Striatal structure and function predict individual biases in learning to avoid pain

Eran Eldar[a,b,1], Tobias U. Hauser[a,b], Peter Dayan[c,2], and Raymond J. Dolan[a,b,2]

[a]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom; [b]Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London WC1B 5EH, United Kingdom; and [c]Gatsby Computational Neuroscience Unit, University College London, London W1T 4JG, United Kingdom

Edited by Thomas D. Albright, The Salk Institute for Biological Studies, La Jolla, CA, and approved March 15, 2016 (received for review October 6, 2015)

Pain is an elemental inducer of avoidance. Here, we demonstrate that people differ in how they learn to avoid pain, with some individuals refraining from actions that resulted in painful outcomes, whereas others favor actions that helped prevent pain. These individual biases were best explained by differences in learning from outcome prediction errors and were associated with distinct forms of striatal responses to painful outcomes. Specifically, striatal responses to pain were modulated in a manner consistent with an aversive prediction error in individuals who learned predominantly from pain, whereas in individuals who learned predominantly from success in preventing pain, modulation was consistent with an appetitive prediction error. In contrast, striatal responses to success in preventing pain were consistent with an appetitive prediction error in both groups. Furthermore, variation in striatal structure, encompassing the region where pain prediction errors were expressed, predicted participants' predominant mode of learning, suggesting the observed learning biases may reflect stable individual traits. These results reveal functional and structural neural components underlying individual differences in avoidance learning, which may be important contributors to psychiatric disorders involving pathological harm avoidance behavior.

avoidance learning | pain | individual differences | striatum | prediction errors

Pain conveys vital feedback on our actions, informing us whether an action compromises our safety and should be avoided. However, learning what to avoid doing, rather than what to do, could lead to maladaptive passive risk-avoidant behavior. For instance, when learning to ski, an overreaction to a painful fall could render a person overly cautious and hinder progress in skill acquisition. Likewise, failed investments might lead to an overconservative passive financial strategy, whereas social rejection might engender reclusive behavior. In some individuals, such as those with avoidant personality disorders, refraining too much from potentially harmful actions can manifest as a stable personality trait (1).

However, an opposite tendency, to learn predominantly from successful actions that helped avoid harm, might lead to maladaptive active behavior. Thus, in soccer, sporadic success in preventing goals by diving to the left or right before seeing where a penalty kick is heading is sufficient for goalkeepers to overwhelmingly prefer this suboptimal active strategy, when in fact the optimal strategy is to passively stay put (2). In the extreme, excessive repetitive activity so as to avoid harm may constitute compulsivity, a debilitating feature of obsessive-compulsive disorder (3).

Complementing previous studies of learning about abstract outcomes (4, 5), we investigated individual biases in learning to avoid pain. To this end, we used a novel gambling task that probes how participants adjust their choices in response to painful electrical shocks and, additionally, how they adjust their choices in response to success in preventing shocks. In line with studies of reward learning (6–9), learning in our task was best explained as driven by an outcome prediction error that reflects the difference between expected and actual outcomes. Consistent with the expression of such a teaching signal, blood-oxygen level-dependent (BOLD) responses to outcomes in the striatum were modulated by expectation. However, striatal response to shocks were qualitatively different in negative learners (i.e., those who predominantly learned from shocks) compared with positive learners (i.e., those who predominantly learned from success in avoiding shocks). Specifically, striatal activity was consistent with an aversive prediction error signal in negative learners and with an appetitive prediction error signal in positive learners. The degree to which a participant tended to learn from success in avoiding than experiencing shocks was predicted by the structure of a participants' striatum, specifically by higher gray matter density where the response to shocks was consistent with a prediction error signal.

Results

Individual Biases in Learning from Pain and Its Prevention. To test for individual differences in learning to avoid pain, we tasked 41 participants to play a card game in which their goal was to minimize the number of painful electrical shocks they might receive. Participants could avoid shocks by gambling that the number they were about to draw will be higher than the number the computer had drawn. An unsuccessful gamble resulted in shock and a successful gamble led to its avoidance. Alternatively, participants could always decline the gamble and opt for a fixed 50% known probability of receiving a shock (Fig. 1*A*). Importantly, participants played with three different decks of cards and

Significance

Our ability to learn how to avoid harm is critical for maintaining physical and mental health. However, excessive harm avoidance can be maladaptive, as evident in psychiatric disorders such as avoidant personality disorder and obsessive-compulsive disorder. We therefore investigated the neural factors underlying individual imbalances in harm avoidance behavior. Our findings show that such imbalances can be predicted by the function and structure of an individual's striatum, a brain region that is critical for goal-directed decisionmaking. Moreover, the neural signals expressed in this region revealed key processes through which individuals learn to avoid harm. These findings highlight a neural basis for imbalanced harm avoidance behavior, extreme forms of which may contribute to psychiatric pathology.

Author contributions: E.E., P.D., and R.J.D. designed research; E.E. and T.U.H. performed research; E.E. analyzed data; and E.E., T.U.H., P.D., and R.J.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: e.eldar@ucl.ac.uk.

[2]P.D. and R.J.D. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1519829113/-/DCSupplemental.

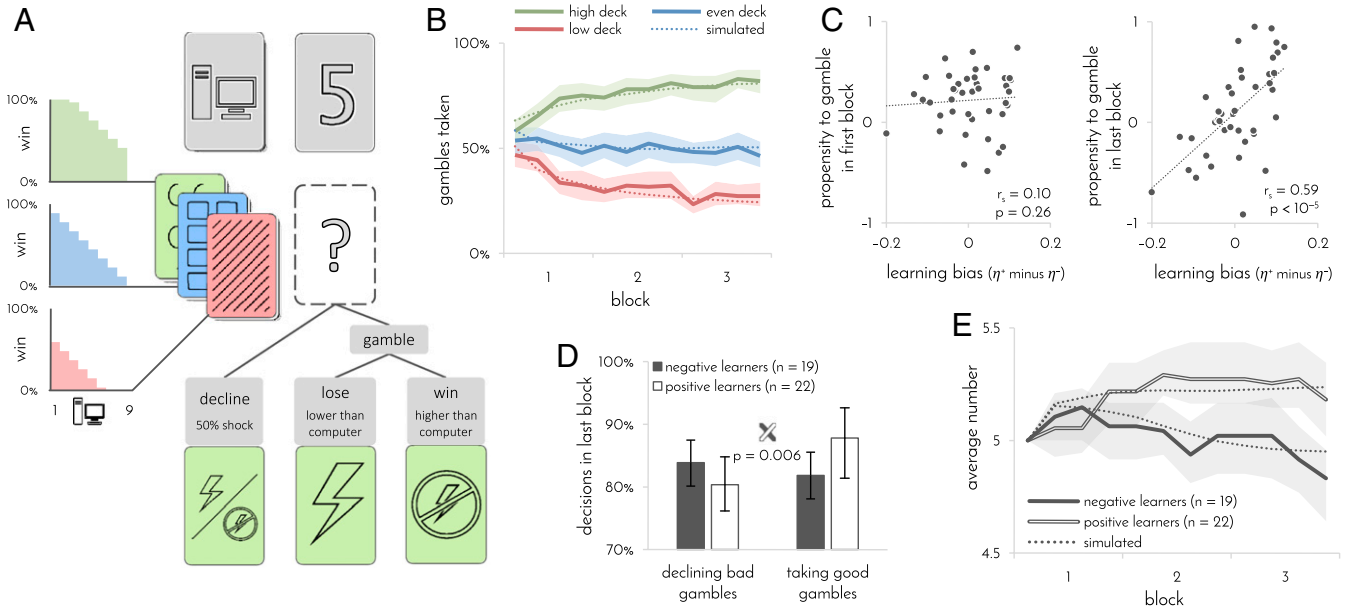

Fig. 1. Experimental design and learning performance. $n = 41$ participants. (*A*) Experimental design. On each trial, participants were presented with one of three possible decks and a number between 1 and 9 drawn by the computer. If participants decided to gamble, a shock was delivered only if the number that they drew was lower than the computer's number. Participants were only informed whether they won or lost the gamble, not which number they drew. Participants had to learn by trial and error how likely gambles were to succeed with each of the three decks. One deck contained a uniform distribution of numbers between 1 and 9 (even deck), one deck contained more 1's (low deck), making gambles 30% less likely to succeed, and one deck contained more 9's (high deck), making gambles 30% more likely to succeed. Opting to decline the gamble resulted in a 50% probability of shock regardless of which numbers were drawn by the computer. (*B*) Gambles taken with each deck as a function of time. Percentages were computed separately for each set of 15 contiguous trials (4 sets/60 trials per block). (*C*) Participants' propensity to gamble in first (*Left*) and last (*Right*) blocks of trials as a function of learning bias. Propensity to gamble was computed by regressing out the effects of deck and computer's number on participant's choices using logistic regression. The numbers 1 and –1 correspond to always and never gamble, respectively. Learning bias was inferred from a participant's choices using the learning model. (*D*) Proportion of bad gambles that were declined and good gambles that were taken in last block of trials as a function of learning bias. Participants with a positive learning bias (positive learners) declined fewer bad gambles and took more good gambles than participants with a negative learning bias (negative learners). Gambles were defined as good or bad based on probability of winning (good: >50%; bad <50%). Error bars: 95% bootstrap CI. (*E*) Average number drawn by computer as a function of time and learning bias. To maintain participants at a 50% gambling rate, numbers increased for positive learners and decreased for negative learners. In *B* and *E*, dotted line indicates simulated task performance of learning model. Shaded areas: 95% bootstrap CI.

had to learn by trial and error how likely a gamble was to be successful with each deck.

In principle, participants can acquire information about the decks from both successful and unsuccessful gambles. Indeed, we observed from their behavior that as they gained experience with the three decks, their willingness to gamble with each deck differed (Fig. 1*B*). A more in-depth analysis indicated that they did not learn from the two types of outcomes to the same degree. Thus, the learning algorithm that best explained participants' choices included two different learning rates, one for learning from shock outcomes (η^-) and one for learning from no-shock outcomes (η^+; log Bayes factor compared with algorithm with a single learning rate = 27.7). In this algorithm, the two learning rates determine the degree to which the two types of outcomes impact on subsequent expectations of gambling with each of the decks, and these expectations in combination with the numbers drawn by the computer determine whether subsequent gambles are taken or declined. The algorithm also accounts for each participant's baseline propensity to take gambles. Learning in the favored algorithm is weighted by associability (10, 11), and the choices made for each deck tend to persist (7) (see *SI Appendix* for details of all learning algorithms and a validation of the model comparison procedure).

Further analysis showed that the difference between the two learning rates captured significant interindividual variation (log Bayes factor of algorithm with two learning rates per participant compared with algorithm with one average learning rate per participant and a group parameter for difference between the learning rates = 20.1). On this basis, we next computed each participant's learning bias as the difference between the two learning rates that best fitted the participant's choices (η^+ minus η^-). This bias reflects the degree to which a participant learned what gambles to take (because they resulted in no-shock outcomes) rather than what gambles to avoid (because they resulted in shocks). A positive learning bias ($\eta^+ > \eta^-$) entails that a propensity to gamble will emerge as the participant is learning from outcomes, whereas a negative learning bias ($\eta^+ < \eta^-$) should engender a propensity to decline a gamble (9). A comparison of participant's raw propensity to gamble at the beginning and end of the experiment confirmed this expectation (Fig. 1*C*). Consequently, by the end of the experiment, participants with a positive learning bias came to take more good gambles, but also decline fewer bad gambles than participants with a negative learning bias (Fig. 1*D*).

To ensure that all participants nevertheless gambled at a similar rate and, thus, received roughly equal amounts of information about the decks, we increased or decreased the numbers the computer drew according to the participants' own gambling rate, while ensuring that the three decks are always matched against similar computer's numbers. Thus, because of their propensity to gamble, positive learners ended up playing against high numbers, whereas negative learners ended up playing against low numbers [Fig. 1*E*; mean difference 0.36, confidence interval (CI) 0.15–0.6, $P = 0.001$, bootstrap test; as a result, positive learners received a slightly higher number of shocks (mean 74.5, CI 72.7–76.5) than negative learners (mean 72.0, CI 70.3–73.3)]. In sum, participants who learned more from painful outcomes developed a propensity to avoid gambling, whereas

participants who learned more from success in preventing pain developed a propensity to gamble.

Striatal Response to Pain Underlies Learning Bias. Behaviorally, learning in our task was best explained as driven by outcome prediction errors (*SI Appendix*, Figs. S1 and S2). Such prediction errors are known to be expressed in the striatum (7–9, 12, 13), and, thus, we focused on examining participants' striatal activity in response to task outcomes. BOLD responses in the striatum (and elsewhere) were overwhelmingly stronger in response to shock than to no-shock outcomes (*SI Appendix*, Fig. S3), likely reflecting sensory and affective processing of the shocks. Thus, to specifically test for prediction error signals, we examined the response to each type of outcome (shock and no-shock) separately, testing whether these responses were modulated by expectation.

We found that BOLD responses to no-shock outcomes in the head of the caudate nucleus were negatively modulated by expectation (derived by using the learning algorithm that best fitted participants' behavior; Fig. 2*A*), consistent with an appetitive prediction error. This signal was not explained away by the effect of associability [$P < 0.05$, familywise error (FWE) corrected]. In contrast, when accounting for prediction errors, we did not find evidence for an associability signal in the striatum or elsewhere ($P > 0.05$ FWE corrected). Moreover, when decomposing the prediction error (PE) signal into the two components that affect a participant's expectation, the one reflecting previous experience with the current deck and the other reflecting the number drawn by the computer, we found a significant effect within the identified striatal region in response to both PE components {Fig. 2*B*; peak Montreal Neurological Institute (MNI) coordinates: deck expectation [−10 6 6], $t_{40} = -4.0$, $P = 0.04$; number [−8 6 2], $t_{40} = 4.2$, $P = 0.03$; P values are FWE corrected across the striatum}.

Whereas responses to no-shock outcomes were similarly modulated by expectations in positive and negative learners, the response to shock outcomes was qualitatively different in the two groups. No prediction error signal was evident in striatal responses to shocks across both groups of participants ($P > 0.05$, FWE small-volume corrected). However, examining the two groups separately revealed that the BOLD response to shocks was positively modulated by expectation in negative learners and negatively modulated by expectation in positive learners (Fig. 2*C*; responses measured in the same area where the response to no-shock outcomes was modulated by expectations; correlation between learning bias and average general linear model (GLM) coefficient r values =0.4, $P = 0.005$, permutation test; partial correlation accounting for number of shocks received: $\rho_s = 0.38$, $P = 0.007$, permutation test; a similar difference between positive and negative learners was evident in the caudate in a voxel-based analysis across the striatum—peak MNI coordinates [−10 −2 18], $P = 0.03$ small-volume corrected, permutation test—as well as when averaging across the whole caudate, $P = 0.05$, Bonferroni corrected for the three subregions of the striatum). In other words, in response to shocks, striatal activity in negative learners was consistent with an aversively signed prediction error, as seen in previous studies of pain prediction (14, 15). By contrast, in positive learners, activity was consistent with an appetitively signed prediction error, as in previous work on pain-relief prediction (16). Importantly, this difference between the groups was not associated with a difference in the mean response to shock compared with no-shock outcomes ($P = 0.60$, bootstrap test, GLM; *SI Appendix*, Fig. S3 *C* and *D*). These results suggest that, across both types of outcomes, positive learners express an appetitively signed prediction error, whereas negative learners express the absolute value of the prediction error.

If striatal prediction errors indeed drive avoidance learning, then the modulation of striatal responses to outcomes by expectation should predict whether a participant subsequently developed a propensity to gamble. Modulation consistent with an appetitive prediction error (signaling which gambles should

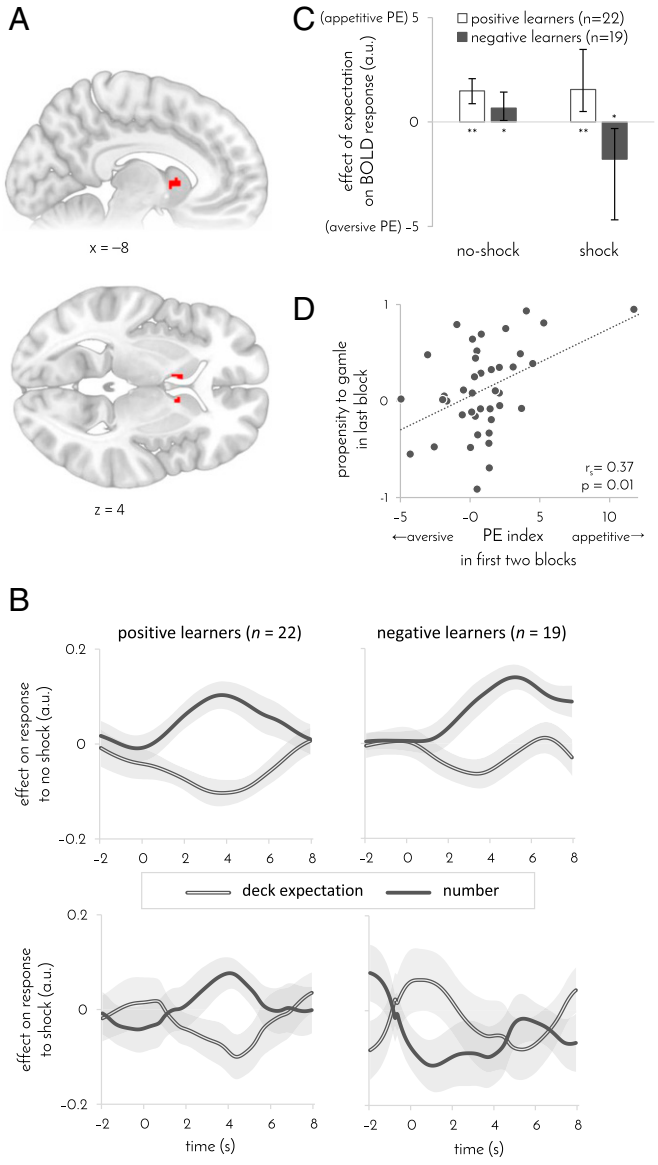

Fig. 2. PE signaling in the striatum. $n = 41$ participants. (*A*) Striatal area where BOLD response to no-shock outcomes was modulated by expectations. Map is masked and FWE corrected for volume of striatum ($P < 0.05$, GLM). Extent: 22 voxels. Peak MNI coordinates: [−6 10 2] $t_{40} = 4.5$, corrected $P = 0.01$; [10 8 4] $t_{40} = 4.7$, corrected $P = 0.007$. x and z denote Montreal Neurological Institute (MNI) coordinates. (*B*) The two components of the prediction error. Effects of deck-based expectation and the number drawn by the computer on BOLD response to shock (*Bottom*) and no shock (*Top*) outcomes in the striatal ROI identified in *A*. Results are shown separately for positive (*Left*) and negative (*Right*) learners. A higher deck expectation and a lower number indicate lower expectation of a shock, and thus, an appetitive PE is consistent with a rise in the effect of number and a dip in the effect of deck expectation on the response to the outcomes. The converse pattern is consistent with an aversive PE. Time 0 indicates outcome onset. Shaded area: SEM. (*C*) Effect of expectations on BOLD response to no-shock and shock outcomes as a function learning bias. Positive values indicate an effect that is consistent with an appetitive PE. $^*P < 0.05$, $^{**}P < 0.005$, error bars: 95% bootstrap CI. (*D*) Propensity to gamble in the last task block as a function of a participant's PE index, computed as the average effect of expectation on striatal response across both types of outcomes. An appetitive PE index predicted subsequent propensity to gamble (computed as in Fig. 1*C*). GLM coefficients in *C* and *D* were taken from the striatal area identified in *A*.

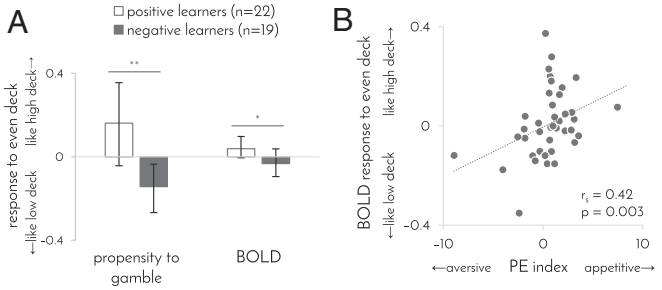

Fig. 3. Behavioral and neural responses to the even deck. (*A*) Propensity to gamble and BOLD response to even deck as a function of learning bias. Propensity to gamble was computed as in Fig. 1*C* but exclusively for the even deck. By contrast, all participants avoided gambling with the low deck (propensity to gamble –0.42, CI –0.54 to –0.30) and favored gambling with the high deck (propensity to gamble 0.72, CI 0.60–0.80). BOLD response of 1 indicates that the response to the even deck was identical to the response to the high deck, whereas a value of –1 indicates that it was identical to the response to the low deck. Similarity of BOLD responses was computed as the Euclidian distance between the two responses' GLM coefficients across all gray matter. Error bars: 95% bootstrap confidence intervals, $*P = 0.05$, $**P = 0.01$, permutation test. (*B*) BOLD response to even deck, compared with high and low decks, as a function of striatal PE index. PE index taken from Fig. 2*D*, and BOLD response was computed as in *A*.

be taken) should lead to a progressively increasing propensity to gamble, whereas modulation consistent with an aversive prediction error (signaling which gambles should be declined) should lead to an increasing propensity to decline. In agreement with this prediction, the degree to which striatal activity was consistent with an appetitive (rather than aversive) prediction error in the first two experimental blocks correlated with participants' propensity to gamble in the last block, even when controlling for striatal activity in the last block (partial $r_s = 0.37$, $P = 0.01$, permutation test). By contrast, the propensity to gamble in the first block was not correlated with the striatal prediction error signaling in the second and third block (Fig. 2*C*; partial $r_s = 0.05$, $P = 0.37$, permutation test, controlling for the prediction error difference in the first block). Thus, we can conclude that striatal responses to outcomes were predictive of whether a participant would subsequently develop a propensity to accept or decline gambles.

Striatal Signals Prefigure Learned Neural Responses. Differences between participants in how they learn should engender not only differences in behavior (i.e., propensity to gamble), but also differences in neural responses to stimuli or contexts about which they are learning (i.e., decks). Such an effect can be expected to be particularly evident for the even deck, because the even deck provided a context where negative learners came to avoid gambling and positive learners came to favor gambling ($P = 0.01$, permutation test; Fig. 3*A*). In simple terms, negative learners came to regard the even deck as a low deck, whereas positive learners came to regard it as a high deck. This result, however, is based on the same behavioral data used to estimate participants' learning biases. Thus, to test the same effect using an alternative independent measure, we examined each participant's BOLD responses to the three decks.

In keeping with the behavioral result, we found that BOLD responses to the even deck were more similar to BOLD responses to the high deck in positive learners and more similar to BOLD responses to the low deck in negative learners (Fig. 3*A*). Moreover, the differences between participants in their representation of the even deck correlated with the degree to which striatal activity was consistent with an appetitive (rather than aversive) prediction errors (Fig. 3*B*; $r_s = 0.42$, $P = 0.003$, permutation test). In fact, striatal activity in the first two experimental blocks predicted subsequent BOLD response to the even deck, measured during the last block (partial $r_s = 0.34$, controlling for striatal activity in last block, $P = 0.02$, permutation test), whereas there was no evidence for the reverse relationship, between the BOLD response to the even deck in the first block and striatal activity in the last two blocks (partial $r_s = 0.03$, controlling for striatal activity in first block, $P = 0.42$, permutation test). Together, these results suggest that differences in prediction error signaling between positive and negative learners shaped their subsequent neural responses to the stimuli about which participants were learning.

Learning Bias Is Predicted by the Structure of the Striatum. If individual differences in learning and striatal function, evident in our task, reflect stable individual traits, then we might also expect underlying differences in striatal structure. Indeed, a positive learning bias was associated with higher gray matter density in the head of the caudate (extent: 1 voxel, MNI coordinates [–17 23 6], $P = 0.03$ FWE small-volume corrected), as measured by using voxel-based morphometry (17) from magnetic transfer anatomical images that are particularly suited for measuring subcortical structures (18). To test whether the overall structure of participants' striatum predicted their learning biases, we used the gray matter density of each participant's 6,315 striatal voxels to predict each participant's learning bias, by means of a regularized regression model that reflected the relationship between learning bias and striatal gray matter density in other participants (i.e., using fivefold cross-validation). Learning biases predicted by striatal structure significantly correlated with the actual biases derived from participants' behavior (Fig. 4*A*; $r = 0.58$, $P = 0.001$, permutation test), with a positive learning bias predicted for 17 of the 22 positive learners, and a negative learning bias predicted for 13 of the 19 negative learners. Furthermore, a predictive relationship between gray matter density and learning bias was mostly negative throughout the putamen and accumbens and mostly positive in the caudate, and thus, it did not involve differences in overall striatal volume (Fig. 4*B*; mean predictive coefficient 0.0, CI –0.0002–0.0001). This predictive spatial pattern was not random, but presented a striking match with the

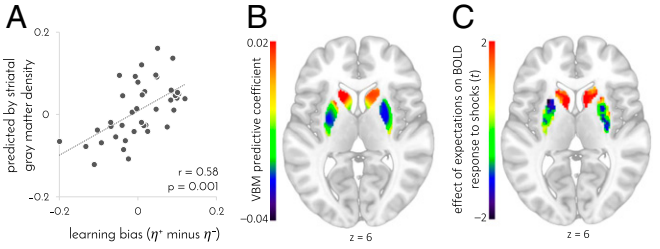

Fig. 4. Striatal gray matter density predicts learning bias. $n = 41$ participants. (*A*) Learning bias predicted by gray matter density in the 6,315 voxels of the striatum as a function of the true learning bias inferred from participants' choices. Learning biases were predicted from gray matter density by using cross-validated regularized linear regression. (*B*) Gray matter density coefficients used to predict learning bias. To create the map, predictive coefficients were averaged across participants and generalized across the striatum by assigning a fraction of each coefficient to each voxel proportionally to the gray matter-density variance shared between that voxel and the coefficient's designated voxel. (*C*) Representation of expectations in the response to shocks. *t* values were computed by using a group-level GLM that included both negative learners, whose BOLD response was regressed against aversively signed prediction errors, and positive learners, whose BOLD response was regressed against appetitively signed prediction errors. There were no significant differences between positive and negative learners within the striatum ($P > 0.5$, FWE small-volume corrected). The map is masked for the volume of the striatum and not thresholded. *z* value in *B* and *C* denotes MNI coordinate.

area where BOLD responses to shocks were modulated by expectation (Fig. 4*C*; regression of individual expectation GLM coefficients on group gray matter-density predictive coefficients: mean 0.44, CI 0.22–0.77, $P = 10^{-5}$, bootstrap test). That is, a more positive learning bias was predicted by higher gray matter density where responses to shocks were consistent with a (appetitive or aversive) prediction error signal, and lower gray matter density in areas that showed no such signal.

Amygdala Specifically Involved in Learning from Pain. A large body of animal and human work implicates the amygdala and periaqueductal gray (PAG) in learning from aversive outcomes and, in particular, in generating aversive prediction errors (19–24). BOLD responses to outcomes in these areas suggests both areas were involved in learning in our task, albeit in different ways. Responses to shock and no-shock outcomes in the PAG were modulated by expectations in the same way as in our striatal ROI (*SI Appendix*, Fig. S3*E*; although we note that the ability of fMRI to discern PAG signals from neighboring structures is limited; ref. 25). In contrast, the amygdala showed no significant modulation in the response to no-shock outcomes, but its response to shocks was consistent with an aversive prediction error across the whole group (extent: 1 voxel, MNI coordinates [28 –6 –18], $t_{40} = 2.8$, $P = 0.04$ corrected for the volume of the amygdala). This latter effect was particularly pronounced in participants with a negative learning bias (Fig. 5*A*), mirroring the modulation of striatal and PAG activity in these same participants.

Interestingly, a positive learning bias was associated with a stronger response to shock than to no-shock outcomes in right posterior insula (extent: three voxels, peak MNI coordinates [46 –32 20], $P = 0.04$ FWE-corrected for all voxels that significantly responded to shocks, shown in *SI Appendix*, Fig. S3*A*). The insula is thought to feed information about salient, painful, and aversive events to amygdala and striatum (26–28). Therefore, we next examined functional connectivity between this area and the striatal and amygdala regions in which response to outcomes was modulated by expectation. Positive learners showed significant functional connectivity between the insula and striatal regions, whereas negative learners showed significant functional connectivity between the insula and amygdala regions (correlation of learning bias with difference between striatal and amygdala functional connectivity: $r_s = 0.42$, $P = 0.005$; Fig. 5*B*). Taken together, these results are consistent with previous suggestions that the amygdala is exclusively involved in learning from increases in aversive outcomes, whereas the striatum and PAG also partake in learning from decreases in aversive outcomes (21, 23, 29).

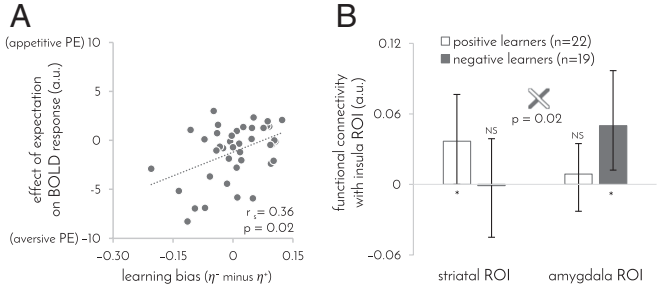

Fig. 5. Individual learning biases outside of the striatum. (*A*) Effect of expectation on BOLD response to shocks as a function of learning bias in the amygdala region where this effect was significant (MNI coordinates [28 –6 –18]). Responses were most consistent with an aversive prediction error in participants who mostly learned from shock outcomes. $n = 41$ participants. (*B*) Functional connectivity between striatal (Fig. 2*A*) and amygdala (*A*) ROIs, and the insula area in which the response to shocks correlated with learning bias (MNI coordinates [46 –32 20]). Error bars: 95% bootstrap CI, $^*P \leq 0.05$, NS: $P > 0.1$.

Discussion

We demonstrate that the two ways through which one can learn to avoid harm are used to different degrees by different individuals. In negative learners—those who primarily learned from being shocked and, thus, developed a propensity to avoid gambles—dorsal striatal and amygdalar responses to shocks were consistent with an aversive prediction error. In contrast, in positive learners—those who primarily learned from their success in preventing shocks and, thus, developed a propensity to gamble—the dorsal striatum's response to shocks was consistent with an appetitive prediction error. This difference in striatal responses to outcomes anticipated observed differences in learned behavior, and in the neural responses to stimuli about which participants were learning. Participants' learning bias was also predicted by the structure of their striatum, indicating that learning biases in our task reflected, at least in part, stable individual traits. Together, the findings reveal neural underpinnings of an elementary behavioral trait that predicts whether an individual learns predominantly what to do to prevent harm or what to avoid doing.

A multitude of animal and human studies implicate the striatum in learning about aversive outcomes (12, 13, 30–34). Striatal areas, including the caudate region identified in our study, have been shown to represent prediction error signals in both classical conditioning (14–16) and instrumental pain avoidance learning (35, 36). However, individual differences in the expression of these prediction error signals have not been studied to our knowledge. The present study was designed to assess the degree to which participants learn what actions have painful outcomes compared with what actions help avoid pain. The latter type of learning, defined in the animal literature as active avoidance learning, is particularly interesting, because it specifies that the very absence of a shock is reinforcing (37). Indeed, our results show that striatal response to outcomes in participants who were biased in favor of active avoidance learning mimicked striatal responses typically seen in studies of reward learning, with no-shock outcomes cast as reward and shock outcomes as reward omission. In contrast, in participants biased in favor of passive avoidance learning (i.e., learning what gambles should be avoided), striatal response to painful outcomes was consistent with an aversive prediction error, as seen in fear conditioning (14).

Thus, our results show that striatal response to pain is consistent with an appetitive prediction error in some individuals and with an aversive prediction error in others. In contrast, striatal responses to successful prevention of pain seem broadly consistent with an appetitive prediction error. Put another way, while some individuals represent an appetitively signed prediction error in response to both types of outcomes, others represent an unsigned prediction error (38) (aversively signed in response to pain and appetitively signed in response to pain prevention). That said, we note that in our experiment, the response to pain was generally stronger than the response to pain prevention, which is inconsistent with a purely appetitive prediction error, although this stronger response could reflect affective and sensory processing of the shocks.

Our findings concur with a view of the striatum as involved in processing both appetitive and aversive outcomes (26, 39). The amygdala, in contrast, was involved in our study solely in processing aversive outcomes, but not their omission. This involvement of the amygdala was most evident in participants who primarily learned from shock outcomes, and underscored by greater functional connectivity with the insula, a region with an established role in processing salient and aversive outcomes (40). These results strongly concur with previous work, indicating that learning from aversive outcomes engages the amygdala, whereas learning from success in avoiding aversive outcomes involves inhibition of the amygdala and activation of the striatum (29, 41-43). That said, we note that animal studies that were able to examine subregions of the striatum and amygdala with greater spatial

resolution reveal a more complex picture (44, 45). Finally, whereas functional connectivity between medial prefrontal cortex, amygdala, and striatum has been shown to mediate avoidance learning (43), it did not mediate learning biases in our task, suggesting that such connectivity might be equally involved in learning from increases and decreases in aversive outcomes.

Several reports have linked variations in the structure of the striatum to individual differences in healthy and pathological decision-making behavior (46, 47) and to the expression of certain pain disorders (48, 49). Of particular relevance to the present study is the observation that obsessive-compulsive disorder (OCD), which features compulsive harm-avoidance behavior, is associated with higher gray matter density in the putamen (50). In our study, higher gray matter density in the putamen (and lower gray matter density in the caudate) predicted better learning from shocks and poorer learning from success in avoiding shocks. It is possible that such insensitivity to safety signals might engender excessively persistent harm avoidance behaviors, which in healthy individuals normally terminate when safety is attained. Thus, this finding raises the interesting possibility that failure to adjust to success in harm avoidance may contribute to compulsivity in OCD.

In conclusion, we describe for the first time to our knowledge individual biases in learning from actual painful outcomes on the one hand and from their prevention on the other. These biases are associated with qualitative differences in striatal prediction error signaling and predicted by differences in striatal structure. Further research should reveal how these functional and structural characteristics map onto psychiatric disorders that feature imbalanced harm avoidance behavior.

Materials and Methods

The experimental protocol was approved by the University of College London local research ethics committee, and informed consent was obtained from all participants. Electrical stimulation was individually titrated to induce a moderate subjective pain level. Participants performed the experiment while being scanned in a Siemens Trio 3T MRI scanner. See *SI Appendix* for further details of the experiment, modeling, and functional MRI procedures.

ACKNOWLEDGMENTS. This work was funded by Wellcome Trust's Cambridge-University College London Mental Health and Neurosciences Network Grant 095844/Z/11/Z (to E.E. and R.J.D.), Wellcome Trust Investigator Award 098362/Z/12/Z (to R.J.D.), the Gatsby Charitable Foundation (P.D.), and Swiss National Science Foundation Grant 151641 (to T.U.H.).

1. American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders* (Am Psychiatr Assoc, Washington, DC), 5th Ed.
2. Bar-Eli M, Azar OH, Ritov I, Keidar-Levin Y, Schein G (2007) Action bias among elite soccer goalkeepers: The case of penalty kicks. *J Econ Psychol* 28:606–621.
3. Abramowitz JS, Taylor S, McKay D (2009) Obsessive-compulsive disorder. *Lancet* 374(9688):491–499.
4. Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science* 306(5703):1940–1943.
5. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci USA* 104(41):16311–16316.
6. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442(7106): 1042–1045.
7. Schönberg 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(47):12860–12867.
8. Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J Neurosci* 28(22):5623–5630.
9. Niv Y, Edlund JA, Dayan P, O'Doherty JP (2012) Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *J Neurosci* 32(2):551–562.
10. Li J, Schiller D, Schoenbaum G, Phelps EA, Daw ND (2011) Differential roles of human striatum and amygdala in associative learning. *Nat Neurosci* 14(10):1250–1252.
11. 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(5):758–767.
12. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482(7383):85–88.
13. Delgado MR, Li J, Schiller D, Phelps EA (2008) The role of the striatum in aversive learning and aversive prediction errors. *Philos Trans R Soc Lond B Biol Sci* 363(1511):3787–3800.
14. Seymour B, et al. (2004) Temporal difference models describe higher-order learning in humans. *Nature* 429(6992):664–667.
15. Schiller D, Levy I, Niv Y, LeDoux JE, Phelps EA (2008) From fear to safety and back: Reversal of fear in the human brain. *J Neurosci* 28(45):11517–11525.
16. Seymour B, et al. (2005) Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nat Neurosci* 8(9):1234–1240.
17. Ashburner J, Friston KJ (2000) Voxel-based morphometry–the methods. *Neuroimage* 11(6 Pt 1):805–821.
18. 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(1):194–198.
19. Phillips RG, LeDoux JE (1992) Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behav Neurosci* 106(2):274–285.
20. Kim JJ, Rison RA, Fanselow MS (1993) Effects of amygdala, hippocampus, and periaqueductal gray lesions on short- and long-term contextual fear. *Behav Neurosci* 107(6):1093–1098.
21. Cole S, McNally GP (2008) Complementary roles for amygdala and periaqueductal gray in temporal-difference fear learning. *Learn Mem* 16(1):1–7.
22. Johansen JP, Tarpley JW, LeDoux JE, Blair HT (2010) Neural substrates for expectation-modulated fear learning in the amygdala and periaqueductal gray. *Nat Neurosci* 13(8):979–986.
23. Yacubian J, et al. (2006) Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *J Neurosci* 26(37):9530–9537.
24. Mobbs D, et al. (2007) When fear is near: Threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science* 317(5841):1079–1083.
25. Linnman C, Moulton EA, Barmettler G, Becerra L, Borsook D (2012) Neuroimaging of the periaqueductal gray: State of the field. *Neuroimage* 60(1):505–522.
26. Navratilova E, Porreca F (2014) Reward and motivation in pain and pain relief. *Nat Neurosci* 17(10):1304–1312.
27. Bushnell MC, Čeko M, Low LA (2013) Cognitive and emotional control of pain and its disruption in chronic pain. *Nat Rev Neurosci* 14(7):502–511.
28. Leong JK, Pestilli F, Wu CC, Samanez-Larkin GR, Knutson B (2016) White-matter tract connecting anterior insula to nucleus accumbens correlates with reduced preference for positively skewed gambles. *Neuron* 89(1):63–69.
29. Rogan MT, Leon KS, Perez DL, Kandel ER (2005) Distinct neural signatures for safety and danger in the amygdala and striatum of the mouse. *Neuron* 46(2):309–320.
30. Salamone JD (1994) The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. *Behav Brain Res* 61(2):117–133.
31. Pezze MA, Feldon J (2004) Mesolimbic dopaminergic pathways in fear conditioning. *Prog Neurobiol* 74(5):301–320.
32. McNally GP, Westbrook RF (2006) Predicting danger: The nature, consequences, and neural mechanisms of predictive fear learning. *Learn Mem* 13(3):245–253.
33. Tom SM, Fox CR, Trepel C, Poldrack RA (2007) The neural basis of loss aversion in decision-making under risk. *Science* 315(5811):515–518.
34. Bolstad I, et al. (2013) Aversive event anticipation affects connectivity between the ventral striatum and the orbitofrontal cortex in an fMRI avoidance task. *PLoS One* 8(6):e68494.
35. Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R (2012) Serotonin selectively modulates reward value in human decision-making. *J Neurosci* 32(17):5833–5842.
36. Roy M, et al. (2014) Representation of aversive prediction errors in the human periaqueductal gray. *Nat Neurosci* 17(11):1607–1612.
37. Mowrer OH, Lamoreaux RR (1946) Fear as an intervening variable in avoidance conditioning. *J Comp Psychol* 39:29–50.
38. Bromberg-Martin ES, Matsumoto M, Hikosaka O (2010) Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron* 68(5):815–834.
39. Ramirez F, Moscarello JM, LeDoux JE, Sears RM (2015) Active avoidance requires a serial basal amygdala to nucleus accumbens shell circuit. *J Neurosci* 35(8):3470–3477.
40. Menon V, Uddin LQ (2010) Saliency, switching, attention and control: A network model of insula function. *Brain Struct Funct* 214(5-6):655–667.
41. Moscarello JM, LeDoux JE (2013) Active avoidance learning requires prefrontal suppression of amygdala-mediated defensive reactions. *J Neurosci* 33(9):3815–3823.
42. Choi JS, Cain CK, LeDoux JE (2010) The role of amygdala nuclei in the expression of auditory signaled two-way active avoidance in rats. *Learn Mem* 17(3):139–147.
43. Collins KA, Mendelsohn A, Cain CK, Schiller D (2014) Taking action in the face of threat: Neural synchronization predicts adaptive coping. *J Neurosci* 34(44):14733–14738.
44. Badrinarayan A, et al. (2012) Aversive stimuli differentially modulate real-time dopamine transmission dynamics within the nucleus accumbens core and shell. *J Neurosci* 32(45):15779–15790.
45. Martinez RC, et al. (2013) Active vs. reactive threat responding is associated with differential c-Fos expression in specific regions of amygdala and prefrontal cortex. *Learn Mem* 20(8):446–452.
46. van den Bos W, Rodriguez CA, Schweitzer JB, McClure SM (2014) Connectivity strength of dissociable striatal tracts predict individual differences in temporal discounting. *J Neurosci* 34(31):10298–10310.
47. Kreitzer AC, Malenka RC (2008) Striatal plasticity and basal ganglia circuit function. *Neuron* 60(4):543–554.
48. Apkarian AV, Hashmi JA, Baliki MN (2011) Pain and the brain: Specificity and plasticity of the brain in clinical chronic pain. *Pain* 152(3, Suppl):S49–S64.
49. Schmidt-Wilcke T, et al. (2007) Striatal grey matter increase in patients suffering from fibromyalgia–a voxel-based morphometry study. *Pain* 132(Suppl 1):S109–S116.
50. Rotge JY, et al. (2010) Gray matter alterations in obsessive-compulsive disorder: An anatomic likelihood estimation meta-analysis. *Neuropsychopharmacology* 35(3):686–691.

SI Appendix

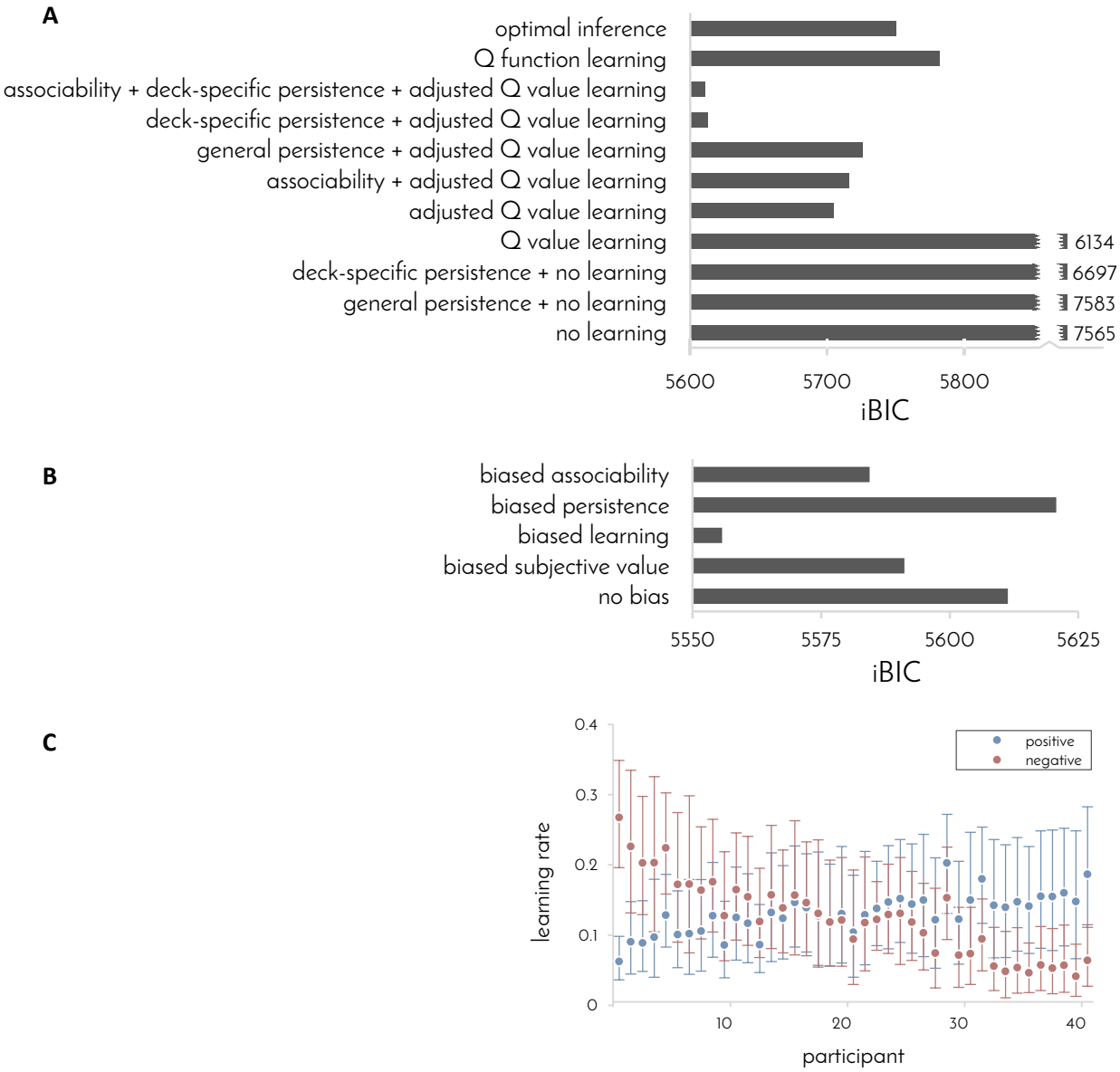

Figure S1. Model comparison and parameter fitting. **(A)** Eleven different learning algorithms were fitted to participants' behavior. Goodness of fit was computed using the integrated Bayesian Information Criterion (iBIC[15]). A difference larger than 10 constitutes very strong evidence in favor of 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).

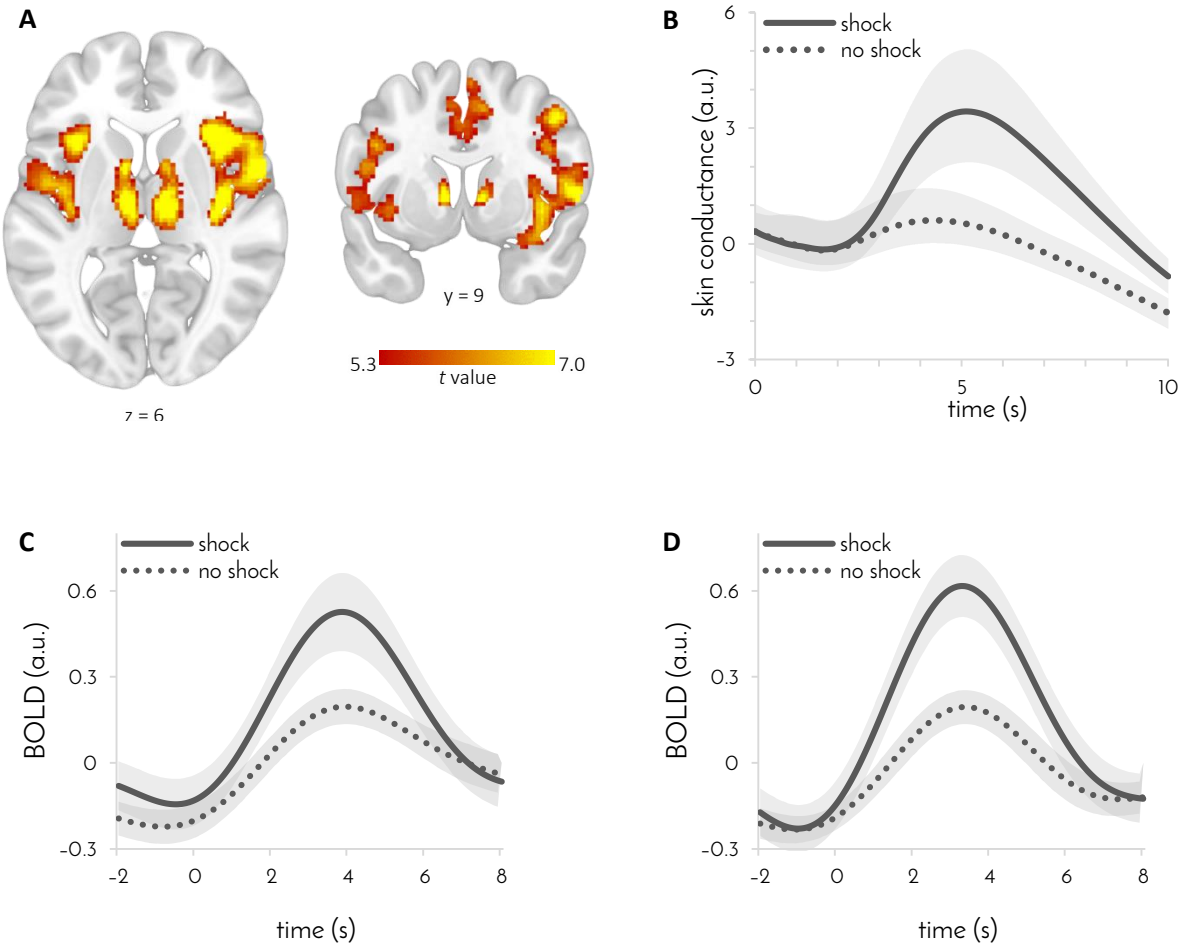

Figure S3. Further analyses of BOLD and skin conductance responses to outcomes. Related to Figure 2. *n* = 41 participants. **(A)** BOLD response to shocks as compared to no-shock gamble outcomes (*n* = 41 participants, $p < 0.05$ FWE corrected). Activated areas include the insula and adjacent somatosensory 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).