Functional specialization within the striatum along both the dorsal/ventral and anterior/posterior axes during associative learning via reward and punishment

Aaron T. Mattfeld, Mark A. Gluck and Craig E.L. Stark

Supplemental Material	http://learnmem.cshlp.org/content/suppl/2011/10/21/18.11.703.DC1.html
References	This article cites 60 articles, 16 of which can be accessed free at: http://learnmem.cshlp.org/content/18/11/703.full.html#ref-list-1
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Research

Functional specialization within the striatum along both the dorsal/ventral and anterior/posterior axes during associative learning via reward and punishment

Aaron T. Mattfeld,[1,2] Mark A. Gluck,[3] and Craig E.L. Stark[1,2,4]

[1]Department of Neurobiology and Behavior, University of California, Irvine, California 92697, USA; [2]Center for Neurobiology of Learning and Memory, Irvine, California 92697, USA; [3]Center for Molecular and Behavioral Neuroscience, Rutgers University-Newark, Newark, New Jersey 07102, USA

The goal of the present study was to elucidate the role of the human striatum in learning via reward and punishment during an associative learning task. Previous studies have identified the striatum as a critical component in the neural circuitry of reward-related learning. It remains unclear, however, under what task conditions, and to what extent, the striatum is modulated by punishment during an instrumental learning task. Using high-resolution functional magnetic resonance imaging (fMRI) during a reward- and punishment-based probabilistic associative learning task, we observed activity in the ventral putamen for stimuli learned via reward regardless of whether participants were correct or incorrect (i.e., outcome). In contrast, activity in the dorsal caudate was modulated by trials that received feedback—either correct reward or incorrect punishment trials. We also identified an anterior/posterior dissociation reflecting reward and punishment prediction error estimates. Additionally, differences in patterns of activity that correlated with the amount of training were identified along the anterior/posterior axis of the striatum. We suggest that unique subregions of the striatum—separated along both a dorsal/ventral and anterior/posterior axis— differentially participate in the learning of associations through reward and punishment.

[Supplemental material is available for this article.]

Flexible instrumental behavior is governed by a dynamic interplay between the tendency to increase the likelihood of making a response following a rewarding outcome and the likelihood of decreasing a response following an aversive outcome (Thorndike 1933; Konorski 1967). Evidence suggests that the striatum, along with its dopaminergic projections, is critical for learning from reward (Schultz 2007). Accumulating data also suggest that similar regions are involved in aversive learning and memory (Bromberg-Martin et al. 2010). While human neuroimaging studies have long supported the claim that activity in the striatum is correlated with reward-related learning and memory (O'Doherty 2004), the specific role for the striatum in aversive learning and memory remains undetermined. The goal of the present study was to use high-resolution functional magnetic resonance imaging (fMRI) to investigate subregional striatal learning-related activity during an associative learning paradigm that separates reward from punishment learning.

Previous human neuroimaging studies have investigated the neural substrates of reward and punishment using a variety of tasks. For example, studies have used gambling tasks (Delgado et al. 2000, 2003; Elliott et al. 2000; Breiter et al. 2001; Yacubian et al. 2006; Liu et al. 2007; Tom et al. 2007), Pavlovian conditioning (Büchel et al. 1998; Jensen et al. 2003, 2007; Seymour et al. 2004, 2005, 2007), "oddball" tasks (Tricomi et al. 2004), reversal learning (Klein et al. 2007; Wheeler and Fellows 2008; Kahnt et al. 2009; Robinson et al. 2010), and instrumental (Kim et al. 2006; Pessiglione et al. 2006) to investigate both reward- and punishment-related neural activity. While evidence

supports the notion that the striatum plays a central role in learning from reward, no general consensus has been reached on its role in aversive learning and memory.

An important limitation of many of the prior studies is that we cannot be certain that activity correlated with aversive *events* (as is typically measured) generalizes to aversive learning. For example, studies that used gambling paradigms or tasks in which the outcomes were predetermined (not contingent on the participants' responses) have consistently observed striatal activity correlated with aversive events (Delgado et al. 2000, 2003; Elliott et al. 2000; Breiter et al. 2001; Tricomi et al. 2004; Yacubian et al. 2006; Liu et al. 2007; Tom et al. 2007). However, since there are no associations or contingencies to learn in these tasks, the activity cannot be tied to learning per se. In contrast, studies of learning that utilized Pavlovian conditioning paradigms have identified ventral striatum activity during aversive learning (Jensen et al. 2003, 2007; Seymour et al. 2004, 2005, 2007). However, similar modulation of striatal activity by aversive feedback has not been observed during instrumental paradigms (Kim et al. 2006; Pessiglione et al. 2006). Human neuroimaging studies using reversal-learning paradigms during a probabilistic selection task in which both rewarded and punished stimuli are presented simultaneously on each trial in a forced-choice design are difficult to interpret within a learning framework. In these tasks, participants are able to adopt either the strategy of choosing rewarded stimuli, or of avoiding punished stimuli (or potentially both strategies) to successfully perform the task (Klein et al. 2007; Wheeler and Fellows 2008; Kahnt et al. 2009; Robinson et al. 2010). Further, activity in the striatum during reversal learning paradigms may reflect switch costs related to reversal events rather than learning-related activity specific to aversive learning. Thus, human neuroimaging studies have provided a mixed picture

[4]Corresponding author.
E-mail cestark@uci.edu.

regarding the role of the striatum in aversive learning and memory.

The purpose of the present study was to utilize an associative learning task—previously used to demonstrate robust learning-related differences for reward and punishment learning in Parkinson's Disease (PD) patients either on or off of their medication (Bódi et al. 2009)—to identify the functional neural correlates of reward and punishment learning in a healthy population. We used high-resolution blood oxygen level-dependent (BOLD) fMRI to identify learning related changes in neural activity during a probabilistic associative learning task that separated the learning of stimuli via reward from stimuli learned via punishment (Fig. 1A; also see the Materials and Methods section for a detailed description of the task). During the probabilistic associative learning task, participants were required to learn to associate specific stimuli with one of two responses (i.e., categories) through trial-and-error. A subset of the stimuli was learned via reward; if participants were correct they received a reward, and if they were incorrect they were provided no feedback. The remaining stimuli were learned via punishment; if they were correct they received no feedback, and if they were incorrect they received punishment. Together, these task manipulations enabled us to isolate activity to both reward-related and punishment-related learning during an associative learning task.

To evaluate both reward- and punishment-based learning-related activity, we performed both trial-wise and model-based analyses. In the trial-wise analyses, we tested for regions that were modulated by learning condition (reward vs. punishment), outcome (correct vs. incorrect), and amount of training (first half of the experiment vs. second half). In the model-based analysis, we used a Q-learning algorithm (Watkins 1989) to derive a trial-specific prediction error (PE) signal for both reward and punishment conditions based on participants' performance. We

correlated the trial-specific PE signal with the fMRI data and tested for regions that were modulated by reward-based PE and punishment-based PE. To increase our power, we constrained our analyses to the striatum and medial temporal lobe (MTL) based on prior evidence implicating these regions in reward and aversive learning and memory (Elliott et al. 2000; Delgado 2007; Delgado et al. 2008). Our results support the notion that the striatum plays a multifaceted role, with distinct subregions—separated along both a dorsal/ventral and anterior/posterior axis—participating in the learning of associations through reward and punishment.

Results

Behavioral performance

Optimal responding

We ran a repeated measures analysis of variance (ANOVA) with learning condition (reward and punishment) and trial blocks as within-subject factors to evaluate changes in performance during the probabilistic associative learning task. Participants showed robust learning for both reward- and punishment-based trials over the course of the experiment (main effect of block, $F_{(3,60)} = 25.83$, $P < 0.0001$). Post hoc trend analyses identified a reliable linear trend in the participants' performance across blocks (linear contrast, $F_{(1,60)} = 498$, $P < 0.0001$). We did not observe a difference in performance for the different learning conditions (main effect of learning condition, $F_{(1,20)} = 0.013$, $P = 0.91$), nor an interaction between condition and block (interaction between block and learning condition, $F_{(3,60)} = 0.331$, $P = 0.80$) (Fig. 1B). From these data, we can conclude that behavioral performance improved over the course of the experiment with no identifiable difference in performance between learning conditions (reward vs. punishment). Therefore, we believe that the feedback was sufficiently motivating and informative.

Reaction time

To analyze the reaction time (RT) data, we similarly used a repeated measures ANOVA with learning condition and blocks as within-subject factors. The results suggested that the time taken to make a decision for both reward- and punishment-based conditions decreased across blocks (main effect of block, $F_{(3,60)} = 54.78$, $P < 0.0001$). A reliable post hoc linear trend analysis supported this conclusion (linear contrast, $F_{(1,60)} = 1658$, $P < 0.0001$). Overall, punishment-based trials took more time than reward-based trials (main effect of learning condition, $F_{(1,20)} = 49.63$, $P < 0.0001$). There was no reliable interaction between condition and block for the RT data (interaction between block and learning condition, $F_{(3,60)} = 2.24$, $P = 0.09$) (Fig. 1C). These data suggest that, while participants were slower on average to respond to punishment-based trials compared to reward-based trials, they decreased their RT for both learning conditions over the course of the experiment.

Learning rate

A potential confounding factor that could account for the differences in observed fMRI data could arise from a difference in learning rates between learning conditions (i.e., stimuli learned via punishment may be slower overall, thereby requiring more trials to reach equivalent performance). To evaluate whether there was a difference in the rate of learning between reward- and punishment-based conditions, we used a state-space logistic regression algorithm (Smith et al. 2004) to calculate a learning curve for each stimulus (Law et al. 2005; Mattfeld and Stark 2010). We derived an area under the curve (AUC) measure from the
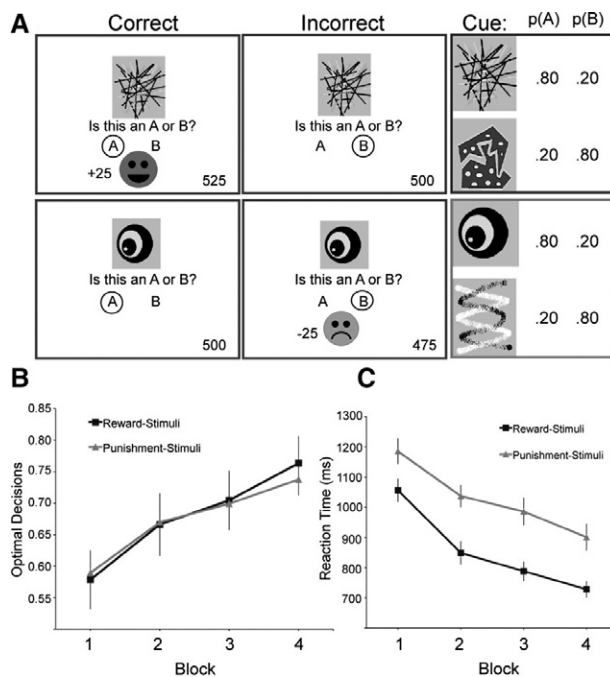

Figure 1. (A) Example of reward-based (top) and punishment-based (bottom) trials in the probabilistic associative learning task. (B) Performance was equivalent for reward-based (dark gray square) and punishment-based (light gray triangle) conditions. (C) Participants were slower overall to respond to punishment-based trials; however, the time to respond to both stimulus types decreased with training. Error bars ± SEM.

stimulus-specific learning curves for reward- and punishment-based stimuli, respectively. No reliable difference between reward- and punishment-based trials AUC was identified ($t_{(20)} = 0.82$, $P = 0.65$).

Imaging results

Reward– vs. punishment–based fMRI activity

First, we identified regions that showed differential learning-related activity for reward- vs. punishment-based learning conditions. We performed a 2×2 repeated measures ANOVA with learning condition (reward and punishment) and outcome (correct and incorrect) as the within-subject factors. The main effect of learning condition revealed activity in the bilateral ventral striatum (predominately within the ventral putamen and extending into the nucleus accumbens) (Table 1). We used a functionally defined region of interest (ROI) analysis, averaging across all voxels that survived our height and spatial extent threshold correction for multiple comparisons, to show that the main effect was driven by greater activity for reward- vs. punishment-based trials regardless of outcome or feedback (Fig. 2A). These data are consistent with previous fMRI studies demonstrating greater activity for reward vs. punishment trials in the striatum (Tricomi et al. 2004).

Outcome (correct vs. incorrect) related fMRI activity

Next, we wanted to identify learning-related activity that correlated with the outcome of an event regardless of reward or punishment condition. Several studies have suggested that the striatum is an integral brain region for processing outcome (Elliott et al. 1997; Delgado et al. 2003; Tricomi and Fiez 2008). When constraining our analysis to the striatum/medial temporal lobe mask, no regions showed a modulation in the form of a main effect of outcome (correct vs. incorrect trials).

Regions showing an interaction of stimulus type and outcome

To identify regions that showed a reliable modulation by feedback, regardless of reward or punishment condition, we selected voxels that were reliably correlated with the interaction between learning condition and outcome. It should be noted that the interaction between outcome and condition reflects feedback-related

activity due to the design of the current study—feedback trials only occurred during incorrect punishment trials and correct reward trials. Feedback trials are integral events for the learning of the appropriate associations during the experiment. Within the striatum we observed activity in the right caudate that was correlated with the interaction term (Table 1). Following a functional ROI analysis, the right caudate showed a greater level of BOLD fMRI activity for trials that received feedback compared to trials that did not receive feedback. Subsequent post hoc pairwise comparisons (Tukey's honestly significant difference [HSD] test) showed a reliable effect in the key contrast; activity in this region was greater for incorrect punishment-based trials vs. correct reward-based trials (Q = 3.8, $P < 0.05$) (Fig. 2B). These results demonstrate that the dorsal caudate is correlated with feedback regardless of reward or punishment condition. Post hoc pairwise comparisons suggest that there is a modest yet reliable difference between incorrect punishment-based trials when compared to correct reward-based trials.

Model–based fMRI activity

To determine whether distinct subregions of the striatum were modulated by reward- and punishment-based PE estimates during an instrumental learning task, we parametrically regressed the fMRI activity with PE estimates derived from a Q-Learning algorithm (Watkins 1989). Given the trial structure of our paradigm and our inability to resolve within-trial events, a temporal difference algorithm would provide the same estimation of the trial-by-trial PE as a delta rule; therefore, we used the simpler delta rule here (Sutton and Barto 1998). We used a repeated measures ANOVA to identify voxels where BOLD fMRI activity was modulated by either a reward-based event occurring, reward-based PE, a punishment-based event occurring, or punishment-based PE. Within the striatal/medial temporal lobe mask, this analysis revealed robust modulation by both reward- and punishment-based PE estimates throughout the striatum (Table 1). Bonferroni corrected post-hoc tests revealed that the anterior head of the right dorsal caudate as well as ventral regions of the bilateral putamen correlated with reward-based PE estimates (all $t_{(27)} > 4.8$, all $P < 0.0001$) and not punishment-based PE estimates (all $t_{(27)} < 1.8$, all $P > 0.07$) (Fig. 3A–C). In contrast, functional ROIs slightly more posterior near the junction between the head and body of the bilateral dorsal caudate largely correlated with punishment-based PE estimates (all $t_{(27)} > 3.2$, all $P < 0.0032$) (Fig. 3D,E). However, the functional ROI in the posterior portion of the left caudate was correlated with reward-based PE estimates as well ($t_{(27)} > 3.9$, $P = 0.0005$) (Fig. 3D). The observed anatomical dissociation between reward and punishment PE neural activity here extends to an instrumental task previous findings identified during a first-order Pavlovian conditioning paradigm (Seymour et al. 2007).

fMRI activity reflecting amount of training

Previous studies using probabilistic category learning paradigms (e.g., Weather Prediction Task) have observed differences in learning between PD patients, MTL amnesics, and controls (Knowlton et al. 1996). However, following extended training, PD patients' performance improved remarkably and became

Table 1. Regions of activation identified in the trial- and model-based analyses

Analysis	Region of activation	MNI (x, y, z)	Volume
Learning condition: reward vs. punishment	Nucleus accumbens/ventral putamen (L)	−15, 6, −6	135 mm³
	Nucleus accumbens/ventral putamen (R)	18, 6, −6	223 mm³
Outcome: correct vs. incorrect	No regions survived corrections for multiple comparisons		
Interaction between learning condition and outcome	Dorsal caudate (R)	8, 0, 10	240 mm³
Model-based analyses	Dorsal caudate head/body (L)—*punishment and reward PE*	−8, 7, 12	584 mm³
	Dorsal caudate head/body (R)—*punishment PE*	11, 7, 13	523 mm³
	Dorsal caudate head (R)—*reward PE*	12, 17, 3	189 mm³
	Ventral putamen (R)—*reward PE*	18, 8, 1	182 mm³
	Ventral putamen (L)—*reward PE*	−17, 8, −5	121 mm³
Training related changes	Dorsal caudate head (R)—*first half < second half*	16, 13, 12	196 mm³
	Dorsal caudate body (R)—*first half > second half*	15, −15, 26	159 mm³
	Hippocampus (R)—*first half < second half*	30, −15, −22	152 mm³

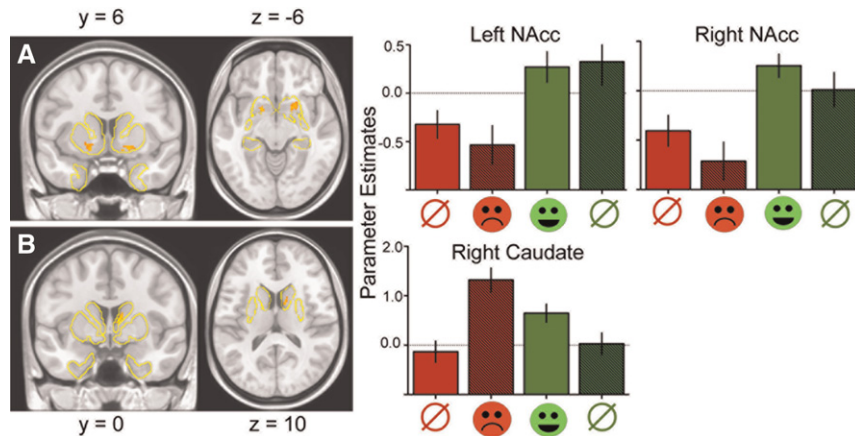(R) Right; (L) left; (PE) prediction error.

Figure 2. (*A*) Activity bilaterally in the ventral striatum was modulated by the main effect of learning condition, showing greater activity for reward- compared to punishment-based events. (*B*) A region in the right dorsal caudate was modulated by the interaction between learning condition and outcome. This region was correlated with feedback showing greater activity for punishment-based feedback (red) compared to reward-based feedback (green). The yellow outline represents the anatomically delineated region of interest used for small volume corrections. (Red Ø) punishment trial correctly responded to—no feedback; (red sad face) punishment trial incorrectly responded to—feedback; (green happy face) reward trial correctly responded to—feedback; (green Ø) reward trial incorrectly responded to—no feedback. Error bars ± SEM.

the control analyses, we included PE and RT regressors in our GLM testing for training-related changes in activity to possibly account for any training-related variance. In both analyses the region-by-training interaction remained robust (all $F_{(2,40)} > 25$, all $P < 0.0001$), suggesting that the observed results are not confounded by differential responding to PE estimates or changes in activity that are tracking changes in RT.

Discussion

BOLD fMRI activity across distinct subregions of the striatum was dynamically modulated during a probabilistic reward- and punishment-based associative learning task. Specifically, the striatum appeared to be functionally divisible along its dorsal/ventral axis with the ventral striatum (regions of the ventral putamen and nucleus accumbens) modulated by learning condition (reward vs. punishment) and reward PE, while the dorsal caudate appeared to be responsive to feedback during both correct reward and incorrect punishment trials. Further, the striatum demonstrated functional specialization along its anterior/posterior axis, with anterior regions modulated by reward PE estimates and late phases of learning, while more posterior regions correlated with punishment PE (at the junction between the head and body of the caudate) and early phases of learning (in the body of the caudate). We suggest that the observed functional specialization along the dorsal/ventral axis likely reflects differentiation within the striatum between regions encoding the valence of events (ventral striatum) vs. regions that play an important role in the learning of associations through trial-and-error by monitoring motivationally salient events (feedback). Moreover, the regional differences in the response as a function of the amount of training suggest the possibility of key interactions between declarative vs. nondeclarative forms of learning and memory.

Activity in the ventral striatum was robustly modulated by learning condition. Bilateral functionally defined ROIs in the ventral putamen showed greater activity for reward compared to punishment-based conditions. These results are consistent with prior human imaging studies showing greater activity for rewarding vs. aversive events (Delgado et al. 2000, 2003; Tricomi et al. 2004). Our results extend these prior findings to an associative learning paradigm. Prior studies that identified differential responding to reward and punishment in the striatum observed this difference primarily during the outcome phase of each trial. Due to constraints in task design, we cannot rule out that outcome-related activity is likely playing an important role here. However, outcome related processing alone does not account for the observed results. For example, reward events that were incorrect and received no feedback showed a greater amount of activity when compared to incorrect punishment events that received feedback. The fact that reward events showed reliably greater activity independent of outcome or feedback supports the notion that the ventral striatum is a key region representing the valence—or the degree of attraction or aversion—of events during learning.

In the dorsal striatum, we observed activity that was modulated by feedback compared to trials that received no feedback. There was slightly greater activity for punishment stimuli that
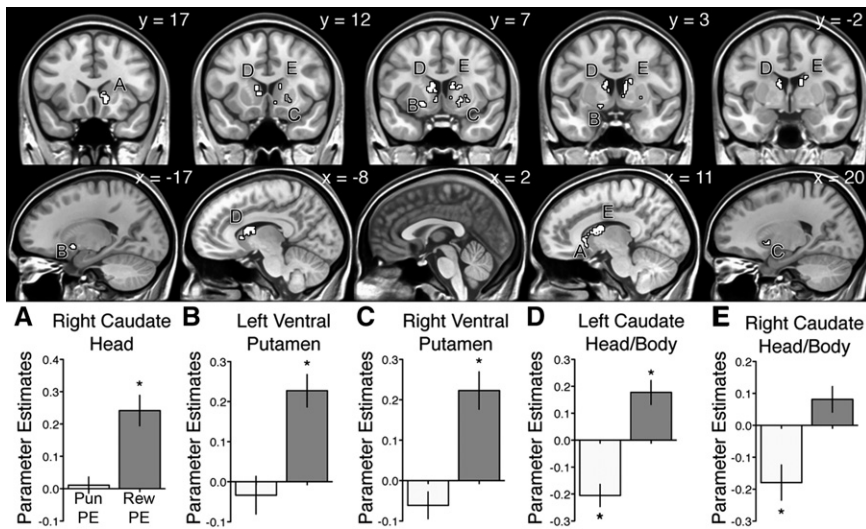
indistinguishable from controls. Here, we wanted to evaluate whether the striatum, during an associative learning paradigm with stochastic feedback, showed differential learning-related activity over the course of training.

To test for an effect of training, we divided our correct and incorrect regressors for reward- and punishment-based conditions into those trials that occurred during the first half of the training (first two runs) and the trials that comprised the second half of training (last two runs). We then selected voxels that showed a main effect of training. Three regions within our striatal/medial temporal lobe mask showed reliable modulation by training following corrections for multiple comparisons: right posterior body of the caudate, right anterior head of the caudate, and right hippocampus (Table 1). Using the regions identified by the main effect of training as functionally defined ROIs as well as their parameter estimates for each condition of interest, we performed a repeated measures ANOVA with learning condition (reward vs. punishment), outcome (correct vs. incorrect), amount of training (first half vs. second half), and region (hippocampus, anterior caudate, and posterior caudate) as within-subjects factors. This analysis showed a reliable region-by-training interaction ($F_{(2,40)} = 25.96$, $P < 0.0001$). The right posterior caudate ROI showed greater activity during the first half of training compared to the second half of training (Fig. 4A), while the right anterior caudate and hippocampus showed remarkably similar results—greater activity for trials occurring during the second half of training vs. the first half (Fig. 4B). Importantly, the amount of training × outcome ($F_{(1,20)} = 1.79$, $P = 0.19$), region × outcome ($F_{(2,40)} = 0.3169$, $P = 0.73$), and amount of training × region × outcome ($F_{(2,40)} = 1.47$, $P = 0.24$) interactions did not reach significance, suggesting that activity related to outcome (correct and incorrect trials) did not change reliably over the course of training in these particular regions. Therefore, activity reflecting the region-by-amount of training interaction is not simply the product of a change in performance with experience.

Additionally, to ensure that the observed results were, in fact, due to training-related changes and not due to changes in response to PE or RT—which are also dynamically changing throughout the experiment—we ran two control analyses. In

Figure 3. The model-based analysis identified regions in the right anterior dorsal caudate (*A*) and bilateral regions of the ventral putamen (*B,C*) that were reliably modulated by reward-based PE estimates. Conversely, bilateral regions of the posterior dorsal caudate near the junction between the head and body of the caudate were reliably modulated by punishment-based PE estimates (*D,E*). *$P < 0.01$ (Bonferroni correction). (Rew) reward; (Pun) punishment; (PE) prediction error; (dark gray bars) reward PE activity; (light gray bars) punishment PE activity. Error bars ± SEM.

were incorrect when compared to reward stimuli that were responded to correctly. The pattern and location of activity correspond very nicely to an ROI identified by Jensen et al. (2007) showing overall greater activity for aversive and appetitive when compared to neutral conditioned stimuli. Interestingly, in their ROI they also observed slightly greater activity to aversive compared to appetitive stimuli. The observed neural modulation in the dorsal caudate makes this region an ideal candidate for learning to associate actions with motivationally salient events (e.g., feedback). In turn, these neural signals are likely used to strengthen or weaken the appropriate associations between actions and their respective outcomes (Barto 1995).

While the dorsal striatum has consistently been identified as a region integral to reward learning (O'Doherty 2004), its observation in aversive learning tasks has been less consistent (Büchel et al. 1998; Becerra et al. 2001; Breiter et al. 2001; Seymour et al. 2005, 2007; Yacubian et al. 2006). The discrepancy between our study showing reliable modulation in the dorsal striatum by punishment and previous studies showing no activity for aversive events may be a result of the differences in task design. Many of the previous tasks used Pavlovian conditioning or gambling paradigms that lack a clear contingency between a participant's action and the outcome. Tricomi et al. (2004) showed that dorsal caudate activity is reliably modulated by the contingency between a participant's response and whether or not they won or lost money. Our paradigm relies on this contingency to facilitate learning to select the optimal reward-based stimuli and avoid the nonoptimal punishment-based stimuli.

Our model-based results are consistent with previous fMRI studies investigating appetitive and aversive PE signals (Seymour et al. 2007). Seymour and colleagues used a Pavlovian conditioning paradigm to associate conditioned stimuli with reward, loss, or both outcomes. Extending their findings to an instrumental learning task, we identified dissociable regions within the striatum that represented reward and punishment PE along not only the anterior/posterior axis of the striatum but also the dorsal/ventral axis. Specifically, reward-based PE, and not punishment-based PE, modulated the anterior head of the dorsal caudate as well as

the bilateral ventral putamen. In contrast, bilateral regions of the dorsal caudate, slightly more posterior near the junction between the head and body of the caudate, showed modulation to punishment-based PE estimates, with the left ROI showing a correlation to both. A notable difference between our results and prior studies is that many of the functional ROIs we identified were largely located in the dorsal caudate. We believe, again, that this is likely the result of task differences. Seymour and colleagues specifically designed their task to omit the opportunity for choice (Seymour et al. 2007). Given the importance of action-outcome contingencies in activating the dorsal caudate, this discrepancy is not surprising.

The anatomically distinct representations of reward and aversive PE along both the anterior/posterior and dorsal/ventral axes of the striatum may be the result of different sources of cortical input related to the processing of reward- and punishment-based events. Investigations of the neural correlates of the BOLD signal suggest that it more likely reflects afferent input as well as intrinsic processing within a brain region (Logothetis et al. 2001). Moreover, the striatum receives largely segregated cortical input (Alexander et al. 1986). Human neuroimaging studies have identified distinct regions across the orbitofrontal cortex (OFC) that are correlated with the processing of both reward- and punishment-related information, with the lateral OFC correlated with punishing outcomes and the medial OFC with rewarding outcomes (O'Doherty et al. 2001). The anterior insula has also been consistently implicated in the representation of punishment and aversive prediction errors (Greenspan and Winfield 1992; Casey et al. 1994; Kim et al. 2006; Pessiglione et al. 2006). We note that the OFC and anterior insula predominately project to the ventral striatum (Chikama et al. 1997; Haber 2003), which potentially accounts for the subtle anatomical separation observed in the Seymour et al. (2007) study. In contrast, our results were predominately located in the dorsal striatum, which receives projections from more lateral regions in the OFC, prefrontal cortex, and posterior regions of the insula (Chikama et al. 1997; Haber 2003). Evidence from prior neuroimaging and tract-tracing studies do not strongly support the conclusion that different cortical contributions are leading to the differences observed in our study. However, given the variable cortical substrate likely involved in reward and punishment learning, we cannot rule out the possibility that the observed anterior/posterior separation in reward and punishment prediction error estimates within the striatum is the result of distinct cortical contributions.

An alternative hypothesis for the anatomical separation may reflect modulation of neural activity by different neuromodulators (Daw et al. 2002; Doya 2002; Bromberg-Martin et al. 2010). Both serotonin (Daw et al. 2002) and distinct populations of dopaminergic neurons (Matsumoto and Hikosaka 2009; for review, see Bromberg-Martin et al. 2010) have been posited to play a role in aversive learning and are differentially expressed throughout the striatum (Heidbreder et al. 1999; Matsumoto and Hikosaka 2009). Specifically, two types of dopamine neurons have been identified: (1) those that are excited by reward and reward-predicting stimuli and inhibited by aversive stimuli; and (2) those
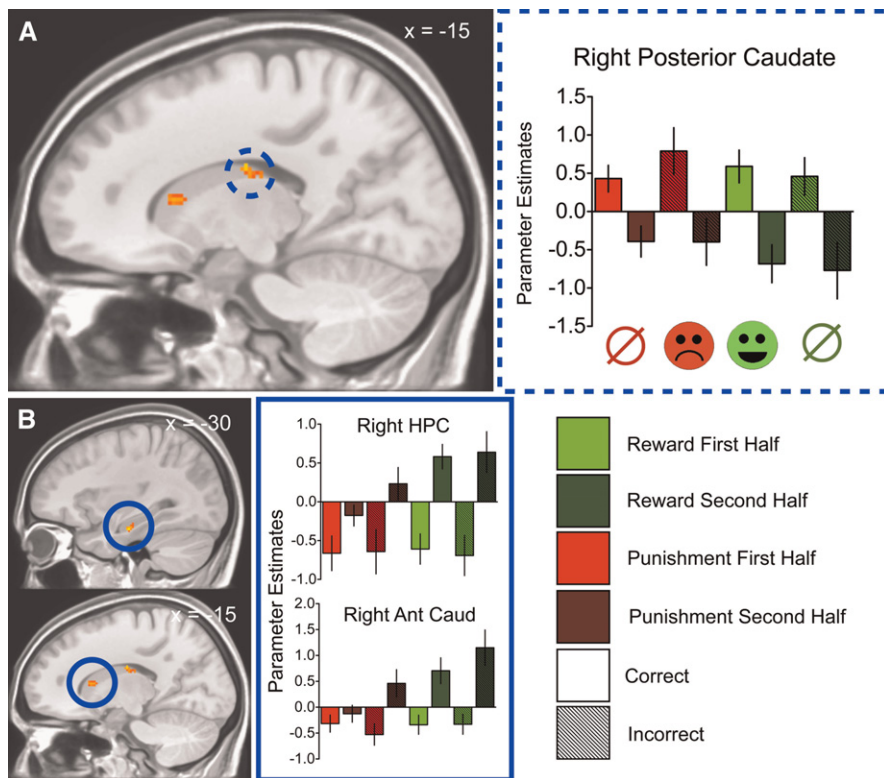
Figure 4. (*A*) An ROI in the right posterior caudate showed a reliable modulation by the amount of training, with greater activity for trials during the first half of training compared to the second half. (*B*) ROIs in the right anterior caudate and right hippocampus were also modulated by the amount of training but showed the opposite pattern of activity. These regions showed greater activity during the second half of training compared to the first half. (Ant) Anterior; (HPC) hippocampus; (Caud) caudate; (red Ø) punishment trial correctly responded to—no feedback; (red sad face) punishment trial incorrectly responded to—feedback; (green happy face) reward trial correctly responded to— feedback; (green Ø) reward trial incorrectly responded to—no feedback. Error bars ± SEM.

has shown a similar dissociation along the anterior/posterior axis of the striatum in nonhuman primates. Seger (2008), using fMRI in humans, has investigated the functional roles of the distinct regions of the striatum during categorization learning. Future studies of associative learning should utilize functional connectivity and pharmacological manipulations along the anterior/posterior axis of the striatum in an effort to improve our understanding of the functionally distinct subregions of the striatum.

Lastly, the reliable region-by-period of training interaction demonstrated a dissociation between regions that were active during learning early in training compared to regions that showed greater activity with extended training. The right posterior caudate was robustly modulated early in training compared to later, while the right anterior caudate and hippocampus showed the opposite pattern of activity, being more active primarily late in training compared with early in training. The training-related results reported here differ from previous studies in both humans (Seger et al. 2010) and monkeys (Hikosaka et al. 1998) that have shown early activation in the anterior striatum, with gradual recruitment of more posterior regions with training. We believe that the discrepancies between our results and prior studies likely stem from differences in task design. For example, Seger et al. (2010) used a deterministic categorization task, while here we employed a probabilistic categorization task. We believe that the probabilistic nature of the feedback in our task makes participants more reliant at the beginning of training on nonexplicit mnemonic strategies, which may involve more posterior regions of the striatum. However, with sufficient training, participants may begin to employ more explicit strategies as they become aware of the stochastic contingencies of the task. For example, goal-directed learning has been shown to be more dependent on the dorsomedial striatum in rats (Yin et al. 2005), which is homologous to the anterior striatum in primates. The pattern of activity observed for the training-related changes in our study is consistent with this hypothesis; however, more direct tests are needed to confirm or reject this possibility.

In summary, these data suggest that the striatum is playing a multifaceted role in the learning of associations through reward and punishment. During the probabilistic associative learning task, both a dorsal/ventral and anterior/posterior functional differentiation emerged from our analyses. The distinction between the dorsal and ventral striatum may reflect unique contributions to the learning of arbitrary associations, with the ventral striatum tracking valence while the dorsal striatum tracks motivationally salient events such as feedback. The anterior/posterior divide for both the model-based and amount of training analyses, on the other hand, may reflect the processing of unique informational content, neuromodulator function, the division of labor between goal-directed and habitual mnemonic function, or, more likely, some combination of all three.

that are excited by both reward and aversive stimuli and their predictors (Matsumoto and Hikosaka 2009). The former are located in the dorsolateral substantia nigra pars compacta which projects mainly to the dorsal striatum in monkeys and rats (Lynd-Balta and Haber 1994; Ikemoto 2007). The latter are located in the ventromedial substantia nigra pars compacta and ventral tegmental area which predominately send projections to the ventral striatum. Matsumoto and Hikosaka (2009) found that the dorsolateral dopaminergic population responded preferentially to motivationally salient stimuli, while the dorsomedial population responded to value-related processes during a Pavlovian conditioning paradigm. The functional dissociation between dopaminergic populations corresponds well with the results reported here obtained by our trial-based analyses, i.e., dorsal caudate was modulated by feedback (motivationally salient events) while the ventral putamen correlated with valence. Further, a heterogeneous dopaminergic population could contribute to our observed anatomical separation in the reward and punishment prediction error modulations.

While our data cannot differentiate between these two alternatives, the results highlight the need for more precise functional delineation of the striatum (see Haber 2003 and Haruno and Kawato 2006 for models positing dorsal/ventral with anterior/posterior integration). To this end, seminal work by Yin et al. (2004, 2005) in rats has demonstrated a functional subdivision between the dorsomedial and dorsolateral striatum during instrumental learning tasks. Work by Miyachi et al. (1997, 2002)

Materials and Methods

Participants

Twenty-eight right-handed volunteers (12 male, mean age 21.7, age range 19–31), with normal or corrected to normal vision and no history of neurological disease or psychiatric illness, participated in the experiment. All gave written informed consent. Participants were recruited from the University of California, Irvine, community and were paid for their participation. Seven participants were excluded from further analysis because they did not have a sufficient number of trials to reliably estimate the hemodynamic response for several conditions of interest, leaving 21 participants for the final analyses.

Experimental task

The behavioral task was previously published in Bódi et al. (2009). Before scanning, participants were informed that they would be presented with four images. Participants were instructed to select which category (A or B) they believed each image belonged to via an MR-compatible button box. They were informed that, following their choices, they would win points, lose points, or receive no feedback. Participants were not told which images were paired with which outcome and were required to learn the associations through trial-and-error. Participants were instructed to win as many points as possible; each participant began the experiment with 500 points.

A trial began with the presentation of one of the four images, with the question "Is this an 'A' or a 'B'?" below the image, followed by the two possible categories. Stimuli were presented for 3000 msec. Each image marked the beginning of either a reward- or a punishment-based trial. Responses were collected during the presentation of the stimulus. Following the participant's response, the selected category was outlined by a circle and was followed by feedback: +25 points and a green happy face for reward-based trials, −25 points and a red sad face for punishment-based trials, or nothing (Fig. 1A). A 1000-msec intertrial-interval separated the presentation of the next stimulus, marking the beginning of the next trial. The total trial length was 4000 msec.

During reward-based trials, when participants selected the optimal category, they received reward feedback with an 80% probability. They received no feedback for the remaining 20% of trials. If the participants selected the nonoptimal category, they received reward feedback 20% of the time, and for the remaining 80% of trials they received no feedback. Similarly, on the punishment-based trials, when participants selected the optimal category, they received no feedback 80% of the time, while on the remaining 20% of trials the same selection was associated with the receipt of punishment. If they selected the nonoptimal category, they received no feedback on 20% of the trials and punishment on the remaining 80% of trials.

In addition to learning trials, perceptual baseline trials were randomly presented during the experiment. During the baseline trials a random visual static pattern was presented along with two boxes that differed slightly in their opacity (target continuously varied between 11% and 17% greater opacity than the nontarget). Participants were instructed to select the brighter of the two boxes. Baseline trials were introduced to induce jitter between trial types, to aid in the estimation of the hemodynamic response, and to establish a reference for the signal for the trial types of interest (Dale and Buckner 1997; Burock et al. 1998).

Participants completed either eight ($n = 21$) or four ($n = 7$, typically resulting from scheduling or scanner difficulties) scanning runs. Participants that completed eight runs were presented with a second stimulus set after the first four runs to increase the number of trials during which participants were actively learning new associations. The order of stimulus sets was counterbalanced across participants. A run consisted of 40 learning trials (each learning image was presented 10 times per run) and 20 baseline trials for a total of 60 trials per run. Trial order was randomly determined. Each run lasted 4 min.

Calculation of the stimulus learning curves

We employed a logistic regression algorithm (Smith et al. 2004) to calculate a learning curve for each stimulus, which was used to derive an area under the curve (AUC) estimate to evaluate learning rate differences for both learning conditions (reward vs. punishment). The binary performance for each stimulus (1, if the response was optimal and 0, if the response was not optimal) was used to calculate the probability of a correct response based on the number of trials using a Gaussian random walk model as the state equation and a Bernoulli model as the observation equation. Representative learning curves can be seen in Supplementary Figure 1.

MRI acquisition

Imaging data were collected on a Philips 3.0 T scanner (Best) equipped with an 8-channel SENSE (Sensitivity Encoding) head coil at the Research Imaging Center (RIC) (Irvine, CA). Functional images were acquired using a T2*-weighted, echoplanar single-shot pulse sequence (|TR, 2000 msec; TE, 26 msec; flip angle, 70°; matrix size, 128 mm × 128 mm; FOV, 180 mm × 180 mm; SENSE factor, 2.5; slice thickness, 1.8 mm; interslice gap, 0.2 mm) with an in-plane acquisition resolution of 1.5 × 1.5 mm. Thirty-five coronal slices were acquired along the long axis of the hippocampus covering the majority of the MTL and the basal ganglia, encompassing approximately half of the cortex. One hundred twenty volumes were acquired during each run. To allow for T1 equilibration, the first four volumes of each run were discarded. Whole-brain anatomical images were acquired using a sagittal T1-weighted magnetization-prepared rapid gradient echo (MP-RAGE) scan (TR, 11 msec; TE, 4.6 msec; flip angle, 18°; matrix size, 320 mm × 264 mm; FOV, 240 mm × 150 mm; slice thickness, 0.75 mm; resolution 0.75 mm isotropic; 200 slices, coregistered with the fMRI data).

Image processing and cross-participant registration

Data analyses were performed using Analysis of Functional Neuroimages (AFNI) software (Cox 1996). Functional volumes from each scanning run were corrected for differences in slice acquisition and realigned to the first volume. Functional data were then coregistered in time and three dimensions. Any time points with >3° rotation or 2 mm translation were eliminated from further analysis. All data for each participant were concatenated across runs. Functional data were iteratively spatially smoothed using AFNI's 3dBlurToFWHM which estimates the intrinsic smoothness of the residuals and then spatially smoothes the functional data until obtaining the targeted smoothness—in our case, an isotropic 3-mm FWHM Gaussian kernel. We functionally defined ROIs by setting a voxel-wise threshold of $P < 0.02$ with a connectivity radius of 1.6 mm and a spatial extent threshold volume of 118 mm^3, resulting in an overall family-wise error corrected alpha-probability of $P < 0.05$ as determined by Monte Carlo simulations (AFNI's AlphaSim program).

Cross-participant alignment began with the whole-brain spatial normalization of each participant's T1-weighted MP-RAGE to the Talairach atlas (Talairach and Tournoux 1988) using a 12-parameter affine transformation matrix. This initial registration provides a rough first pass, removing large spatial shifts between participants prior to our subsequent region of interest alignment (ROI-AL) approach (Stark and Okado 2003). Fine-tuned cross-participant registration was accomplished using Advanced Normalization Tools (ANTs), a powerful diffeomorphic alignment algorithm (Avants et al. 2008) that creates a 3D vector field to map each participant's brain space into a template space. We used each participant's structural scan to create a custom central tendency template. Following the template generation, each participant's MP-RAGE was warped into the custom template space. The resulting transformation parameters were subsequently applied to the functional data.

Statistical analysis of functional imaging data

Subject-specific behavioral design matrices were created containing the following regressors: reward trials that were optimally

responded to and received feedback (RO+; *correct*); reward trials that were optimally responded to and did not receive feedback (RO−; *were correct but told incorrect*); reward trials that were not optimally responded to and did not receive feedback (RNO−; *incorrect*); reward trials that were not optimally responded to and received feedback (RNO+; *were incorrect but told correct*); punishment trials that were optimally responded to and did not receive feedback (PO−; *correct*); punishment trials that were optimally responded to and received feedback (PO+; *were correct but told incorrect*); punishment trials that were not optimally responded to and did not receive feedback (PNO−; *were incorrect but told correct*); and punishment trials that were not optimally responded to and received feedback (PNO+; *incorrect*). Nuisance regressors coding for drift in the MR signal were also included in the design matrix.

A deconvolution approach based on multiple linear regression was used to analyze the fMRI data for each participant. The hemodynamic response for each vector is estimated using nine time-shifted design matrices modeled as tent functions, estimating the BOLD activity from 0 to 16 sec after the onset of the trial. The resulting β coefficients represent activity vs. baseline for each vector at a given time point in each voxel. The summed β coefficients (4–12 sec after trial onset) were used as the model's estimate for each regressor of interest. The resulting summed β coefficients for trial types RO+, RNO−, PO−, and PNO+ were used for group-level analyses—repeated-measures ANOVA with factors *learning condition* (reward-based, punishment-based) and *outcome* (correct, incorrect), testing for significant effects across the group.

In a separate yet related analysis, we assessed the BOLD activity correlated with the amount of training. To perform this analysis, we ran a separate multiple regression model using regressors similar to those of our original analysis; however, we further split the conditions of interest according to whether they were experienced during the first two runs of the experiment (first half) or the last two runs of the experiment (second half) for each stimulus set. For those participants who received eight experimental runs, the second four runs were split in a manner similar to the original four runs—doubling the number of events that constituted our conditions of interest. At the group-level, we utilized a linear mixed-effects model to identify reliable differences across the group for *learning condition* (reward, punishment), *outcome* (correct, incorrect), and *amount of training* (first half, second half). Separate control analyses were performed to account for the possibility that the resulting BOLD modulation was, in part, due to differential responding to PE estimates or correlations with RT. In the control analyses, regressors for demeaned PE estimates and RT measurements were added to the GLM to account for these potential confounding factors.

Last, to assess BOLD activity related PE estimates, we modeled each participant's performance using a Q-learning algorithm and derived trial-specific PE estimates using a delta learning rule (Widrow and Hoff 1960). We used demeaned PE estimates as auxiliary behavioral information to parametrically assess changes in BOLD activity correlated with PE estimates. The trial-by-trial PE estimates were convolved with a canonical basis function. The resulting time-series was correlated with each participant's functional data using multiple linear regression. The subject-specific design matrices included regressors for reward-based events occurring, punishment-based events occurring, PE estimates for reward-based trials, PE estimates for punishment-based trials, along with regressors of no interest coding for drift in the MR signal. The β coefficients for the reward- and punishment-based PE estimates were used for subsequent repeated measures ANOVA at the group level.

Computational learning model

We modeled each participant's performance using a Q-learning algorithm (Watkins 1989). Action values were updated using the delta rule (Widrow and Hoff 1960): $A_i(t + 1) = A_i(t) + \eta[\delta(t)]$; t is the trial number, i is the selected category (A or B), η is the learning rate, and δ is the error estimate reflecting the difference between expected and received outcomes, $E—A_i$. The expected value, E,

was 1 for reward-based trials that received feedback, −1 for loss-based trials that received feedback, and 0 for any trials that did not receive feedback. We used the trial-specific error estimate, δ, for subsequent parametric analysis of the fMRI data. The temporal structure of each trial only facilitated the estimation of the trial-by-trial PE estimate; therefore, we used the delta rule rather than the temporal difference algorithm here (Sutton and Barto 1998).

In the model, the associative value for each category began at 0.50. We used a softmax action selection algorithm (Sutton and Barto 1998) to determine the probability that each category would be selected based on the trial-specific associative values: $P_i(t) = \exp[\beta A_i(t)]/\sum \exp[\beta A_i(t)]$, where β is the inverse temperature amplifying the differences between action values for high values of β or making the actions equally likely for low values of β.

The best fitting learning rate, η, and inverse temperature, β, were determined using maximum likelihood estimations. Across participants, we systematically incremented η (0.01 to 1) and β (0.5 to 5), calculating for each pair of estimated parameters the negative log-likelihood of the probability that the current response would be made by the participant: $L = \sum -\ln P_i(t)$. For each participant, we estimated the best fitting parameters by selecting the set of model parameters that gave the lowest value of L. A single set of model parameters (the average of the parameters across all participants) were used to derive the final prediction error estimates: η = 0.41; β = 3.73 (Supplemental Table 1).

Acknowledgments

References

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

Avants B, Duda JT, Kim J, Zhang H, Pluta J, Gee JC, Whyte J. 2008. Multivariate analysis of structural and diffusion imaging in traumatic brain injury. *Acad Radiol* **15:** 1360–1375.

Barto AG. 1995. Adaptive critics and the basal ganglia. In *Models of information processing in the basal ganglia* (ed. JC Houk, et al.), pp. 215–232. The MIT Press, Cambridge, MA.

Becerra L, Breiter HC, Wise R, Gonzalez RG, Borsook D. 2001. Reward circuitry activation by noxious thermal stimuli. *Neuron* **32:** 927–946.

Bódi N, Kéri S, Nagy H, Moustafa A, Meyers CE, Daw N, Dibó G, Takáts A, Bereczki D, Gluck M. 2009. Reward-learning and the novelty-seeking personality: A between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. *Brain* **132:** 2385–2395.

Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P. 2001. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* **30:** 619–639.

Bromberg-Martin ES, Matsumoto M, Hikosaka O. 2010. Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron* **68:** 815–834.

Büchel C, Morris J, Dolan RJ, Friston KJ. 1998. Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron* **20:** 947–957.

Burock MA, Buckner RL, Woldorff MG, Rosen BR, Dale A. 1998. Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *Neuroreport* **9:** 3735–3739.

Casey KL, Minoshima S, Berger KL, Koeppe RA, Morrow TJ, Frey KA. 1994. Positron emission tomographic analysis of cerebral structures activated specifically by repetitive noxious heat stimuli. *J Neurophysiol* **71:** 802–807.

Chikama M, McFarland NR, Amaral DG, Haber SN. 1997. Insular cortical projections to functional regions of the striatum correlate with cortical cytoarchitectonic organization in the primate. *J Neurosci* **17:** 9686–9705.

Cox RW. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* **29:** 162–173.

Dale AM, Buckner RL. 1997. Selective averaging of rapidly presented individual trials using fMRI. *Hum Brain Mapp* **5:** 329–340.

Daw N, Kakade S, Dayan P. 2002. Opponent interactions between serotonin and dopamine. *Neural Netw* **15:** 603–616.

Delgado MR. 2007. Reward-related responses in the human striatum. *Ann NY Acad Sci* **1104:** 70–88.

Delgado MR, Nystrom LE, Fissell C, Nolls C, Fiez JA. 2000. Tracking the hemodynamic responses to reward and punishment in the striatum. *J Neurophysiol* **84:** 3072–3077.

Delgado MR, Locke HM, Stenger VA, Fiez JA. 2003. Dorsal striatum responses to reward and punishment: Effects of valence and magnitude manipulations. *Cogn Affect Behav Neurosci* **3:** 27–38.

Delgado MR, Li J, Schiller D, Phelps EA. 2008. The role of the striatum in aversive learning and aversive prediction errors. *Phil Trans R Soc B* **363:** 3787–3800.

Doya K. 2002. Metalearning and neuromodulation. *Neural Netw* **15:** 495–506.

Elliott R, Frith CD, Dolan RJ. 1997. Differential neural response to positive and negative feedback in planning and guessing tasks. *Neuropsychologia* **35:** 1395–1404.

Elliott R, Friston KJ, Dolan RJ. 2000. Dissociable neural responses in human reward systems. *J Neurosci* **20:** 6159–6165.

Greenspan JD, Winfield JA. 1992. Reversible pain and tactile deficits associated with a cerebral tumor compressing the posterior insula and parietal operculum. *Pain* **50:** 29–39.

Haber SN. 2003. The primate basal ganglia: Parallel and integrative networks. *J Chem Neuroanat* **26:** 317–330.

Haruno M, Kawato M. 2006. Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Netw* **19:** 1241–1254.

Heidbreder CA, Hedou G, Feldon J. 1999. Behavioral neurochemistry reveals a new functional dichotomy in the shell subregion of the nucleus accumbens. *Prog Neuropsychopharmacol Biol Psychiatry* **23:** 99–132.

Hikosaka O, Miyashita K, Miyachi S, Sakai K, Lu X. 1998. Differential roles of the frontal cortex, basal ganglia, and cerebellum in visuomotor sequence learning. *Neurobiol Learn Mem* **70:** 137–149.

Ikemoto S. 2007. Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Res Rev* **56:** 27–78.

Jensen J, McInotsh AR, Crawley AP, Mikulis DJ, Remington G, Kapur S. 2003. Direct activation of the ventral striatum in anticipation of aversive stimuli. *Neuron* **40:** 1251–1257.

Jensen J, Smith AJ, Willeit M, Crawley AP, Mikulis DJ, Vitcu I, Kapur S. 2007. Separate brain regions code for salience vs. valence during reward prediction in humans. *Hum Brain Mapp* **28:** 294–302.

Kahnt T, Park SQ, Cohen MX, Beck A, Heinz A, Wrase J. 2009. Dorsal striatal-midbrain connectivity in humans predicts how reinforcements are used to guide decisions. *J Cogn Neurosci* **21:** 1332–1345.

Kim H, Shimojo S, O'Doherty JP. 2006. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biology* **4:** e233. doi: 10.1371/journal.pbio.0040233.

Klein TA, Neumann J, Reuter M, Hennig J, von Cramon DY, Ullsperger J. 2007. Genetically determined differences in learning from errors. *Science* **318:** 1642–1645.

Knowlton BJ, Mangels JA, Squire LR. 1996. A neostriatal habit learning system in humans. *Science* **273:** 1399–1402.

Konorski J. 1967. *Integrative activity of the brain: An interdisciplinary approach*. University of Chicago Press, Chicago.

Law JR, Flannery MA, Wirth S, Yanike M, Smith AC, Frank LM, Suzuki WA, Brown EN, Stark CEL. 2005. FMRI activity during the gradual acquisition and expression of paired-associate memory. *J Neurosci* **25:** 5720–5729.

Liu X, Powell DK, Wang H, Gold BT, Corbly CR, Joseph JE. 2007. Functional dissociation in the frontal and striatal areas for processing of positive and negative reward information. *J Neurosci* **27:** 4587–4597.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A. 2001. Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412:** 150–157.

Lynd-Balta E, Haber SN. 1994. The organization of midbrain projections to the striatum in the primate: Sensorimotor-related striatum versus ventral striatum. *Neuroscience* **59:** 625–640.

Matsumoto M, Hikosaka O. 2009. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* **459:** 837–841.

Mattfeld AT, Stark CEL. 2010. Striatal and medial temporal lobe functional interactions during visuomotor associative learning. *Cereb Cortex* **21:** 647–658.

Miyachi S, Hikosaka O, Miyashita K, Karadi Z, Rand MK. 1997. Differential roles of monkey striatum in learning of sequential hand movement. *Exp Brain Res* **115:** 1–5.

Miyachi S, Hikosaka O, Lu X. 2002. Differential activation of monkey striatal neurons in the early and late stages of procedural learning. *Exp Brain Res* **146:** 122–126.

O'Doherty JP. 2004. Reward representations and reward-related learning in the human brain: Insights from neuroimaging. *Curr Opin Neurobiol* **14:** 769–776.

O'Doherty JP, Kringelbach ML, Rolls ET, Hornak J, Andrews C. 2001. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci* **4:** 95–102.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. 2006. Dopamine-dependent prediction errors underpin reward-seeking behavior in humans. *Nature* **442:** 1042–1045.

Robinson OJ, Frank MJ, Sahakian BJ, Cools R. 2010. Dissociable responses to punishment in distinct striatal regions during reversal learning. *NeuroImage* **51:** 1459–1467.

Schultz W. 2007. Multiple dopamine functions at different time courses. *Annu Rev Psychol* **30:** 259–288.

Seger CA. 2008. How do the basal ganglia contribute to categorization? Their role in generalization, response selection, and learning via feedback. *Neurosci Biobehav Rev* **32:** 265–278.

Seger CA, Peterson EJ, Cincotta CM, Lopez-Paniagua D, Anderson CW. 2010. Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and Granger causality modeling. *NeuroImage* **50:** 644–656.

Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS. 2004. Temporal difference models describe higher-order learning in humans. *Nature* **429:** 664–667.

Seymour B, O'Doherty JP, Klotzenburg M, Wiech K, Frackowiak R, Friston KJ, Dolan R. 2005. Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nat Neurosci* **8:** 1234–1240.

Seymour B, Daw N, Dayan P, Singer T, Dolan RJ. 2007. Differential encoding of losses and gains in the human striatum. *J Neurosci* **27:** 4826–4831.

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, Graybiel AM, Suzuki WA, Brown EN. 2004. Dynamic analysis of learning in behavioral experiments. *J Neurosci* **24:** 447–461.

Stark CEL, Okado Y. 2003. Making memories without trying: Medial temporal lobe activity associated with incidental memory formation during recognition. *J Neurosci* **23:** 6748–6753.

Sutton RS, Barto AG. 1998. *Reinforcement learning: An introduction*. The MIT Press, Cambridge, MA.

Talairach J, Tournoux P. 1988. *A coplanar sterotaxic atlas of the human brain*. Thieme Medical, New York.

Thorndike EL. 1933. A proof of the law of effect. *Science* **77:** 173–175.

Tom SM, Fox CR, Trepel C, Poldrack RA. 2007. The neural basis of loss aversion in decision-making under risk. *Science* **315:** 515–518.

Tricomi EM, Fiez JA. 2008. Feedback signals in the caudate reflect goal achievement on a declarative memory task. *NeuroImage* **41:** 1154–1167.

Tricomi EM, Delgado MR, Fiez JA. 2004. Modulation of caudate activity by action contingency. *Neuron* **41:** 281–292.

Watkins CHCJ. 1989. Learning with delayed rewards. PhD thesis, Cambridge University, Cambridge, UK.

Wheeler EZ, Fellows LK. 2008. The human ventromedial frontal lobe is critical for learning from negative feedback. *Brain* **131:** 1323–1331.

Widrow B, Hoff ME. 1960. Adaptive switching circuits. *IRE Wescon Convention Record*. 96–104.

Yacubian J, Gläscher J, Schroeder K, Sommer T, Braus DF, Büchel C. 2006. Dissociable systems for gain- and loss-related value predictions and errors of prediction in the human brain. *J Neurosci* **26:** 9530–9537.

Yin HH, Knowlton BJ, Balleine BW. 2004. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci* **19:** 181–189.

Yin HH, Knowlton BJ, Balleine BW. 2005. Blockade of NMDA receptors in the dorsomedial striatum prevents action-outcome learning in instrumental conditioning. *Eur J Neurosci* **22:** 505–512.