actual painful outcomes compared to learning from success in preventing pain, we designed a card game, inspired by previous work on reward learning[6,7], in which participants' goal was to minimize the number of painful electrical shocks they could receive. The game consisted of 180 trials, divided into three 60-trial blocks. On each trial, participants were first shown which one of three possible decks (each having distinct color and pattern) they will be playing with. After a short interval (2 to 5 s, uniformly distributed), the computer drew a number between 1 and 9 and participants had up to 2.5 s to choose whether they wanted to gamble that the number that they draw will be higher than the computer's number. If participants chose to gamble, they avoided a shock if the number that they drew was indeed higher than the computer's number, and they received a shock if it was lower (as well as in half of the trials in which the numbers were equal). Conversely, if participants declined the gamble, they received a shock with a fixed 50% probability that was known to the participants. Not making any choice always resulted in a shock. Feedback was provided 700 ms following each choice and consisted of a 'shock', 'no-shock' or 'shock/no-shock' visual symbol (**Figure 1**A) accompanied, when appropriate, by electrical stimulation (the drawn number was not shown). Trials in which no choice was made (less than 1% of trials) were excluded from all subsequent analyses. Critically, participants were told that each of the three decks contained a different proportion of high and low numbers, and thus, they had to learn by trial and error how likely a gamble was to succeed with each of the decks. Unbeknownst to participants, one deck contained a uniform distribution of numbers between 1 and 9 ('even deck'), one deck contained more 1's than other numbers ('low deck'), making gambles 30% less likely to succeed, and one deck contained more 9's than other numbers ('high deck'), making gambles 30% more likely to succeed. In the first 15 trials, the computer drew the numbers 4, 5, and 6 three times each, and the other numbers once each. To make sure that all participants take a gamble in approximately 50% of trials, in each subsequent set of 15 trials, the numbers that the computer drew three times were increased by one (e.g., [4, 5, 6] → [5, 6, 7]) if participants took two thirds or more of the gambles against these numbers in the previous 15 trials, or decreased by one if participants took a third or less of the gambles. Participants' decks were pseudorandomly ordered while ensuring that the three decks were matched against similar computers' numbers and that no deck appeared in successive trials more than the other decks.

Electrical stimulation. Participants underwent an established individual pain titration procedure[8,9] with a Digitimer DS7a electric stimulator (Welwyn Garden City, UK). Following a brief overview of the equipment and titration process, an electrode was placed on the back of the participant's left hand. Titration began with a low-current electric shock (0.1 mA) and participants were asked to rate their experience of pain on a visual 22-point scale (ranging from 0 = no sensation to 5 = mildly painful to 10 = intolerable). The initial rating was followed by a series of shocks, increasing in small milliamp increments. Subjective ratings of pain were collected after each shock until a rating of 6 was reached. The final shock intensity was then used throughout the experiment. Habituation to stimulation over the course of the experiment, as measured by how participants rated the shock again at the end of the experiment, was generally mild (mean rating change –0.12). Absolute shock intensities and levels of habituation did not differ significantly between participants later classified as positive and negative learners ($p > 0.1$, bootstrap test).

Pre-task training. Before performing the experiment, to familiarize participants with the basic structure of the task, participants received training outside the scanner without electrical shock feedback. Training consisted of 60 trials involving a single 'even' deck and visual feedback indicating the number that participants drew.

Post-task questionnaire. Following the experiment, participants were asked to rate each deck as to whether it contained mostly low or mostly high numbers on a visual 22-point scale (ranging from 0 = only low numbers to 1 = only high numbers). Rating confirmed that participants learned the task well (low deck 0.22 CI 0.17 to 0.29; even deck 0.43 CI 0.37 to 0.47; high deck 0.81 CI 0.74 to 0.86), and the ratings did not differ between participants later classified as positive and negative learners ($p > 0.1$, bootstrap test). No participant reported being aware that the computer's numbers were adjusted to the participant's choices.

Propensity to gamble. To compute a participant's propensity to take or avoid gambles, we fitted to participant's decisions a logistic regression model comprised of three terms: an intercept, the computer's numbers (scaled to range between –1 (for the number 9) and 1 (for the number 1)), and the participant's deck (-1 for low, 0 for even and 1 for high). Propensity to gamble was then computed by applying the logistic function to the intercept alone and scaling the result to range between –1 and 1. This measure indicates the participant's tendency to take or avoid gambles when the odds of winning and losing are equal (i.e., when playing with the even deck against the number 5).

Learning algorithms. To determine what learning algorithm participants used to perform the task, we compared five different algorithms in terms of how well they explained participant's choices. In all algorithms, the probability of taking each gamble was modeled by applying the logistic function to a term that represented available evidence.

Algorithm 1 ('no learning') is oblivious to previous experience with the decks, and it computes the evidence as $\beta + \beta' N_t$, where N_t is the computer's number at trial t, scaled between –1 and 1 as above, β' is an inverse temperature parameter, and β is a decision bias parameter.

Algorithm 2 ('no learning + general persistence') tends to repeat recently taken actions[10]. To this end, it maintains a persistence variable p_a for each action a ('gamble' and 'decline'). p_t^a is set to one when the action is taken, and decays exponentially through multiplication by a free parameter λ otherwise. The evidence is then computed as $\beta + \beta' N_t + \beta'' \Delta p_t$, where $\Delta p_t = p_t^{\text{gamble}} - p_t^{\text{decline}}$, and β'' is a free parameter that controls persistence strength.

Algorithm 3 ('no learning + deck-specific persistence') tends to repeat actions recently taken with each deck. Thus, it maintains a persistence variable $p_t^{d,a}$ for each deck-action pair (d, a), and the evidence is computed with respect to the current deck as $\beta + \beta' N_t + \beta'' \Delta p_t^{d_t}$, where $\Delta p_t^{d_t} = p_t^{d_t,\text{gamble}} - p_t^{d_t,\text{decline}}$.

Algorithm 4 ('Q value learning') tracks the expected outcome of gambles with each deck *d* by means of a Q value as follows: $Q_{t+1}^{d_t} = Q_t^{d_t} + \eta\delta_t$, where $\delta_t = r_t - Q_t^{d_t}$ is the difference between the actual (r_t) and expected ($Q_t^{d_t}$) outcome of a gamble (i.e., the outcome prediction error, ignoring the effect of the computer's number), $r_t = 1$ stands for shock, $r_t = -1$ stands for no shock, and η is a learning rate parameter. The evidence is then computed as $\beta + \beta' N_t + \beta'' Q_t^{d_t}$.

Algorithms 5 ('adjusted Q value learning') is similar to algorithm 4, except that prediction errors are computed with respect to expectations that also factor in the computer's number: $\delta_t = r_t - Q_t^{d_t} - \frac{\beta'}{\beta''} N_t$. This way, the algorithm learns more about the decks from outcomes that are more surprising (i.e., from no-shock outcomes of gambles taken against higher numbers, and from shock outcomes of gambles taken against lower numbers).

Algorithm 6 ('adjusted Q value learning + associability') is similar to algorithm 5, except that learning is modulated by an associability variable α_t^d, computed as a running average of the absolute value of recent prediction errors for each deck (i.e., $\alpha_{t+1}^{d_t} = \alpha_t^{d_t} + \eta'\left(|\delta_t| - \alpha_t^{d_t}\right)$), where η' is the associability update rate[11,12]. Thus, Q values were updated as $Q_{t+1}^{d_t} = Q_t^{d_t} +$

$\alpha_t^{d_t}\eta\delta_t$. Associability was initialized as a free parameter in between 0 and the maximal possible prediction error.

Algorithm 7 ('adjusted Q value learning + general persistence') is similar to algorithm 5, except that it tends to repeat recent actions similarly to algorithm 2. Thus, it computes the evidence as $\beta + \beta' N_t + \beta'' Q_t^{d_t} + \beta''' \Delta p_t$.

Algorithm 8 ('adjusted Q value learning + deck-specific persistence') is similar to algorithm 5, except that it tends to repeat actions recently taken with each deck similarly to algorithm 3. Thus, it computes the evidence as $\beta + \beta' N_t + \beta'' Q_t^{d_t} + \beta''' \Delta p_t^{d_t}$.

Algorithm 9 ('adjusted Q value learning + deck-specific persistence + associability') is similar to algorithm 8, except that learning is modulated by associability as in algorithm 6.

Algorithm 10 ('Q function learning') learns a two-parameter logistic function for each deck, consisting of an intercept $a_{t+1}^{d_t} = a_t^{d_t} + \eta\delta_t$, and a slope $b_{t+1}^{d_t} = b_t^{d_t} + \eta' N_t \delta_t$, where δ_t is computed by applying the logistic function to $a_t^{d_t} + b_t^{d_t} N_t$ and subtracting this quantity from 0 in the case of a shock outcome or from 1 in the case of a shock outcome. These update equations constitute a simplification of the Iteratively Reweighted Least Squares (IRLS) maximum likelihood estimation for logistic regression[13]. The evidence is then computed as $\beta + \beta'\left(a_t^{d_t} + b_t^{d_t} N_t\right)$.

Algorithm 11 ('optimal inference') makes full use of all available evidence given what participants knew about the task. On each trial, the algorithm infers the maximum a posteriori solution for the logistic function corresponding to each deck, given all previously observed outcomes and Gaussian priors on the intercept and slope variables (intercept prior mean = 0 and slope prior mean = 2.29, which fit the training deck; intercept and slope variance determined by free parameters). The evidence is then computed as in Algorithm 10. This algorithm was implemented by estimating through slice sampling[13] on each trial the Bayesian logistic regression solution given all previously observed gamble outcomes.

Learning/persistence biases. After identifying the best-fitting learning algorithms (Algorithm 8: 'adjusted Q value learning + deck-specific persistence' and Algorithm 9: 'adjusted Q value learning + deck-specific persistence + associability'), we tested whether the algorithms' ability to explain participants' choices would be improved by implementing a learning/persistence bias in favor of gambling or declining a gamble. The models already include a decision bias parameter that allows them to favor either gambling or declining to begin with, but a learning/persistence bias can make such a tendency evolve over time. Thus, we compared the basic algorithm ('no bias') to four variants of the same algorithm, each of which involves a different type of additional bias. Variant 1 ('biased subjective value') is allowed to weight shock and no-shock outcomes differently by means of a subjective value bias parameter ψ. Thus, $\mathrm{r_t}$ is set as $\sqrt{\psi}$ for no-shock outcomes and as $-\frac{1}{\sqrt{\psi}}$ for shock outcomes, such that ψ reflects the ratio between the subjective value of no-shock and shock outcomes. Variant 2 ('biased learning') is allowed to learn at a different rate from shock and no shock outcomes. Therefore, this variant includes two learning rate parameters, one for shock outcomes (η^-) and one for no-shock outcomes (η^+). Variant 3 ('biased persistence') allows differential persistence in gambling and declining. Therefore, this variant includes two persistence decay parameters, one for gambling and one for declining. Variant 4 ('biased associability', for Algorithm 9 only) is allowed to update associability at a different rate following shock and no shock outcomes. Therefore, this variant includes two associability update rate parameters, one for gambling and one for declining. Variant 2 of Algorithm 9 turned out to be the best-fitting model (see Model fitting and Model comparison below), and thus, individually fitted positive and negative learning rate parameters were used to classify participants as positive ($\eta^+ > \eta^-$) and negative ($\eta^+ < \eta^-$) learners.

Model fitting. To fit the parameters of the different learning algorithm to participants' choices, we used a hierarchical expectation-maximization approach[13]. We first modeled each of the parameters using some initial prior distribution at the group level. We then sampled 100,000 random parameterizations from these priors, computed the likelihood of observing participants' choices given each parametrization, and used the computed likelihoods as importance weights[61] to resample (and accordingly reparameterize) the group-level prior distributions. These steps were iteratively repeated until convergence. Finally, to obtain the best-fitting parameters for each individual participant, we computed a weighted mean of the final batch of 100,000 parametrizations, in which each parameterization was weighted by the likelihood it assigned to the individual participant's choices. Learning rate parameters were modeled with beta distributions (initialized with $\alpha = 1$, $\beta = 1$), inverse temperature and variance parameters were modeled with gamma distributions (initialized with $k = 3$, $\theta = 3$), the bias parameter was modeled with a normal distribution (initialized with $\mu = 0$ and $\sigma = 1$), and the subjective-value bias parameter was modeled with a log-normal distribution (initialized with $\mu = 0$ and $\sigma = 1$).

Model comparison. To compare between pairs of models, in terms of how well each model accounted for participants' choices, we estimated the log Bayes factor[14] by means of an integrated Bayesian Information Criterion[15] (iBIC). We estimated the evidence in favor of each model ($\mathcal{L}$) as the mean likelihood of the model given 100,000 random parameterizations drawn from the fitted group-level priors[13]. We then computed the iBIC by penalizing the model evidence to account for model complexity as follows: $\mathrm{iBIC} = -2\ln\mathcal{L} + k\ln n$, where *k* is the number of fitted parameters and *n* is the number of participant choices used to compute the likelihood. Lower iBIC values indicate a more parsimonious model fit.

fMRI data acquisition. Whole-brain T2*-weighted echo-planar imaging (EPI) data were acquired using a Siemens Trio 3T scanner, using a 32-channel headcoil. The sequence chosen was selected to minimize dropout in the striatum, anterior cingulate and amygdala[16]. Each volume contained 37 slices of 3-mm isotropic data; echo time = 30 ms, repetition time = 2.56 s per volume, echo spacing of 0.5 ms, slice tilt of −30° (T > C), Z-shim of −0.4 mT/m ms, ascending slice acquisition order. The mean number of volumes acquired per partcipant was 867 (the total number of volumes acquired varied depending on participants' choice times). To account for T1 saturation effects, the first six volumes of each session, taken before the experiment was started, were discarded.

Structural MRI data acquisition. Magnetic Transfer (MT) maps, which are particularly suitable for structural measurements of subcortical regions[17], were calculated from a multi-parameter protocol based on a 3D multi-echo fast low angle shot (FLASH) sequence[18]. Three co-localized 3D multi-echo FLASH datasets were acquired in sagittal orientation with 1 mm isotropic resolution (176 partitions, field of view (FOV) = 256 × 240 mm^2, matrix 256 × 240 × 176) and non-selective excitation with predominantly proton density weighting (PDw: TR/α = 23.7 ms/6°), T1 weighting (18.7 ms/20°), and MT weighting (23.7 ms/6°; excitation preceded by an off-resonance Gaussian MT pulse of 4 ms duration, 220° nominal flip angle, 2 kHz frequency offset). The signals of six equidistant bipolar gradient echoes (at 2.2 ms to 14.7 ms echo time) were averaged to increase the signal-to-noise ratio. Semi-quantitative MT parameter maps, corresponding to the additional saturation created by a single MT pulse, were calculated by means of the signal amplitudes and T1 maps[19], eliminating the influence of relaxation and B1 inhomogeneity[20].

Field maps. Whole-brain field maps (3-mm isotropic) were acquired to allow for subsequent correction in geometric distortions in EPI data at high field strength. Acquisition parameters were 10-ms/12.46-ms echo times (short/long respectively), 37-ms total EPI readout time, with positive/up phase encode direction and phase-encode blip polarity −1.

Physiological monitoring. During scanning sessions, peripheral measurements of participants' pulse, breathing and skin conductance were made together with scanner slice synchronization pulses using Spike2 data acquisition system (Cambridge Electronic Design Limited, Cambridge UK). The cardiac pulse signal was measured using an MRI compatible pulse oximeter (Model 8600 F0, Nonin Medical, Inc. Plymouth, MN) attached to the participant's left index finger. The respiratory signal, thoracic movement, was monitored using a pneumatic belt positioned around the abdomen close to the diaphragm. Skin conductance was recorded on the tips of the left middle and ring fingers using EL509 electrodes (Biopac Systems Inc., Goleta, CA, USA) and 0.5%-NaCl electrode paste (GEL101; Biopac). Constant voltage (2.5 V) was provided by a custom-build coupler, whose output was converted to an optical pulse with a minimum frequency of 100 Hz at 0 µS to avoid aliasing, and then converted to a digital signal (Micro1401, CED, Cambridge, UK). Temperature and relative humidity of the experimental room was kept at 20 °C and 50%.

fMRI preprocessing. The following pre-statistics processing was applied in SPM12 (Wellcome Trust Centre for Neuroimaging) using default settings: slice-timing correction, motion correction, field-map-based distortion correction, co-registration with structural MRI and normalization to MNI space, spatial smoothing using a Gaussian kernel of 8.0 mm Full-Width at Half Maximum (FWHM), and high-pass temporal filtering with a cutoff frequency of 0.0078 Hz.

fMRI General Linear Model (GLM). To examine BOLD responses to the different decks, as well as the representation of prediction error signals, we performed a GLM analysis using SPM12 that included separate regressors indicating onsets of the appearance of the low, even and high decks, the computer's number draw, the participant's decision, and the four different types of outcomes (shock or no-shock outcomes of taken or declined gambles). In addition, the GLM included parametric regressors indicating the computer's number when number was drawn, the participants' choice at the time of decision, and the participant's prediction errors at gamble outcomes. Prediction errors were computed by applying the learning model, instantiated with mean group parameters, to the participant's sequence of stimuli and outcomes. Mean group parameters were used in line with previous work[21-27] in order to regularize individual estimates, which are otherwise noisy, as well as to ensure that a participant's behavioral data do not bias the results of the participant's GLM analysis. This latter concern is particularly relevant to studies of individual differences in fMRI, in which different parameterizations of the model will return different results for the same fMRI dataset. Thus, when using individual parameterizations, it is uncertain whether inter-individual differences in the results are due to differences in brain activity or due to differences in the parameterization of the model. The GLM also included 18 regressors for cardiac and respiratory phases to correct for physiological noise[28] and 6 motion parameters regressors to correct for motion-induced noise. In addition to this primary GLM, to test whether the BOLD response to outcomes reflected both previous experience with the decks and the computer's numbers, we used an additional GLM with similar regressors but including two parametric regressors at gamble outcome onset, one indicating the Q value of the current deck as derived from the model, and another one indicating the number drawn by the computer, orthogonalizing in turn the two regressors with respect to one another. Group-level significance of prediction error GLM coefficients was tested with FWE correction for the volume of the striatum, or, when examining BOLD response in a region of interest as a whole, by averaging the coefficients extracted from all voxels that comprise the region and then using a bootstrap test, Bonferroni-corrected for the number of regions. Anatomical regions of interest were identified using MNI coordinates provided with SPM12 by Neuromorphometrics, Inc. (Somerville, MA, USA) under academic subscription. Statistical brain maps were imaged using MRIcroGL (http://www.mccauslandcenter.sc.edu/mricrogl/) and overlaid on high-resolution anatomical images provided with the software.

fMRI time course analysis. To assess the time course of the effects of different components of the prediction error on the BOLD response to outcomes, we regressed the preprocessed BOLD signal (averaged across the functionally defined striatal ROI) for each time point from 2 s prior to outcome onset to 8 s following outcome onset against the model-derived deck Q value and the number drawn by the computer. The BOLD signal was upsampled to 100 Hz to allow averaging across trials with disparate fMRI acquisition timings. Both the BOLD signal and the regressors were z-scored. The two regressors were orthogonalized with respect to one another. The regression was performed separately for each type of outcome and for each functional MRI run (each run corresponded to an experimental block), and regression coefficients were averaged across runs.

fMRI functional connectivity analysis. To examine functional connectivity with striatal and amygdala areas in which responses to outcomes were modulated by expectations, we fit a GLM that included as regressors the preprocessed BOLD signal from three areas: 1. Striatal area where responses to no-shock outcomes were modulated by expectations ($p < 0.05$ FWE small-volume corrected). 2. Amygdala area where responses to shock outcomes were modulated by expectations ($p < 0.05$ FWE small-volume corrected). 3. Average gray matter signal. Thus, the coefficients fitted to the first two regressors reflected functional connectivity specific to either the striatal or amygdala ROI, accounting for variance shared between these regressors as well as with the global gray-matter signal. The GLM also included 18 physiological regressors and 6 motion parameters regressors to correct for these sources of noise.

fMRI response to decks. To examine the similarity between the BOLD response to the even deck and the BOLD response to the low and high decks, we computed for each participant the Euclidean distance between the vector of gray-matter GLM coefficients for the even deck and the GLM coefficients for the low ($\mathrm{D}_{\mathrm{even/low}}$) and high ($\mathrm{D}_{\mathrm{even/high}}$) decks. We then computed the even deck similarity index as $\frac{\mathrm{D}_{\mathrm{even/low}}-\mathrm{D}_{\mathrm{even/high}}}{\mathrm{D}_{\mathrm{even/low}}+\mathrm{D}_{\mathrm{even/high}}}$. A similarity index of 1 indicated identity to the high deck and a value of –1 indicated identity to the low deck.

Skin conductance analysis. We tested the effect of outcomes on skin conductance using SCRalyze (http://scralyze.sourceforge.net), which employs a GLM for event-related evoked skin conductance responses[29]. Skin conductance time series were filtered with a bidirectional first order Butterworth band pass filter with cut-off frequencies of 5 and 0.0159 Hz, and then modeled using the same GLM used for the fMRI analysis.

Voxel-based morphometry. To compute gray matter density maps, we segmented the MT maps into different tissue classes – gray matter, white matter and non-brain voxels (cerebrospinal fluid, skull) – and then normalized the tissue maps to MNI space using the Dartel algorithm in SPM12 with default settings. Subsequently, the tissue maps were scaled by the Jacobian determinants from the final normalization step, so as to preserve the total volume of tissue in each structure[30], and then smoothed by convolution with an isotropic Gaussian kernel of 3 mm FWHM.

Learning biases prediction. To predict participants' learning biases (η^{+} minus η^{-}), we used gray matter density data from the 6,315 voxels that comprised the striatum (corresponding to the caudate, putamen and accumbens labels in the MNI atlas) as 6,315 predictors in a regularized linear regression model. Predictions were generated in a 5-fold cross validation scheme, predicting the learning biases of each fifth of the participants using a regression model that was fitted to the rest of the participants[31]. Regularization was performed using the Least Absolute Shrinkage and Selection Operator (LASSO) method[32]. We used 5 different settings of the LASSO shrinkage factor (1, 0.1, 0.01, 0.001, 0.0001) and found that 0.0001 yielded the highest correlation between predicted and actual values. We corrected for multiple comparisons using a permutation test, in which the null distribution was generated by permuting the vector of actual learning biases 10,000 times, and applying the same procedure described above to predict each permuted vector with each of the 5 shrinkage factors while taking the highest correlation coefficient found for each permutation. To ensure that predictions did not simply reflect global effects of participant age, sex or whole-brain gray matter volume, we regressed all variance that could be explained by these variables out of the predicted learning biases.

Statistical analysis. Since many of the variables of interest were not normally distributed, we report non-parametric statistics throughout the manuscript. Bias-corrected and accelerated bootstrapping[33] with 10,000 samples was used to generate 95% confidence intervals and to test the significance of differences between two vectors or between a single vector and zero. Randomization tests[34] with 10,000 permutations were used to test significance of correlations. All correlation coefficients denote Spearman rank correlations, except for the correlation between predicted and actual learning biases which denotes Pearson linear correlation, since learning biases were predicted using a linear regression model. All non-directional tests are two tailed and all directional tests are one tailed. All data analysis was performed using MATLAB (Mathworks, Natick, MA, USA).

SI References

1. Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* 54, 1063.
2. Chiappelli J, Nugent KL, Thangavelu K, Searcy K, Hong LE (2013) Assessment of trait and state aspects of depression in schizophrenia. *Schizophrenia Bull* 40, 132–142.
3. Eckblad M, Chapman LJ (1986) Development and validation of a scale for hypomanic personality. *J Abnorm Psychol* 95, 214.
4. Poreh AM, et al. (2006) The BPQ: A scale for the assessment of borderline personality based on DSM-IV criteria. *J Pers Disord* 20, 247–260.
5. Spielberger CD (2010) *State-Trait Anxiety Inventory* (John Wiley & Sons, Hoboken, NJ).
6. Pizzagalli DA, Iosifescu D, Hallett LA, Ratner KG, Fava M (2008) Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task. *J Psychiat Res* 43, 76–87.
7. Preuschoff K, Bossaerts P, Quartz SR (2006) Neural differentiation of expected reward and risk in human subcortical structures. *Neuron* 51, 381–390.
8. Vlaev I, Seymour B, Dolan RJ, Chater N (2009) The price of pain and the value of suffering. *Psychol Sci* 20, 309–317.
9. Crockett MJ, Kurth-Nelson Z, Siegel JZ, Dayan P, and Dolan RJ (2014) Harm to others outweighs harm to self in moral decision making. *P Natl Acad Sci* 111, 17320–17325.
10. Schonberg T, Daw ND, Joel D, O'Doherty JP (2007) Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J Neurosci* 27, 12860–12867.
11. Li J, Schiller D, Schoenbaum G, Phelps EA, Daw, ND (2011) Differential roles of human striatum and amygdala in associative learning. *Nat Neurosci* 14, 1250–1252.
12. Boll S, Gamer M, Gluth S, Finsterbusch J, Büchel C (2013) Separate amygdala subregions signal surprise and predictiveness during associative fear learning in humans. *Eur J Neurosci* 37, 758–767.
13. Bishop CM (2006) *Pattern Recognition and Machine Learning* (Springer, Heidelberg, Germany).
14. Kass RE, Raftery AE (1995) *Bayes factors*. J Am Stat Assoc 90, 773–795.
15. Huys QJ, et al. (2012) Bonsai trees in your head: how the Pavlovian system sculpts goaldirected choices by pruning decision trees. *PLoS Comp Biol* 8, e1002410.
16. Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. *Neuroimage* 33, 493–504.

17. Helms G, Draganski B, Frackowiak R, Ashburner J, Weiskopf N (2009) Improved segmentation of deep brain grey matter structures using magnetization transfer (MT) parameter maps. *Neuroimage* 47, 194–198.

18. Weiskopf N, Helms G (2008). Multi-parameter mapping of the human brain at 1mm resolution in less than 20 minutes. In *Proceedings of the 16th Annual Meeting ISMRM*.

19. Helms G, Dathe H, Dechent P (2008) Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. *Magnet Reson Med* 59, 667–672.

20. Helms G, Dathe H, Kallenberg K, Dechent P (2008) High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. *Magnet Reson Med* 60, 1396–1407. 14

21. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441, 876–879.

22. Wittmann BC, Daw ND, Seymour B, Dolan RJ (2008) Striatal activity underlies novelty based choice in humans. *Neuron* 58, 967–973.

23. Pine A, Shiner T, Seymour B, Dolan RJ (2010) Dopamine, time, and impulsivity in humans. *J Neurosci* 30 8888–8896.

24. Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R (2012) Serotonin selectively modulates reward value in human decision-making. *J Neurosci* 32, 5833–5842.

25. Voon V, et al. (2010). Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron* 65, 135–142.

26. Hauser TU, Iannaccone R, Walitza S, Brandeis D, Brem S (2015) Cognitive flexibility in adolescence: Neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *NeuroImage* 104, 347–354.

27. Eldar E, Niv Y (2015) Interaction between emotional state and learning underlies mood instability. *Nat Comm* 6, 6149.

28. Hutton C, et al. (2011) The impact of physiological noise correction on fMRI at 7T. *Neuroimage* 57, 101–112.

29. Bach DR, Flandin G, Friston KJ, Dolan RJ (2010) Modelling event-related skin conductance responses. *Int J Psychophysiol* 75, 349–356.

30. Ashburner J, Friston KJ (2000) Voxel-based morphometry — the methods. *Neuroimage* 11, 805–821.

31. Kohavi RA (1995) Study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence, Vol. 2*.

32. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J Roy Stat Soc B* 58, 267–288.

33. Efron B (1987) Better bootstrap confidence intervals. *J Am Stat Assoc* 82, 171–185.

34. Edgington E, Onghena P (2007) *Randomization tests* (CRC Press, Boca Raton, FL).