

Amygdala and Ventral Striatum Make Distinct Contributions to Reinforcement Learning

Highlights

- The amygdala is necessary for deterministic and stochastic reinforcement learning
- The ventral striatum is only necessary for learning stochastic reward associations
- Amygdala and ventral striatum lesions decrease choice consistency
- Ventral striatum lesions hasten choice reaction times leading to more errors

Authors

Vincent D. Costa, Olga Dal Monte,
Daniel R. Lucas, Elisabeth A. Murray,
Bruno B. Averbeck

Correspondence

vincent.costa@nih.gov

In Brief

The amygdala is not ascribed a role in reinforcement learning despite its known involvement in Pavlovian processes that influence choice. Costa et al. demonstrate that the amygdala and ventral striatum are both critical in learning to choose based on past outcomes.



Amygdala and Ventral Striatum Make Distinct Contributions to Reinforcement Learning

Vincent D. Costa,^{1,2,*} Olga Dal Monte,¹ Daniel R. Lucas,¹ Elisabeth A. Murray,¹ and Bruno B. Averbeck¹

¹Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892-4415, USA

²Lead Contact

*Correspondence: vincent.costa@nih.gov

<http://dx.doi.org/10.1016/j.neuron.2016.09.025>

SUMMARY

Reinforcement learning (RL) theories posit that dopaminergic signals are integrated within the striatum to associate choices with outcomes. Often overlooked is that the amygdala also receives dopaminergic input and is involved in Pavlovian processes that influence choice behavior. To determine the relative contributions of the ventral striatum (VS) and amygdala to appetitive RL, we tested rhesus macaques with VS or amygdala lesions on deterministic and stochastic versions of a two-arm bandit reversal learning task. When learning was characterized with an RL model relative to controls, amygdala lesions caused general decreases in learning from positive feedback and choice consistency. By comparison, VS lesions only affected learning in the stochastic task. Moreover, the VS lesions hastened the monkeys' choice reaction times, which emphasized a speed-accuracy trade-off that accounted for errors in deterministic learning. These results update standard accounts of RL by emphasizing distinct contributions of the amygdala and VS to RL.

INTRODUCTION

Reinforcement learning (RL) refers to learning the value of states in the environment and actions taken in those states. A common assumption about RL is that dopamine neurons in the midbrain signal reward prediction errors (RPEs; Schultz, 2015), which drive plasticity in the striatum to facilitate learning (e.g., Collins and Frank, 2014; Gurney et al., 2015; Houk et al., 1995). This view is supported by evidence that RPEs are correlated with ventral striatum (VS) blood-oxygen-level-dependent (BOLD) responses (e.g., O'Doherty et al., 2003, 2004; Rutledge et al., 2010) and fluctuations in striatal dopamine levels (e.g., Day et al., 2007; Hart et al., 2014). However, VS inactivation or lesions do not impair instrumental learning or conditioned reinforcement (Cardinal et al., 2002; Floresco, 2015; Stern and Passingham, 1996). This raises a long-standing question about the causal role of the VS in RL. It also highlights a need to understand

how other parts of the mesolimbic dopamine system contribute to RL.

The amygdala is often overlooked or assigned a modulatory role in RL (Haber and Behrens, 2014). The amygdala, however, receives dopaminergic inputs (Haber and Fudge, 1997) and is highly involved in Pavlovian learning of appetitive and aversive associations (Baxter and Murray, 2002; Janak and Tye, 2015; Moscarello and LeDoux, 2014). Although the amygdala exerts Pavlovian control over operant behavior (Cardinal et al., 2002; Robinson et al., 2014; Stuber et al., 2011), it remains unclear whether it contributes directly to choice behavior. There is evidence across species both in favor of (Hampton et al., 2007; Rygula et al., 2015; Seymour and Dolan, 2008) and against (Izquierdo et al., 2013; Izquierdo and Murray, 2007; Jang et al., 2015) the amygdala playing a role in RL during choice tasks.

To determine the causal roles of the amygdala and VS in RL, the present study used computational modeling to quantify the effect of targeted excitotoxic lesions of the amygdala or VS on monkeys' choice behavior in a two-arm bandit task. We show that VS is necessary for learning stochastic reward associations and that the amygdala contributes substantially to RL in both stochastic and deterministic environments.

RESULTS

We tested rhesus macaques on a two-armed bandit reversal learning task using deterministic or stochastic reward schedules (Figure 1A). The subjects included four monkeys with bilateral excitotoxic lesions of the amygdala (Figure 1B), three monkeys with bilateral excitotoxic lesions of the VS (Figure 1C), and four unoperated controls. The monkeys were first tested on randomly interleaved blocks following one of three probabilistic reward schedules: 80%/20%, 70%/30%, and 60%/40%. Learning was impaired in both lesion groups compared to controls. To determine whether their deficits were specific to the stochastic task, we next tested the monkeys on the same task using a deterministic reward schedule (100%/0%). The task was carried out in blocks of 80 trials. On each trial, a pair of visual cues were displayed as choice options. One cue was associated with a greater probability of reward delivery than the other. The monkeys were allowed to select one cue per trial by making a saccade to the selected cue. By sampling the stimulus-reward outcomes across trials, they could determine which of the two

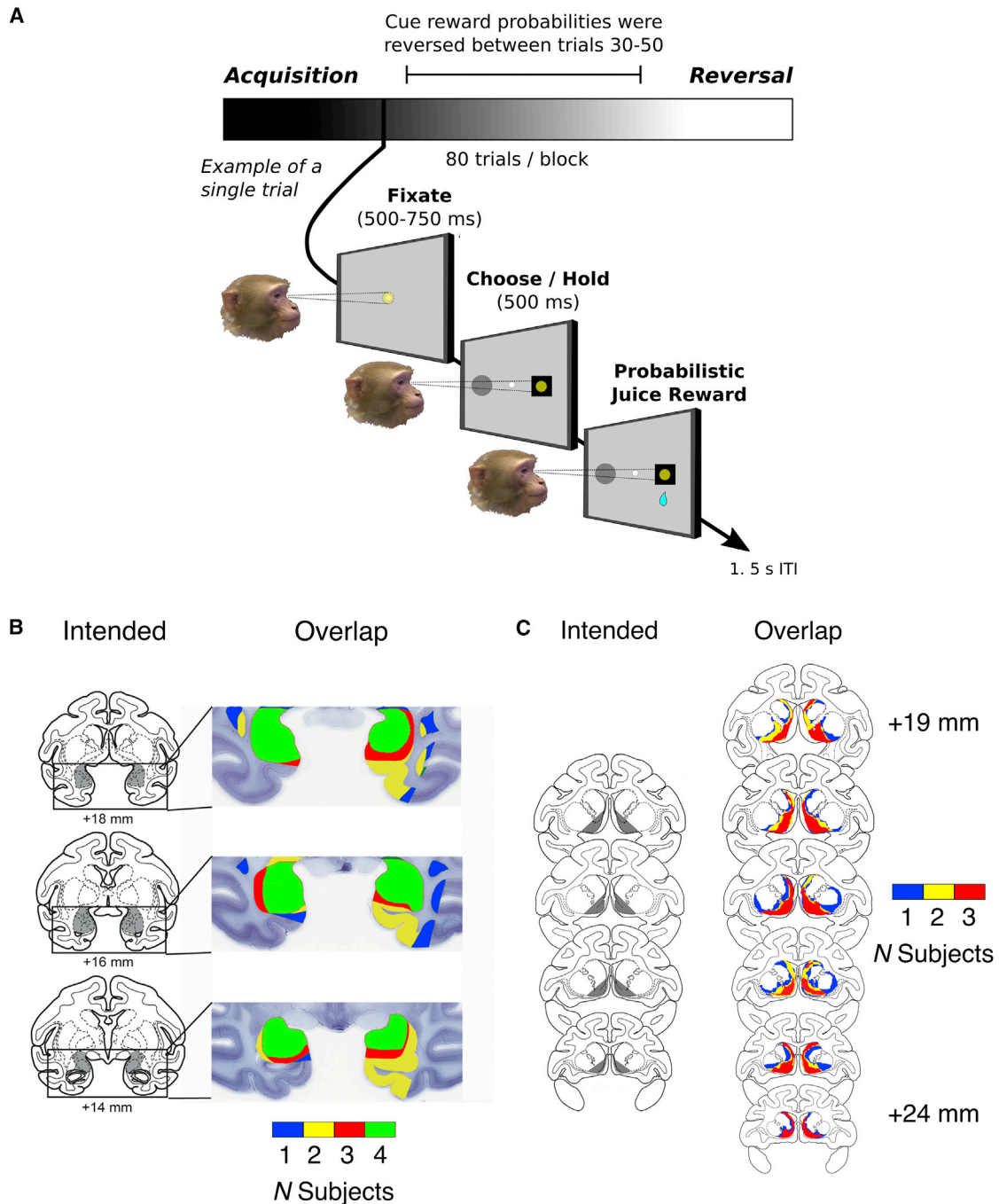


Figure 1. Two-Arm Bandit Task and Lesion Maps

(A) The task unfolded over a block of 80 trials. In each trial, the animals first fixated on a central point, followed by the presentation of two stimuli left and right of fixation. The animals selected one of the stimuli by making a saccade to it and fixating on the chosen cue. In each block, one object was rewarded more often than the other according to the current schedule (100%/0%, 80%/20%, 70%/30%, or 60%/40%). Stimulus-reward mappings were reversed randomly between trials 30 and 50. A new stimulus pair was introduced at the start of each block, and multiple blocks were completed per session.

(B and C) Lesion extent mapped for animals with bilateral excitotoxic lesions of amygdala (B) or ventral striatum (C).

cues was more often rewarded. The stimulus-reward mappings of the two options were reversed once per block on a randomly chosen trial between trials 30 and 50. At the end of the block, two

new cues were introduced (signaling the start of a new block), and the monkeys again had to discover by trial and error which of the two options was more often rewarded.

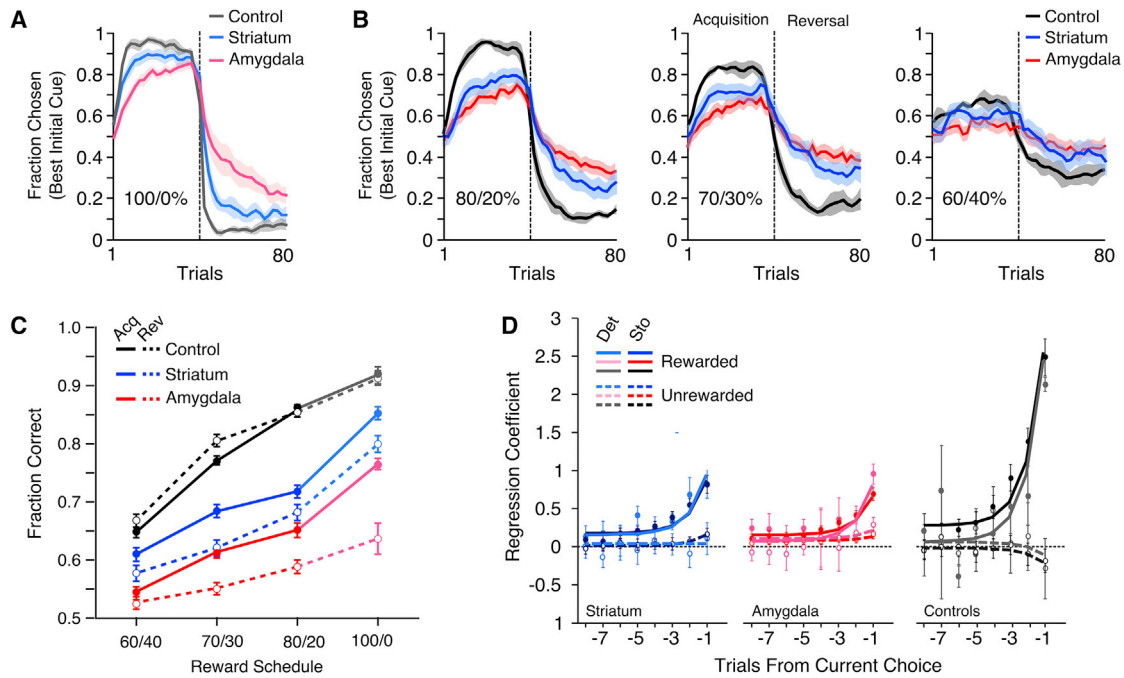


Figure 2. Choice Behavior for the Lesion and Control Groups

(A and B) Fraction of times each group chose the cue with a higher initial value in the deterministic (A) and stochastic (B) tasks. Dark lines show means, and shaded regions show ± 1 SEM (computed across sessions). Dotted vertical lines show reversal point. Choice curves were smoothed with a moving average window of six trials. Because the number of trials before and after acquisition varied across blocks, trial number was normalized to be between 0 and 1 for each phase and then averaged.

(C) Overall fraction of correct choices for each group by reward schedule and learning phase (see Table S1).

(D) Logistic regression coefficients (symbols) predicting the monkeys' current choice based on previous trials and outcomes. Two predictors were used to indicate whether prior selection of each option was rewarded or unrewarded. Positive regression coefficients for either predictor indicate a greater likelihood of choosing an option that resulted in reward or no reward. Lines indicate exponential fits of the estimated regression coefficients.

To assess motivation, we quantified the percentage of aborted trials per session due to a failure to acquire/hold central fixation or a failure to make a choice-related saccade. The lesion (VS: 12.83%; amygdala: 12.78%) and control (11.16%) groups did not differ in the percentage of trials aborted due to fixation errors ($F(2,204) = 1.77$, $p = 0.173$). In addition, the lesion (VS: 0.31%; amygdala: 0.26%) and control (0.01%) groups did not differ in the percentage of trials aborted due to choice errors ($F(2,205) = 1.82$, $p = 0.165$). We also computed how quickly the monkeys acquired central fixation. Fixation reaction times (RTs) were faster in the VS lesion group compared to the amygdala ($t(197) = 17.4$, $p < 0.001$) or control ($t(157) = 16.23$, $p < 0.001$) groups. Fixation RTs were similar in the amygdala and control groups ($t(204) = -1.15$, $p = 0.251$; Table S1). Therefore, lesion effects on RL are not attributable to reduced motivation.

Choice Behavior

To visualize the monkeys' choice behavior, we aligned the trials in each block around the reversal point estimated by a Bayesian change-point analysis (Figure 2; Supplemental Experimental Procedures). In both the deterministic (Figure 2A) and stochastic (Figure 2B) tasks, monkeys learned to select the better option and to switch their choice behavior once they detected the contingency reversal. To quantify lesion effects on choice behavior, we computed the fraction of correct choices in the acquisition

and reversal phases (Figure 2C). Group performance differed in the deterministic task when averaged across acquisition and reversal ($F(2,41) = 74.59$, $p < 0.001$). The striatal lesion group performed worse than the controls ($t(19) = -7.79$, $p < 0.001$), and the monkeys with amygdala lesions performed worse than those with striatal lesions ($t(33) = -6.23$, $p < 0.001$).

Choice accuracy also differed by group in the stochastic task (Figure 2C; $F(2,232) = 289.71$, $p < 0.001$). Again, monkeys with striatal lesions chose the correct option less often than controls ($t(145) = -13.77$, $p < 0.001$) but more often than animals with amygdala lesions ($t(156) = 7.41$, $p < 0.001$). The control group also showed improved performance in the reversal phase compared to the acquisition phase, whereas in the VS and amygdala lesion groups, performance declined in the reversal phase (group \times phase, $F(2,232) = 15.57$, $p < 0.001$). Although the reward schedule influenced performance ($F(2,461) = 407.57$, $p < 0.001$), controls showed greater linear improvement relative to the VS ($t(142) = 7.7$, $p < 0.001$) or amygdala ($t(160) = 11.48$, $p < 0.001$; group \times schedule, $F(4,461) = 29.05$, $p < 0.001$) lesion groups. Schedule-related performance increases were equivalent in the two lesion groups ($t(154) = 1.53$, $p = 0.131$).

We also used a logistic regression model to quantify how previous trial choices and outcomes influenced the monkeys' ongoing choice behavior (Figure 2D). A positive regression

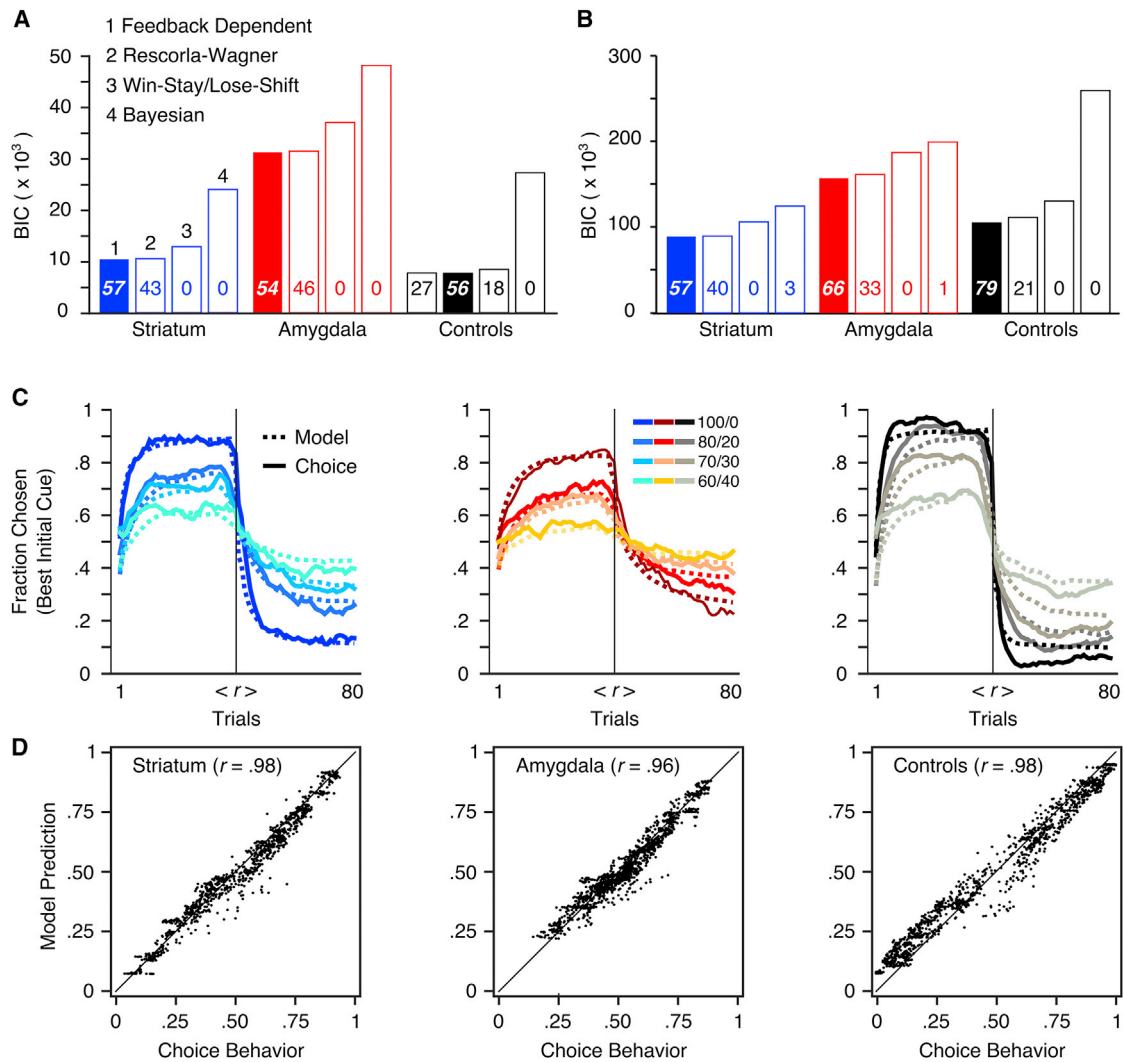


Figure 3. Model Comparisons Indicated that Choice Behavior Was Best Fit by a Feedback-Dependent RL Model

(A and B) Bayesian information criterion (BIC) estimates for different models fit to each groups' choice behavior in the deterministic (A) and stochastic (B) tasks (see Table S2). Filled bars indicate the model with the lowest BIC estimate summed across sessions. Numbers within each bar represent percentage of sessions in which BIC was lowest for that model.

(C) Feedback-dependent model predictions overlaid on choice behavior for each group and reward schedule.

(D) Scatterplots of trial-by-trial estimates of model predictions and monkeys' choice behavior, averaged per session.

coefficient indicated that the monkeys chose the same rewarded or unrewarded option that they chose on a past trial. In both tasks, the monkeys' current choices were influenced more by past choices that were rewarded versus unrewarded. In the deterministic (group \times outcome, $F(2,42) = 4.85$, $p = 0.009$) and stochastic (group \times outcome, $F(2,208) = 261.07$, $p < 0.001$) tasks, regression coefficients differed between the lesion and control groups as a function of outcome. Across tasks, the mean regression coefficient weighting rewarded outcomes was larger in controls compared to the VS (deterministic: $t(19) = 7.12$, $p < 0.001$; stochastic: $t(143) = 13.96$, $p < 0.001$) or amygdala (deterministic: $t(30) = 11.37$, $p < 0.001$; stochastic: $t(163) = 18.67$, $p < 0.001$) lesion groups. In the deterministic task, the mean regression coefficient for unrewarded outcomes

did not differ between the lesion and control groups ($F(2,42) = 3.09$, $p = 0.056$). In the stochastic task, however, unrewarded outcomes were weighted less in controls compared to the amygdala ($t(164) = -10.6$, $p < 0.001$) or VS ($t(144) = -6.52$, $p < 0.001$) lesion groups.

Reinforcement Learning Model Selection

We assessed whether an RL model best fit the monkeys' choice behavior compared to alternative models (Figure 3; Table S2). We compared Bayesian Information Criterion (BIC) estimates for the following four models: (1) a feedback-dependent RL model that incorporated different learning rate parameters for positive and negative outcomes, (2) a non-feedback-dependent RL model

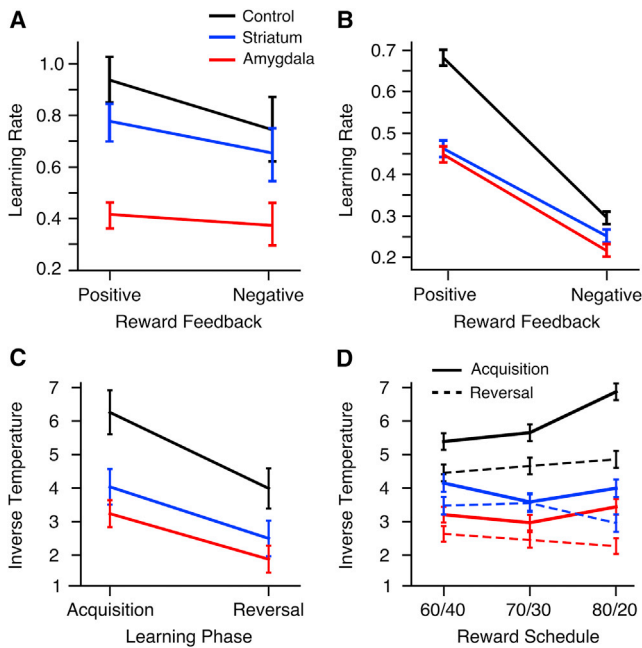


Figure 4. Learning Rates and Choice Consistency

(A and B) Feedback-dependent learning rates for the lesion and control groups when the choice behavior is fit with an RL model in the deterministic (A) and stochastic (B) tasks.

(C and D) Inverse temperature values for the lesion and control groups in the deterministic (C) and stochastic (D) tasks broken out by learning phase. Error bars are ± 1 SEM. See Table S3 for average parameter values.

(e.g., Rescorla-Wagner), (3) a win-stay/lose-shift model, and (4) a Bayesian ideal observer model (Supplemental Experimental Procedures). BIC estimates were lowest overall when we fit the feedback-dependent RL model to each group's choice behavior in the deterministic (Figure 3A) and stochastic (Figure 3B) tasks. Overlay of averaged model predictions and actual choice behavior showed that this model accurately fit the monkeys' choices (Figure 3C), which was also seen in scatterplots of predicted and measured choice behavior (Figure 3D). Therefore, we proceeded by comparing feedback-dependent RL model parameter fits for the lesion and control groups.

Feedback-Dependent Learning

We first analyzed how quickly the monkeys learned to update the expected value of each choice option based on positive and negative feedback. Positive and negative learning rates were fit for choices that resulted in reward or no reward, respectively. In the deterministic task (Figure 4A), positive and negative learning rates did not differ ($F(1,42) = 3.44$, $p = 0.071$). In the stochastic task (Figure 4B), learning was driven more by positive, rather than negative, feedback ($F(1,231) = 273.84$, $p < 0.001$), and mean learning rates were higher in the reversal phase compared to the acquisition phase ($F(1,234) = 45.16$, $p < 0.001$). There were no effects of reward schedule on learning rates in the stochastic task ($p > 0.6$ for all main effects and interactions).

Next, we examined group differences in learning rates. In the deterministic task, the groups differed in their mean learning

rate ($F(2,42) = 12.93$, $p < 0.001$; Figure 4A). Specifically, the mean learning rate was reduced in monkeys with amygdala lesions compared to those with VS lesions ($t(34) = -3.99$, $p < 0.001$) or intact controls ($t(30) = -4.23$, $p < 0.001$). The mean learning rate did not differ between the VS lesion and control groups ($t(19) = 0.85$, $p = 0.431$). Thus, in a deterministic reward environment, learning rates were only impaired in the monkeys with amygdala lesions.

When we analyzed the learning rates in the stochastic task, mean learning rates were lower in both the amygdala and VS lesion groups ($F(2,234) = 37.18$, $p < 0.001$; Figure 4B). Learning from positive feedback was reduced in the monkeys with amygdala ($t(165) = -8.64$, $p < 0.001$) or VS ($t(146) = -8.14$, $p < 0.001$) lesions compared to controls. Learning from negative feedback was also reduced in the amygdala lesion group ($t(166) = -3.42$, $p < 0.001$) relative to controls. Negative learning rates did not differ between the VS lesion and control groups ($t(146) = 2.4$, $p = 0.116$). To summarize, monkeys with amygdala lesions showed deficits in feedback-dependent learning in both the stochastic and deterministic tasks, whereas monkeys with VS lesions only showed learning deficits in the stochastic task relative to controls.

Because the VS lesion group was more affected in the stochastic task, we also examined how learning rates changed across the two tasks. For the stochastic task, we averaged learning rates across schedules and directly compared them to learning rates in the deterministic task. Negative learning rates were lower in the stochastic versus the deterministic task ($F(1,41) = 25.93$, $p < 0.001$), and this effect was consistent in the lesion and control groups (group \times task, $F(2,41) = 1.19$, $p = 0.314$). However, there were clear task-related differences between the lesion and control groups in learning from positive feedback (group \times task, $F(2,41) = 8.96$, $p < 0.001$). For the VS lesion ($t(11) = -4.58$, $p < 0.001$) and control ($t(8) = -6.59$, $p < 0.001$) groups, positive learning rates decreased in the stochastic versus the deterministic task. This was not the case for the amygdala lesion group, in which positive learning rates were equivalent in the stochastic and deterministic tasks ($t(22) = 0.43$, $p = 0.666$). Thus, task effects were confined to learning from positive feedback, supporting the differences found between the lesion and control groups in each task.

Choice Consistency

We next quantified choice consistency using the inverse temperature parameter β from the RL model. This parameter quantified how consistently the monkeys chose the higher-value option; therefore, smaller inverse temperature values were indicative of noisier choice behavior. For both the deterministic ($F(1,42) = 11.32$, $p = 0.002$) and stochastic ($F(1,232) = 39.74$, $p < 0.001$) tasks, choice consistency was higher in the acquisition phase compared to the reversal phase (Figures 4C and 4D).

In both the deterministic ($F(2,42) = 11.16$, $p < 0.001$) and stochastic ($F(2,228) = 39.74$, $p < 0.001$) tasks, choice consistency was reduced in the lesion groups relative to controls (Figures 4C and 4D). Compared to controls, monkeys with amygdala lesions (deterministic: $t(31) = -4.42$, $p < 0.001$; stochastic: $t(155) = 14.28$, $p < 0.001$) or VS lesions (deterministic: $t(20) = -2.62$, $p = 0.016$; stochastic: $t(143) = -6.7$, $p < 0.001$) less consistently chose the higher-value option. In the deterministic

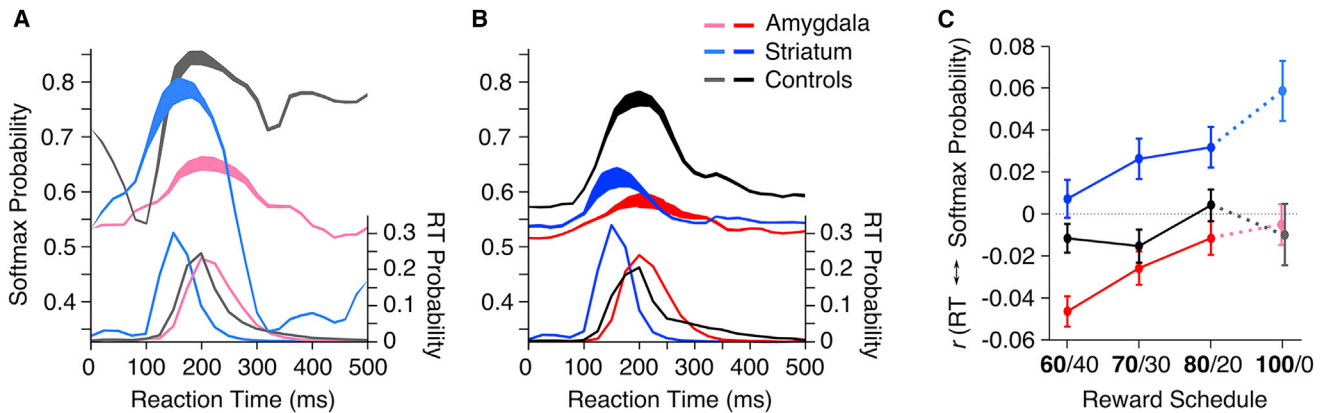


Figure 5. Striatal Lesion Effects on Choice Reaction Times and Speed-Accuracy Trade-Off

(A and B) Upper plots with the ordinate oriented to the left illustrate how the chosen softmax probability varied as a function of choice RT in the deterministic (A) and stochastic (B) tasks. Larger softmax probabilities indicate more consistent selection of higher-value option. Gaussian kernel regression was used to estimate mean softmax probabilities as a function of RT (see [Experimental Procedures](#)). The lower plots with the ordinate oriented to the right show RT probability distributions for each group (20 ms bins; see [Table S1](#)). Line thickness of the estimated softmax probability indicates relative density of the RT probability distribution at that bin.

(C) Linear correlation coefficient between chosen softmax probabilities and RTs for each reward schedule and lesion group.

task, the lesion groups showed equivalent decreases in choice consistency ($t(33) = -1.82$, $p = 0.078$), whereas in the stochastic task, the amygdala lesion group was less consistent than the VS lesion group ($t(158) = -2.44$, $p = 0.016$). Also, in the stochastic task, the reward schedule modulated choice consistency only in the control group (group \times schedule, $F(4,476) = 2.51$, $p = 0.041$). In controls (schedule, $F(2,155) = 5.09$, $p = 0.007$), choice consistency was higher in the 80%/20% schedule compared to the 70%/30% ($t(156) = 2.38$, $p = 0.018$) or 60%/40% ($t(156) = 3.03$, $p = 0.002$) schedules but did not differ in the latter two schedules.

When we examined how inverse temperature values changed across the two tasks, we found that they were slightly higher in the stochastic task compared to the deterministic task ($F(1,216) = 4.79$, $p = 0.031$), and this effect was consistent in the lesion and control groups (group \times task, $F(2,181) < 1$, $p = 0.723$). Therefore, the monkeys with VS and amygdala lesions exhibited noisier choice behavior in general.

Reaction Time Effects on Choice Consistency

Choice RTs differed between the lesion and control animals ([Figure 5](#); [Table S1](#)). The monkeys with VS lesions were faster compared to the monkeys with amygdala lesions or controls in both the deterministic ($F(2,42) = 84.81$, $p < 0.001$) and the stochastic ($F(2,228) = 802.6$, $p < 0.001$) tasks. Consistent with this result, the choice RT distribution of the VS lesion group was shifted left compared to the control (Kolmogorov-Smirnov [KS] test deterministic: $p < 0.001$; stochastic: $p < 0.001$) or amygdala lesion (deterministic: $p < 0.001$; stochastic: $p < 0.001$) groups. Choice RTs did not differ between the amygdala lesion and control groups in either task (all p values > 0.14). In addition, RTs did not differ by task ($F(1,250) < 1$, $p = 0.797$), and there were no effects of reward schedule on choice RTs in any group.

The observation of faster choice RTs in the VS lesion group suggests that their less consistent selection of high-value options was related to a speed-accuracy trade-off (i.e., faster re-

sponses led to less accurate choices). To visualize how choice RTs were related to choice accuracy, we extracted the softmax probability for each of the monkeys' choices from the RL model. These estimates indicated the probability that monkeys' choices would be rewarded on a given trial. We then used Gaussian kernel regression to determine the function relating RTs to choice probabilities for each block, sampled the function in equally spaced 20 ms RT bins, and averaged the function estimates across blocks. This was done for both the deterministic and stochastic tasks ([Figures 5A](#) and [5B](#)). In each task, it was clear that choice accuracy increased along with RT from about 100 ms up to approximately 225 ms, after which it plateaued before decreasing as RT further increased. This relationship between RTs and choice probabilities was consistent in all three groups, although there were vertical shifts in the curves.

To quantify the relationship between RTs and choice probabilities, we computed the correlation coefficient for each block of trials and averaged them by reward schedule and session ([Figure 5C](#)). The correlation magnitude increased along with the reward schedule ($F(2,457) = 4.16$, $p = 0.016$). The lesion and control groups differed in the magnitude and sign of the correlation coefficient relating choice RT and accuracy. In the deterministic task (group, $F(2,42) = 5.87$, $p = 0.006$), direct comparison of correlation coefficients by group confirmed a larger, positive correlation in the VS lesion group compared to the amygdala lesion ($t(37) = 2.46$, $p = 0.023$) or control ($t(20) = 3.19$, $p = 0.005$) groups. Similarly, in the stochastic task, collapsed across schedules, there was a larger, positive correlation in the VS lesion group compared to the amygdala ($t(153) = 7.46$, $p < 0.001$) or control ($t(143) = 3.12$, $p < 0.001$) groups.

The VS lesion group's choice RTs were faster overall and showed a speed-accuracy trade-off. Their less consistent selection of high-value options might therefore be due to the VS lesion

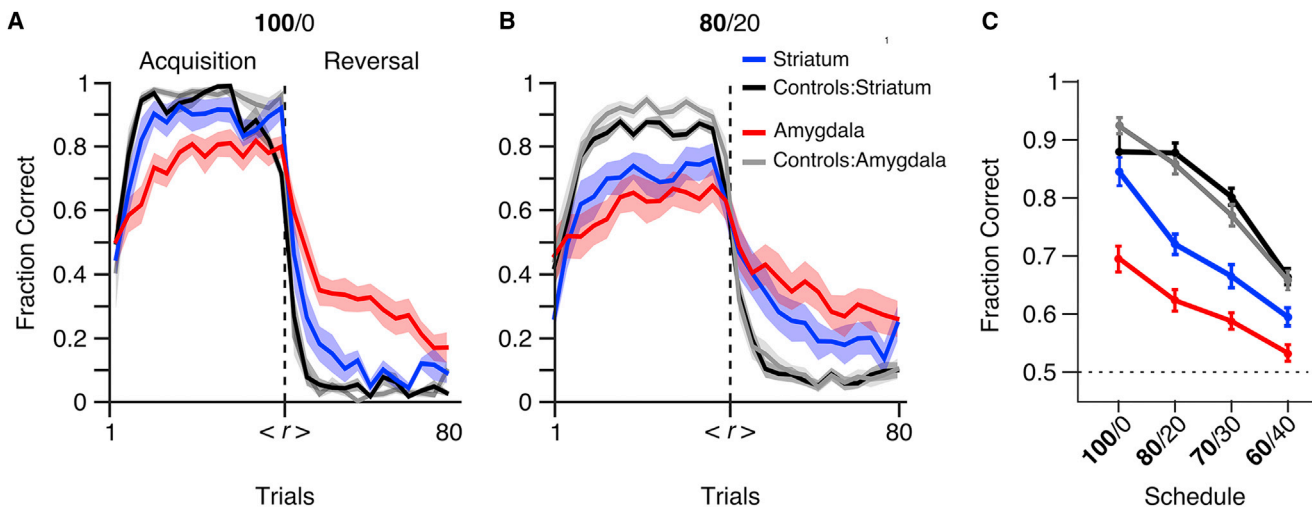


Figure 6. Faster RTs in the VS Lesion Group Modulated Choice Consistency in the Deterministic Task

(A) Fraction of times the monkeys in the lesion and control groups chose the cue with a higher initial value in the deterministic task when RT distributions were matched. RTs were matched pairwise between each of the animals in the lesion and control groups. Specifically, the data labeled Striatum and Controls:Striatum plot the behavioral performance when the RT distributions of these two groups were matched and likewise for the data labeled Amygdala and Controls:Amygdala. (B) Fraction of times the animals in each group chose the cue with a higher initial value in the 80%/20% blocks of the stochastic task when RT distributions were matched as described in (A).

(C) Fraction of correct choices for the lesion and control groups broken out by reward schedule when group RT distributions were matched. Shaded regions and error bars indicate ± 1 SEM.

group acting impulsively—especially in the deterministic task, where striatal lesions had a minor effect on choice consistency relative to controls. To test this hypothesis, we matched the choice RT distributions of the lesion groups and controls pairwise and computed the fraction of times the monkeys in each group chose the higher-value option in each task.

When the choice RT distributions of the VS and control groups were matched in the deterministic task (Figures 6A and 6C), the fraction of correct choices was equivalent in the VS lesion and control groups ($t(19) = 1.294$, $p = 0.211$). This indicates that the effect of VS lesions on choice consistency in the deterministic task was limited to trials in which the animals responded too quickly. In the stochastic task, however, equating group choice RT distributions did not mitigate previously identified differences in choice accuracy between the VS lesion group and controls ($F(1,141) = 36.6$, $p < 0.001$; Figures 6B and 6C). For example, in the 80%/20% schedule, choice performance was consistently lower in the VS lesion group compared to controls after matching the choice RT distributions ($t(143) = -5.73$, $p < 0.001$; Figure 6B).

The amygdala lesion and control groups did not differ in their choice RT distributions. Thus, we did not expect matching these groups' choice RTs would eliminate differences in choice consistency. Even after choice RTs were matched, monkeys with amygdala lesions showed decreased choice consistency relative to controls in the deterministic ($t(30) = -5.88$, $p < 0.001$) and stochastic ($F(1,149) = 196.6$, $p < 0.001$) tasks.

Pearce-Hall Error Learning

RPEs can promote learning directly or indirectly by altering attention to reward predictive cues (Roesch et al., 2012). The amygdala appears to utilize error signals to modulate attention during

learning (Roesch et al., 2012; Holland and Schiffino, 2016) versus direct encoding of RPEs in VS (Li et al., 2011) and dopamine neurons (Roesch et al., 2012). Therefore, we fit a hybrid Pearce-Hall (PH) model (Li et al., 2011) to the monkeys' choices to examine how amygdala and VS lesions affected attentional gating of learning. This model contains an associability parameter that modulates the learning rate as a function of the absolute magnitude of past RPEs. When RPEs are large, associability (i.e., attention) is increased. The model has two free parameters, κ and η (Table S3). The κ parameter refers to a fixed scalar that modulates value updating of each cue, akin to the learning rate parameter in the feedback-dependent RL model (to control model complexity, we did not fit separate learning rate parameters for positive and negative feedback). The η parameter is a decay parameter that controls the temporal dynamics of associability over time.

The fitted κ parameters were consistent with the learning rate results found using the feedback-dependent RL model. In the deterministic task, κ (Figure 7A) was reduced in the monkeys with amygdala lesions ($t(30) = 2.32$, $p = 0.027$) relative to controls. For the VS lesion group, the value of κ did not differ from the amygdala lesion ($t(33) < 1$, $p = 0.517$) or control ($t(19) = -1.56$, $p = 0.136$; group, $F(2,42) = 2.62$, $p = 0.084$) groups. In the stochastic task, by contrast, κ was reduced in both the amygdala ($t(163) = -5.13$, $p < 0.001$) and VS ($t(145) = -3.78$, $p < 0.001$) lesion groups compared to controls ($F(2,233) = 14.16$, $p < 0.001$).

In the deterministic task, η values only marginally differed between the lesion and control groups ($F(2,41) = 3.08$, $p = 0.057$; Figure 7B). Planned post-hoc comparisons of η did indicate that it was reduced in monkeys with amygdala lesions relative to those with VS lesions ($t(33) = -2.6$, $p = 0.0138$). However, η values in either lesion group were equivalent to that estimated

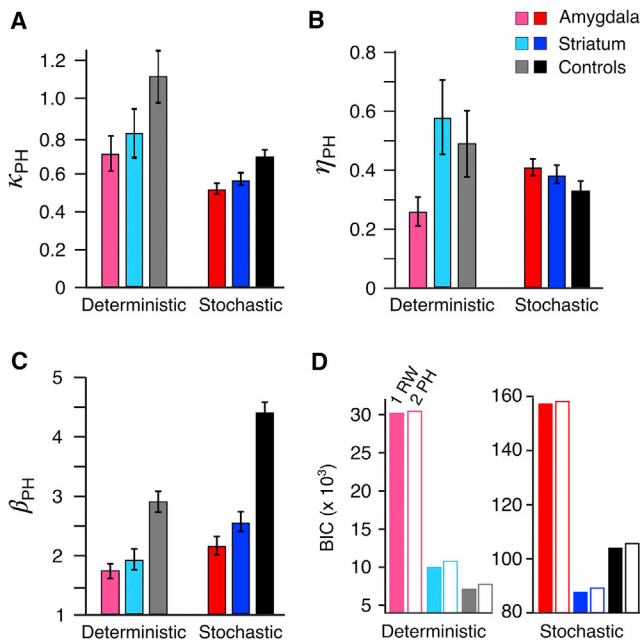


Figure 7. Pearce-Hall Model Parameter Fits to Choice Behavior (A–C) Fitted parameter estimates for κ (fixed learning rate; A), η (decay rate of associability; B), and β (choice consistency; C). See Table S3 for average parameter values. Error bars indicate ± 1 SEM. (D) BIC estimates for each group and task when choices were fit using a PH or RW model.

for controls (p values > 0.13). In the stochastic task, there were again only marginal differences in η values between the lesion and control groups ($F(2,235) = 2.94$, $p = 0.054$). Planned post-hoc comparisons indicated a larger η value in the amygdala lesion group compared to controls ($t(166) = 2.49$, $p = 0.0163$) but not in comparison to the VS lesion group ($t(157) < 1$, $p = 0.345$). Thus, we found relatively weak evidence that the amygdala or VS affected associability.

Because we fit the PH model to the monkeys' choice behavior, we also fit an inverse temperature parameter to measure choice consistency. Similar to our results using the feedback-dependent RL model, inverse temperature values were lower in the animals with lesions of the amygdala or VS compared to controls both in the deterministic ($F(2,41) = 6.01$, $p = 0.005$) and stochastic ($F(2,212) = 62.13$, $p < 0.001$) tasks.

The PH hybrid model is an extension of the Rescorla-Wagner (RW) learning algorithm. The RW algorithm has a fixed learning rate and no associability term. To determine which model (PH or RW) was more parsimonious, we computed BIC estimates to summarize how well each model fit the monkeys' choices (Figure 7D). In each task and in all three groups, the RW model with one less parameter yielded a more parsimonious fit compared to the PH model.

DISCUSSION

We combined computational modeling with selective lesions of the amygdala or VS to determine the causal role of these regions

in two facets of RL, feedback-dependent learning and choice consistency. Relative to controls, learning from positive feedback was generally impaired in monkeys with amygdala lesions, whereas monkeys with VS lesions were only impaired when learning stochastic associations. Both the VS and amygdala lesion groups chose higher-value options less consistently than controls. Although in the deterministic task, decreased consistency in the VS lesion group was attributable to their overall faster choice RTs. In addition, we did not find strong evidence that the primate amygdala critically encodes Pearce-Hall associability in bandit tasks.

Ventral Striatum Is Only Necessary for Learning Stochastic Reward Associations

Neurocomputational accounts of RL implicate VS in value updating (Collins and Frank, 2014; Schultz, 2015). On the other hand, inactivation of VS in rodents does not impair instrumental learning or conditioned reinforcement (Cardinal et al., 2002; Dalton et al., 2014; Floresco, 2015). Likewise, excitotoxic VS lesions in monkeys do not affect particular forms of stimulus-reward learning (Stern and Passingham, 1996). In all these cases, only deterministic learning was examined. Therefore, our finding that VS lesions had a greater impact on feedback-dependent learning in the stochastic versus the deterministic task highlights the necessity of the VS in learning probabilistic reward associations. Moreover, it reconciles the results of prior VS inactivation studies with existing theories about neural implementations of RL.

It was recently shown that nucleus accumbens (NAc) shell, but not core, preferentially mediates probabilistic learning (Dalton et al., 2014). In rats, shell inactivation reduces win-stay behavior and increases errors to criterion. This result parallels our findings that VS lesions reduced sensitivity to positive feedback and decreased choice consistency in the stochastic task. Although our VS lesions subsume the shell and core (Friedman et al., 2002), future research should determine whether our effects are specifically due to shell lesions.

Although we confirm that VS mediates stochastic RL, this does not necessarily mean that value updating in the striatum relies on dopaminergic RPEs. When control monkeys were tested on the same stochastic, two-armed bandit task, learning from positive feedback was heightened when levodopa was used to enhance dopaminergic function compared to dopamine antagonism with haloperidol. Yet, neither drug modulated the monkeys' learning rates compared to saline (Costa et al., 2015). Systemic blockade of the dopamine transporter similarly does not modulate learning rates during a probabilistic, three-arm bandit task (Costa et al., 2014). Thus, whether the role of the VS in stochastic learning is dopamine related remains an open question.

Another issue is whether stochastic learning deficits in the VS lesion group are specific to acquiring stimulus-outcome associations, as opposed to action-outcome associations. D_2 receptor antagonism in dorsal striatum disrupts action selection based on reinforcement of past choices (Lee et al., 2015) and dopamine terminals in dorsomedial striatum preferentially encode choices relative to specific movements (Parker et al., 2016). Thus, adaptations of the current task to study action-outcome learning may reveal greater dependence on dorsal versus ventral striatum.

Ventral Striatum Modulates Speed-Accuracy Trade-Off

VS lesions hastened choice RTs compared to the amygdala lesion or control groups. The faster choice RTs in the VS lesion group also made them more susceptible to a speed-accuracy trade-off, particularly in the deterministic task. In rodents, NAc core lesions increase delay discounting (Cardinal et al., 2001) and impair performance when reward is contingent on withholding a response for a fixed period of time (Pothuizen et al., 2005). VS lesions also disrupt the temporal specificity of dopaminergic RPEs induced by changes in reward timing (Takahashi et al., 2016). Altogether, these findings suggest that VS regulates temporal integration of reward information to guide choice behavior.

Amygdala Directly Contributes to Reinforcement Learning

The amygdala plays a fundamental role in Pavlovian learning and the valuation of sensory stimuli (Baxter and Murray, 2002; Janak and Tye, 2015; Salzman and Fusi, 2010; Seymour and Dolan, 2008). Pavlovian associations encoded in amygdala are also known to modulate instrumental behavior (Cardinal et al., 2002; Corbit and Balleine, 2005; Holland and Gallagher, 2003) through circuit interactions involving the ventral tegmental area and VS (Corbit et al., 2007; Stuber et al., 2011). Therefore, the amygdala is well positioned to guide choice behavior.

Yet, it is unclear whether the amygdala directly contributes to value-guided decision making. This is in part due to mixed evidence across species about the necessity of the amygdala in mediating object discrimination reversal learning (Chau et al., 2015; Hampton et al., 2007; Izquierdo et al., 2013; Izquierdo and Murray, 2007; Rudebeck and Murray, 2008; Stalnaker et al., 2007). Our results are consistent with evidence of impaired discrimination learning in human patients with amygdala damage (Hampton et al., 2007) and in monkeys with aspiration or radio frequency amygdala lesions (Jones and Mishkin, 1972; Schwartzbaum and Poulos, 1965). They are inconsistent with evidence that excitotoxic amygdala lesions do not disrupt object discrimination reversal learning (Izquierdo and Murray, 2007). Our data also counter evidence that the amygdala encodes inflexible value representations that impede reversal learning and credit assignment (Chau et al., 2015; Stalnaker et al., 2007).

The apparent contradiction between our results and past studies is likely due in part to methodological differences. For example, monkeys with excitotoxic amygdala lesions perform better than controls on a deterministic object reversal learning task (Jang et al., 2015; Rudebeck and Murray, 2008). This benefit is limited to the first few reversals (Jang et al., 2015), which suggests that amygdala contributions to RL may differ in contexts of unexpected versus expected uncertainty. Another difference is that prior studies tested animals with extensive, if not exclusive, training on deterministic stimulus-reward associations (Izquierdo and Murray, 2007). It is therefore possible that the monkeys in the amygdala lesion group might have shown little or no deficit in the deterministic task if they had only experienced deterministic stimulus-outcome pairings. However, the general and sizable deficit in learning from positive feedback seen in the amygdala lesion group argues against this idea.

Our results do support a nascent view of amygdala function that emphasizes its active role in decision making (Grabenhorst et al., 2012; Seymour and Dolan, 2008; Phelps et al., 2014). Amygdala activity tracks signals relevant to economic choice (Grabenhorst et al., 2012; Hernádi et al., 2015), model-based RL (Prévost et al., 2013), and abstract context representations (Saez et al., 2015). Human patients with amygdala damage show learning deficits in both stochastic and deterministic bandit tasks (Hampton et al., 2007). Also, serotonin depletion in the marmoset amygdala delays probabilistic discrimination learning by reducing sensitivity to reward and punishment (Rygula et al., 2015). Thus, when learning to choose between two objects, the amygdala is needed to form strong feedback-dependent stimulus-reward associations in order to identify and consistently choose objects that are most valuable. However, it is likely that value representations formed in the amygdala drive choice behavior via projections to the VS and dopamine subpopulations.

Lesion Effects on Pearce-Hall Error Learning in Choice Contexts

Pearce-Hall learning theory describes how associability (i.e., amount of attention paid to a cue) influences the rate at which conditioned associations are learned (Li et al., 2011; Pearce and Hall, 1980). Experiments designed around this view suggest that amygdala computes associability (Holland and Schiffrino, 2016; Roesch et al., 2012). Likewise, in the deterministic task, we found that associability was reduced in the monkeys with amygdala lesions relative to those with VS lesions. However, one caveat to this finding is that associability encoding in the lesion groups was equivalent to controls. Moreover, associability encoding did not differ between the amygdala and striatal lesion groups in the stochastic task. Therefore, it is difficult to draw firm conclusions about the relative roles of the amygdala and VS in signaling associability.

Previous studies have found that amygdala and VS differentially encode associability when associability is computed by fitting a hybrid PH model to Pavlovian-conditioned autonomic responses (Atlas et al., 2016; Li et al., 2011; Zhang et al., 2016). In these cases, conditioned responses were better fit by the PH model than the RW model. The opposite was true when we fit the same models to the monkeys' choice data. The RW model with one fewer parameter yielded more parsimonious fits than the hybrid PH model. Modeling learning in terms of preparatory versus consummatory responses may reveal differential roles for the amygdala in signaling associability (Zhang et al., 2016). We might also have detected deficits in PH error learning if the monkeys were tested on choice-related variants of serial prediction or unblocking tasks (Holland and Schiffrino, 2016). At the very least, our results highlight a need to better understand how the amygdala and VS contribute to PH error learning, particularly in stochastic environments where surprise-induced learning is not necessarily advantageous (Courville et al., 2006).

CONCLUSION

Traditional views of reinforcement learning have held that the ventral striatum integrates dopaminergic RPEs to facilitate learning but have not ascribed a specific role for the

amygdala. Our results suggest that standard accounts of RL should be revised to reflect a key role for the amygdala in learning to choose between stimuli and to account for the relative role of the VS in mediating stochastic versus deterministic learning. Moreover, since amygdala projections to VS preferentially facilitate appetitive behaviors (Namburi et al., 2015), our data offer a new perspective on how interactions between the amygdala and ventral striatum enable reinforcement learning.

EXPERIMENTAL PROCEDURES

Subjects

Eleven male rhesus macaques, weighing 6.5 to 10.5 kg, were studied. Three monkeys received bilateral excitotoxic lesions of the VS, four received bilateral excitotoxic lesions of the amygdala, and four were retained as unoperated controls. All monkeys were placed on water control for the duration of the study and, on test days, earned all of their fluid through performance on the task. All experimental procedures were performed in accordance with the Guide for the Care and Use of Laboratory Animals and were approved by the National Institute of Mental Health Animal Care and Use Committee.

Surgery

For detailed surgical information, see [Supplemental Experimental Procedures](#). In brief, each monkey was first implanted with a titanium head restraint device. The monkeys assigned to each lesion group then received targeted bilateral excitotoxic lesions of the amygdala or VS. All testing occurred after the lesion surgeries.

Lesion Assessment

For each monkey, lesions of amygdala or VS were quantitatively assessed from postoperative MRI scans. The extent of damage was evaluated from T2-weighted scans obtained within 10 days of surgery. For each operated animal, MRI scan slices were matched to drawings of coronal sections from a standard rhesus monkey brain at 1 mm intervals. Each lesion was subsequently plotted onto the standard sections to map its location and extent (Figure 1B).

Experimental Setup

The monkeys completed an average of 16.81 (SD = 6.72) blocks per session of a two-arm bandit task. Each block consisted of 80 trials and involved a single reversal of the stimulus-reward contingencies (Figure 1). On each trial, the monkeys had to first acquire and hold a central fixation point (500–750 ms). After the monkey fixated for the required duration, two stimuli appeared to the left and right (6° visual angle) of the central fixation point. Stimuli varied in shape and color, and their trial-to-trial screen locations were randomized. Monkeys chose between stimuli by making a saccade to one of the two stimuli and fixating on that cue for 500 ms. One stimulus was assigned a high reward probability and the other a low reward probability. Juice reward was delivered at the end of each trial according to the assigned reward schedule, followed by a fixed 1.5 s intertrial interval. A failure to acquire fixation within 5 s, to hold central fixation, or to make a choice within 1 s resulted in an immediate repeat of the previous trial. The trial on which the cue-reward mapping reversed within each block was selected randomly from a uniform distribution across trials 30 to 50. The reversal trial did not depend on the monkey reaching a performance criterion. The reward schedule was randomly selected at the start of each block and remained constant within a block. We ran two experiments. In the first experiment, we used three reward schedules: 80%/20%, 70%/30%, and 60%/40%. In the second experiment, we used a deterministic reward schedule (100%/0%).

The pair of stimuli, presented as choice options per block, always differed in color and shape. In addition, each stimulus combination was crossed with reward schedule and which shape was initially assigned a higher value (e.g., whether the blue square was the best choice before or after the reversal). This resulted in 12 block combinations in the deterministic version of the task and in 36 block combinations in the stochastic version of the task. Block

presentations were fully randomized without replacement. This ensured that a specific stimulus-reward combination was never repeated until all block combinations were experienced. Although combinations were potentially repeated across sessions, upon inspection, there was no evidence of improved performance across sessions.

Stimulus presentation and behavioral monitoring were controlled by a PC computer running MonkeyLogic (Asaad and Eskandar, 2008). Stimuli were displayed on an LCD monitor (1024 × 768 resolution) situated 40 cm from the monkeys' eyes. Eye movements were monitored at 1 kHz (Arrington Research). On rewarded trials, apple juice (0.085 mL) was delivered through a pressurized plastic tube gated by a computer-controlled solenoid valve (Mitz, 2005).

Task Training

All monkeys that received lesions were trained and tested following surgery. Initial training focused on having the monkeys learn the task structure using a deterministic reward schedule. They were then trained on a probabilistic version of the task until they routinely completed 15–20 blocks per session and demonstrated stable performance on interleaved blocks of the three probabilistic reward schedules. Data from these training sessions were not included in the reported analyses. Once testing began, the monkeys completed the stochastic and deterministic tasks in a separate series of daily sessions. All of the monkeys completed the deterministic task after they completed testing on the stochastic task. Prior to testing on the deterministic task, the lesion and controls groups did not differ in the number of sessions ($F(2,8) = 1.48$, $p = 0.234$) or total number of blocks completed ($F(2,8) < 1$, $p = 0.554$) on the stochastic task. In particular, the amygdala and VS lesion groups did not differ (number sessions: $t(5) < 1$, $p = 0.377$; total number of blocks: $t(5) < 1$, $p = 0.818$). Therefore, group differences in the deterministic task cannot be attributed to time in training.

Computational Modeling

We fit three models to the choice data. The first was a causal Bayesian model (see [Supplemental Experimental Procedures](#)) that we used to estimate reversal points in the stimulus-reward mappings. We then fit a stateless temporal difference learning model. Because it is stateless, this model is equivalent to an RW model (Rescorla and Wagner, 1972). Finally, we fit a hybrid PH model (Li et al., 2011; Pearce and Hall, 1980).

Reinforcement Learning Model

We split the trials in each block into acquisition and reversal phases using the reversal point calculated with the Bayesian model. We then fit separate RL models to each phase. This was done for each session and separately for each schedule. We used a standard RL model to estimate learning from positive and negative feedback, as well as to estimate the inverse temperature. Specifically, value updates were given by:

$$v_i(k+1) = v_i(k) + \alpha_f(R - v_i(k)) \quad \text{Equation 1}$$

where v_i is the value estimate for option i , R is the reward feedback for the current choice, and α_f is the feedback-dependent learning rate parameter, where f indexes whether the current choice was rewarded ($R = 1$) or not ($R = 0$). In other words, on each trial, α_f is one of two fitted values used to scale prediction errors based on reward feedback for the current choice. These value estimates were then passed through a logistic function to generate choice probability estimates:

$$d_1(k) = \frac{1}{1 + e^{\beta(v_2(k) - v_1(k))}}, \quad d_2(k) = 1 - d_1(k) \quad \text{Equation 2}$$

The likelihood is then given by:

$$f(D | \beta, \alpha_f) = \prod_k [d_1(k)c_1(k) + d_2(k)c_2(k)] \quad \text{Equation 3}$$

where $c_1(k)$ had a value of 1 if option 1 was chosen on trial k and $c_2(k)$ had a value of 1 if option 2 was chosen. Otherwise, they had a value of 0. Standard function optimization techniques were used to maximize the log-likelihood of the data given the parameters. As estimation can settle on local minima, we used 20 initial values for the parameters. The maximum of the log-likelihood across fits was then used.

When we fit the RL model to the acquisition phase, starting values were reset to 0.5 at the beginning of each block, unless the stimulus in the new block was one of the same stimuli from the previous block. This happened in some cases because we used a small set of stimuli that were paired and counterbalanced with reward schedule. When a stimulus in the new block was the same as the stimulus from the previous block, we carried the value estimate from the previous block forward. For the reversal phase, starting values for each block corresponded to end values of the corresponding acquisition phase.

Pearce-Hall Error Learning

We also fit a hybrid PH model (Li et al., 2011). In this model, learning is dynamically gated by two free parameters, κ and η . Value updates are given by:

$$v_i(k+1) = v_i(k) + \kappa\alpha(k)\delta \quad \text{Equation 4}$$

$$\delta = R - v_i(k) \quad \text{Equation 5}$$

$$\alpha(k+1) = \eta|\delta| + (1 - \eta)\alpha(k) \quad \text{Equation 6}$$

where $v_i(k)$ is the value estimate for option i , R is the reward feedback for the current choice, δ is the reward prediction error of the chosen option, κ is a constant learning rate parameter, $\alpha(k)$ is the associability of the chosen option, and η is a weighting factor that accounts for the gradual update in associability across trials. To limit free parameters in the model, we assigned each option a starting value of 0.5 and set $\alpha(0)$ to 1. Also, because of the structure of this model, it was not straightforward to fit it separately to the acquisition and reversal phases. Therefore, we fit one model across the entire block of trials.

Just as was done for the feedback-dependent RL model, value estimates were passed through a logistic function (Equation 2) to generate choice probability estimates. The likelihood is then given by:

$$f(D | \beta, \eta, \kappa) = \prod_k [d_1(k)c_1(k) + d_2(k)c_2(k)] \quad \text{Equation 7}$$

As described above, standard function optimization techniques were used to maximize the log of the likelihood of the data given the parameters.

For comparison with the PH model, we also fit an RW model to the choice data. This was done by setting η in the above model to a fixed value of 0. In this case, the RW model was also not fit separately to each phase.

Saccadic Reaction Times

We computed choice RTs on a trial-by-trial basis. Choice RTs were defined as the amount of time between the onset of the choice options and initiation of a saccade targeting either option. RT probability distribution functions were constructed (Figures 5A and 5B) by binning RTs in 20 ms bins.

Speed-Accuracy Trade-Off

To examine the speed-accuracy trade-off, we first extracted the chosen softmax probability, d_i , from the feedback-dependent RL model for each of the monkeys' choices. Kernel regression, a non-parametric technique used to compute the conditional expectation of a random variable (Hastie et al., 2009), was used to visualize how choice RTs were related to chosen softmax probabilities (Figures 5A and 5B). We specified a Gaussian kernel (20 ms), and the estimated function was sampled in evenly spaced 20 ms bins from 0 to 500 ms. We quantified the speed-accuracy trade-off by estimating correlation coefficients for each block and averaged them across schedules for each session.

Aborted Trials due to Fixation or Choice Errors

Fixation errors were defined as a failure to acquire fixation up to 1 s after onset of the fixation cue or a failure to hold fixation up until presentation of the two choice options. Choice errors were defined as a failure to saccade to one of the two choice options up to 1 s after onset of the choice options. Each error type was quantified as a percentage of the total number of trials completed per session.

Classical Statistics

Each dependent variable was entered into mixed effects ANOVAs implemented in MATLAB. When appropriate, schedule, learning phase, feedback type, and monkey were specified as fixed effects and session as a random fac-

tor nested under monkey. Post hoc analyses of significant main effects used Fisher's least significant difference test to correct for multiple comparisons (Levin et al., 1994). Post hoc tests of significant interactions consisted of computing univariate ANOVAs for component effects corrected for multiple comparisons.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.neuron.2016.09.025>.

AUTHOR CONTRIBUTIONS

Conceptualization, V.D.C., O.D.M., E.A.M., and B.B.A.; Methodology, V.D.C., O.D.M., E.A.M., and B.B.A.; Software, V.D.C. and B.B.A.; Formal Analysis, V.D.C. and B.B.A.; Investigation, V.D.C., O.D.M., and D.R.L.; Writing—Initial Draft, V.D.C. and B.B.A.; Writing—Review and Editing, V.D.C., O.D.M., E.A.M., and B.B.A.; Resources, E.A.M. and B.B.A.

ACKNOWLEDGMENTS

This work was supported by the Intramural Research Program of the National Institute of Mental Health. We thank Laura Kennerly, Laura Kakalios, Kati Rothenhoefer, Raquel Vicario-Feliciano, and Pam Noble for assistance with data collection; Stephanie Huard for assistance in constructing the lesion density maps; and Andrew Mitz for technical assistance.

Received: February 25, 2016

Revised: July 28, 2016

Accepted: August 31, 2016

Published: October 6, 2016

REFERENCES

- Asaad, W.F., and Eskandar, E.N. (2008). A flexible software tool for temporally precise behavioral control in Matlab. *J. Neurosci. Methods* 174, 245–258.
- Atlas, L.Y., Doll, B.B., Li, J., Daw, N.D., and Phelps, E.A. (2016). Instructed knowledge shapes feedback-driven aversive learning in striatum and orbitofrontal cortex, but not the amygdala. *eLife* 5, e15192.
- Baxter, M.G., and Murray, E.A. (2002). The amygdala and reward. *Nat. Rev. Neurosci.* 3, 563–573.
- Cardinal, R.N., Pennicott, D.R., Sugathapala, C.L., Robbins, T.W., and Everitt, B.J. (2001). Impulsive choice induced in rats by lesions of the nucleus accumbens core. *Science* 292, 2499–2501.
- Cardinal, R.N., Parkinson, J.A., Hall, J., and Everitt, B.J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci. Biobehav. Rev.* 26, 321–352.
- Chau, B.K., Sallet, J., Papageorgiou, G.K., Noonan, M.P., Bell, A.H., Walton, M.E., and Rushworth, M.F. (2015). Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques. *Neuron* 87, 1106–1118.
- Collins, A.G., and Frank, M.J. (2014). Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev.* 121, 337–366.
- Corbit, L.H., and Balleine, B.W. (2005). Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of Pavlovian-instrumental transfer. *J. Neurosci.* 25, 962–970.
- Corbit, L.H., Janak, P.H., and Balleine, B.W. (2007). General and outcome-specific forms of Pavlovian-instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral tegmental area. *Eur. J. Neurosci.* 26, 3141–3149.
- Costa, V.D., Tran, V.L., Turchi, J., and Averbeck, B.B. (2014). Dopamine modulates novelty-seeking behavior during decision making. *Behav. Neurosci.* 128, 556–566.

- Costa, V.D., Tran, V.L., Turchi, J., and Averbeck, B.B. (2015). Reversal learning and dopamine: a Bayesian perspective. *J. Neurosci.* *35*, 2407–2416.
- Courville, A.C., Daw, N.D., and Touretzky, D.S. (2006). Bayesian theories of conditioning in a changing world. *Trends Cogn. Sci.* *10*, 294–300.
- Dalton, G.L., Phillips, A.G., and Floresco, S.B. (2014). Preferential involvement by nucleus accumbens shell in mediating probabilistic learning and reversal shifts. *J. Neurosci.* *34*, 4618–4626.
- Day, J.J., Roitman, M.F., Wightman, R.M., and Carelli, R.M. (2007). Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* *10*, 1020–1028.
- Floresco, S.B. (2015). The nucleus accumbens: an interface between cognition, emotion, and action. *Annu. Rev. Psychol.* *66*, 25–52.
- Friedman, D.P., Aggleton, J.P., and Saunders, R.C. (2002). Comparison of hippocampal, amygdala, and perirhinal projections to the nucleus accumbens: combined anterograde and retrograde tracing study in the macaque brain. *J. Comp. Neurol.* *450*, 345–365.
- Grabenhorst, F., Hernádi, I., and Schultz, W. (2012). Prediction of economic choice by primate amygdala neurons. *Proc. Natl. Acad. Sci. USA* *109*, 18950–18955.
- Gurney, K.N., Humphries, M.D., and Redgrave, P. (2015). A new framework for cortico-striatal plasticity: behavioural theory meets in vitro data at the reinforcement-action interface. *PLoS Biol.* *13*, e1002034.
- Haber, S.N., and Behrens, T.E. (2014). The neural network underlying incentive-based learning: implications for interpreting circuit disruptions in psychiatric disorders. *Neuron* *83*, 1019–1039.
- Haber, S.N., and Fudge, J.L. (1997). The primate substantia nigra and VTA: integrative circuitry and function. *Crit. Rev. Neurobiol.* *11*, 323–342.
- Hampton, A.N., Adolphs, R., Tyszka, M.J., and O'Doherty, J.P. (2007). Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. *Neuron* *55*, 545–555.
- Hart, A.S., Rutledge, R.B., Glimcher, P.W., and Phillips, P.E. (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J. Neurosci.* *34*, 698–704.
- Hastie, T., Tibshirani, R., and Friedman, J.H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition (Springer).
- Hernádi, I., Grabenhorst, F., and Schultz, W. (2015). Planning activity for internally generated reward goals in monkey amygdala neurons. *Nat. Neurosci.* *18*, 461–469.
- Holland, P.C., and Gallagher, M. (2003). Double dissociation of the effects of lesions of basolateral and central amygdala on conditioned stimulus-potentiated feeding and Pavlovian-instrumental transfer. *Eur. J. Neurosci.* *17*, 1680–1694.
- Holland, P.C., and Schiffino, F.L. (2016). Mini-review: prediction errors, attention, and associative learning. *Neurobiol. Learn. Mem.* *131*, 207–215.
- Houk, J.C., Adamas, J.L., and Barto, A.G. (1995). A model of how the basal ganglia generates and uses neural signals that predict reinforcement. In *Models of Information Processing in the Basal Ganglia*, J.C. Houk, J.L. Davis, and D.G. Beiser, eds. (MIT Press), pp. 249–274.
- Izquierdo, A., and Murray, E.A. (2007). Selective bilateral amygdala lesions in rhesus monkeys fail to disrupt object reversal learning. *J. Neurosci.* *27*, 1054–1062.
- Izquierdo, A., Darling, C., Manos, N., Pozos, H., Kim, C., Ostrander, S., Cazares, V., Stepp, H., and Rudebeck, P.H. (2013). Basolateral amygdala lesions facilitate reward choices after negative feedback in rats. *J. Neurosci.* *33*, 4105–4109.
- Janak, P.H., and Tye, K.M. (2015). From circuits to behaviour in the amygdala. *Nature* *517*, 284–292.
- Jang, A.I., Costa, V.D., Rudebeck, P.H., Chudasama, Y., Murray, E.A., and Averbeck, B.B. (2015). The role of frontal cortical and medial-temporal lobe brain areas in learning a Bayesian prior belief on reversals. *J. Neurosci.* *35*, 11751–11760.
- Jones, B., and Mishkin, M. (1972). Limbic lesions and the problem of stimulus-reinforcement associations. *Exp. Neurol.* *36*, 362–377.
- Lee, E., Seo, M., Dal Monte, O., and Averbeck, B.B. (2015). Injection of a dopamine type 2 receptor antagonist into the dorsal striatum disrupts choices driven by previous outcomes, but not perceptual inference. *J. Neurosci.* *35*, 6298–6306.
- Levin, J.R., Serlin, R.C., and Seaman, M.A. (1994). A controlled, powerful multiple-comparison strategy for several situations. *Psychol. Bull.* *115*, 153–159.
- Li, J., Schiller, D., Schoenbaum, G., Phelps, E.A., and Daw, N.D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nat. Neurosci.* *14*, 1250–1252.
- Mitz, A.R. (2005). A liquid-delivery device that provides precise reward control for neurophysiological and behavioral experiments. *J. Neurosci. Methods* *148*, 19–25.
- Moscarello, J.M., and LeDoux, J. (2014). Diverse effects of conditioned threat stimuli on behavior. *Cold Spring Harb. Symp. Quant. Biol.* *79*, 11–19.
- Namburi, P., Beyeler, A., Yorozu, S., Calhoun, G.G., Halbert, S.A., Wichmann, R., Holden, S.S., Mertens, K.L., Anahtar, M., Felix-Ortiz, A.C., et al. (2015). A circuit mechanism for differentiating positive and negative associations. *Nature* *520*, 675–678.
- O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron* *38*, 329–337.
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* *304*, 452–454.
- Parker, N.F., Cameron, C.M., Taliaferro, J.P., Lee, J., Choi, J.Y., Davidson, T.J., Daw, N.D., and Witten, I.B. (2016). Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* *19*, 845–854.
- Pearce, J.M., and Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* *87*, 532–552.
- Phelps, E.A., LeDoux, J.E., and LeDoux, J.E. (2014). Emotion and decision making: multiple modulatory neural circuits. *Annu. Rev. Neurosci.* *37*, 263–287.
- Pothuizen, H.H., Jongen-Rêlo, A.L., Feldon, J., and Yee, B.K. (2005). Double dissociation of the effects of selective nucleus accumbens core and shell lesions on impulsive-choice behaviour and salience learning in rats. *Eur. J. Neurosci.* *22*, 2605–2616.
- Prévost, C., McNamee, D., Jessup, R.K., Bossaerts, P., and O'Doherty, J.P. (2013). Evidence for model-based computations in the human amygdala during Pavlovian conditioning. *PLoS Comput. Biol.* *9*, e1002918.
- Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory*, A.H. Black and W.F. Pokasy, eds. (Appleton-Century-Crofts), pp. 64–99.
- Robinson, M.J., Warlow, S.M., and Berridge, K.C. (2014). Optogenetic excitation of central amygdala amplifies and narrows incentive motivation to pursue one reward above another. *J. Neurosci.* *34*, 16567–16580.
- Roesch, M.R., Esber, G.R., Li, J., Daw, N.D., and Schoenbaum, G. (2012). Surprise! Neural correlates of Pearce-Hall and Rescorla-Wagner coexist within the brain. *Eur. J. Neurosci.* *35*, 1190–1200.
- Rudebeck, P.H., and Murray, E.A. (2008). Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *J. Neurosci.* *28*, 8338–8343.
- Rutledge, R.B., Dean, M., Caplin, A., and Glimcher, P.W. (2010). Testing the reward prediction error hypothesis with an axiomatic model. *J. Neurosci.* *30*, 13525–13536.
- Rygula, R., Clarke, H.F., Cardinal, R.N., Cockcroft, G.J., Xia, J., Dalley, J.W., Robbins, T.W., and Roberts, A.C. (2015). Role of central serotonin in

- anticipation of rewarding and punishing outcomes: effects of selective amygdala or orbitofrontal 5-HT depletion. *Cereb. Cortex* 25, 3064–3076.
- Saez, A., Rigotti, M., Ostojic, S., Fusi, S., and Salzman, C.D. (2015). Abstract context representations in primate amygdala and prefrontal cortex. *Neuron* 87, 869–881.
- Salzman, C.D., and Fusi, S. (2010). Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annu. Rev. Neurosci.* 33, 173–202.
- Schultz, W. (2015). Neuronal reward and decision signals: from theories to data. *Physiol. Rev.* 95, 853–951.
- Schwartzbaum, J.S., and Poulos, D.A. (1965). Discrimination behavior after amygdectomy in monkeys: learning set and discrimination reversals. *J. Comp. Physiol. Psychol.* 60, 320–328.
- Seymour, B., and Dolan, R. (2008). Emotion, decision making, and the amygdala. *Neuron* 58, 662–671.
- Stalnaker, T.A., Franz, T.M., Singh, T., and Schoenbaum, G. (2007). Basolateral amygdala lesions abolish orbitofrontal-dependent reversal impairments. *Neuron* 54, 51–58.
- Stern, C.E., and Passingham, R.E. (1996). The nucleus accumbens in monkeys (*Macaca fascicularis*): II. emotion and motivation. *Behav. Brain Res.* 75, 179–193.
- Stuber, G.D., Sparta, D.R., Stamatakis, A.M., van Leeuwen, W.A., Hardjoprajitno, J.E., Cho, S., Tye, K.M., Kempadoo, K.A., Zhang, F., Deisseroth, K., and Bonci, A. (2011). Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. *Nature* 475, 377–380.
- Takahashi, Y.K., Langdon, A.J., Niv, Y., and Schoenbaum, G. (2016). Temporal specificity of reward prediction errors signaled by putative dopamine neurons in rat VTA depends on ventral striatum. *Neuron* 91, 182–193.
- Zhang, S., Mano, H., Ganesh, G., Robbins, T., and Seymour, B. (2016). Dissociable learning processes underlie human pain conditioning. *Curr. Biol.* 26, 52–58.