# Ventral striatum's role in learning from gains and losses

Craig A. Taswell[a], Vincent D. Costa[a], Elisabeth A. Murray[a], and Bruno B. Averbeck[a,1]

[a]Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892-4415

Edited by L. F. Abbott, Columbia University, New York, NY, and approved November 12, 2018 (received for review June 11, 2018)

**Adaptive behavior requires animals to learn from experience. Ideally, learning should both promote choices that lead to rewards and reduce choices that lead to losses. Because the ventral striatum (VS) contains neurons that respond to aversive stimuli and aversive stimuli can drive dopamine release in the VS, it is possible that the VS contributes to learning about aversive outcomes, including losses. However, other work suggests that the VS may play a specific role in learning to choose among rewards, with other systems mediating learning from aversive outcomes. To examine the role of the VS in learning from gains and losses, we compared the performance of macaque monkeys with VS lesions and unoperated controls on a reinforcement learning task. In the task, the monkeys gained or lost tokens, which were periodically cashed out for juice, as outcomes for choices. They learned over trials to choose cues associated with gains, and not choose cues associated with losses. We found that monkeys with VS lesions had a deficit in learning to choose between cues that differed in reward magnitude. By contrast, monkeys with VS lesions performed as well as controls when choices involved a potential loss. We also fit reinforcement learning models to the behavior and compared learning rates between groups. Relative to controls, the monkeys with VS lesions had reduced learning rates for gain cues. Therefore, in this task, the VS plays a specific role in learning to choose between rewarding options.**

ventral striatum | reinforcement learning | appetitive | aversive | neuroeconomics

Adaptive behavior requires that organisms choose wisely to gain rewards and avoid punishment. Reinforcement learning (RL) refers to the behavioral process of learning about the value of choices, based on choice outcomes. From an algorithmic point of view, rewards and punishments exist on opposite sides of a single value axis. Simple distinctions between rewards and punishments, however, and their theoretical expression on a single value axis, hide the considerable complexities that underlie appetitive and aversive RL. Most notably, both rewards and punishments come in many forms. Food, sex, and ascending the social hierarchy are rewarding. Correspondingly, loss of cached food, pain, and social defeat are punishing (1). Whether threat, pain, and loss of accumulated reward drive learning via the same neural systems, at any level, is unclear. Furthermore, even when gains and losses are expressed with money, which has objective value, they can have differential subjective effects on behavior (2–4).

Studies of RL often use paradigms in which participants learn to choose options on the basis of reward frequency or reward magnitude (5–7). These studies have shown that the striatum, and the dopamine input to the striatum, underlies learning to select rewarding options. Theoretical models of RL extend directly to learning from losses, and therefore striatal-mediated learning may generalize to these conditions (8). This hypothesis is supported by work that has shown that dopamine neurons, which provide reward prediction error (RPE) signals to the striatum, increase their firing rates when rewards are unexpectedly delivered and decrease their firing rate when rewards are unexpectedly omitted (9, 10). However, some studies have explicitly examined learning from gains and losses (as opposed to reward omission) and found that they are mediated by partially overlapping, but partially distinct, systems that cross cortical and subcortical circuits. For example, single-neuron studies in macaques have shown that the dorsolateral prefrontal cortex, as well as the anterior cingulate cortex, encode both losses and gains in a competitive game in which conditioned reinforcers could be gained and lost (11). In other work, the medial orbitofrontal cortex was found to encode gains and avoidance of losses, both of which have positive value (12). This study also found that appetitive RPEs in reward trials (i.e., increases with unexpected rewards) correlated with the extent of activation in the ventral striatum (VS), whereas RPEs in aversive trials (i.e., increases with unexpected punishments) correlated with activation in the insula, consistent with other work (5). In addition to the work in macaques and humans, work in rodents, which has used various paradigms including conditioned place aversion and Pavlovian threat of shock, has shown that basolateral amygdala circuits through the VS encode reward-mediated approach behavior, whereas circuits through the central nucleus of the amygdala encode avoidance (13–15). Related experiments focusing on circuitry have found that dopamine inputs to the infralimbic (IL)/prelimbic (PL) regions of medial prefrontal cortex also encode avoidance behavior (16). Thus, there is evidence that both overlapping and distinct systems underlie learning from rewards and punishments, using some paradigms.

To examine the role of the VS in learning from both gains and losses, we adapted a previously used token reward system (11) to two-armed bandit RL tasks. In the tasks, rhesus monkeys made choices among options and received tokens for their choices. The tokens were represented by circles on the bottom of the screen, and the animals periodically received juice in exchange for accumulated tokens. The use of tokens, which are secondary

## Significance

**A broad set of neural circuits, including the amygdala and frontal-striatal systems, has been implicated in mediating learning from gains and losses. The ventral striatum (VS) has been implicated in several aspects of this process. Here, we examined the specific contribution of the VS to learning from gains vs. losses. We found that the VS plays a role in learning to choose between two options that vary in gains but plays no role in learning to choose between two options when one or both is associated with a loss. Computational modeling supported this by showing that animals with VS lesions specifically learned slowly when choosing between gains but not losses.**

Author contributions: C.A.T., V.D.C., E.A.M., and B.B.A. designed research; C.A.T. performed research; C.A.T. analyzed data; and C.A.T., V.D.C., E.A.M., and B.B.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

[1]To whom correspondence should be addressed. Email: bruno.averbeck@nih.gov.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1809833115/-/DCSupplemental.

Published online December 13, 2018.

reinforcers, allowed us to study the effects of gains and losses on choices using one and the same unit of value. We ran four variants of the task to address specific questions. Three variants used deterministic outcomes, and one used stochastic. We compared the behavioral performance of three monkeys with lesions of the VS and four unoperated controls.

## Results

The monkeys were run on a series of four tasks. In each task, trials involved a forced choice between two images. Selection of a particular image led to increases or decreases in accumulated tokens (Fig. 1*A*). The outcome of each trial following a choice was realized on the monitor screen as a change in the number of tokens the animal had accumulated. Every four to seven trials, with the interval chosen randomly, we cashed out the accumulated tokens. During cash-out, the monkeys received one drop of juice for each token. The animals had to learn over trials to select the image from the pair that maximized their gains and minimized their losses. When the monkeys had no tokens and they chose a loss cue, there was no change in the tokens. The animals could also not incur negative token counts.

The tasks were run in a fixed sequence (Fig. 1*A*). Each task evaluated the monkeys' choices during learning and performance on novel or familiar stimulus–outcome associations. In the novel blocks, the monkeys learned stimulus–reward associations for a novel set of images. In the familiar blocks (*SI Appendix*, Figs. S1–S4), the monkeys chose between stimuli they had repeatedly sampled over the course of prior experimental sessions. The stimulus–outcome associations of these familiar choice options were fixed for the duration of the experiment. Novel and familiar blocks were randomly interleaved each day. The novel blocks allowed us to examine the rate at which cue–reward associations were learned, whereas the familiar blocks allowed us to examine asymptotic performance with overlearned cue–reward associations.

**Deterministic RL of Stimulus–Outcome Associations.** We first evaluated the ability of the monkeys with or without lesions of the VS to learn deterministic stimulus–outcome associations. In this task, at the beginning of each novel block, the monkeys encountered four images they had not seen before. Each image was associated with a fixed, deterministic gain or loss of tokens (+2, +1, −1, or −2 tokens). Two of the images were presented as choice options on each trial. This resulted in six unique pairs of images, which we refer to as conditions. In each block, conditions were randomly interleaved over intervals of 12 trials until each condition occurred 18 times in novel (Fig. 2) blocks and 6 times in familiar (*SI Appendix*, Fig. S1) blocks.

In novel blocks, the monkeys learned the stimulus–outcome associations efficiently. With experience, they were able to choose the better option of the pair on a high proportion of trials (Fig. 2). There were differences in performance across conditions [Condition; $F_{(5,20)} = 140$, $P < 0.001$] and differences in performance across trials in the different conditions [Condition × Trial; $F_{(85,38)} = 5.7$, $P < 0.001$]. The monkeys performed best in the conditions in which there was a loss paired with the largest reward. For example, they most often picked the best cue when choosing between the +2 and −1 and +2 and −2 conditions. This effect was driven largely by the frequency with which they experienced the outcomes associated with each cue and the differences in the values of the cues. The animals most frequently picked the +2 cue across all conditions, and therefore most frequently received feedback on its value, and the value of this cue would also asymptote at +2.

In task 1, there were no differences between groups [Group; $F_{(1,9)} = 0.1$, $P = 0.7611$] and no differences between the groups across conditions [Group × Condition; $F_{(5,22)} = 0.5$, $P = 0.7801$].

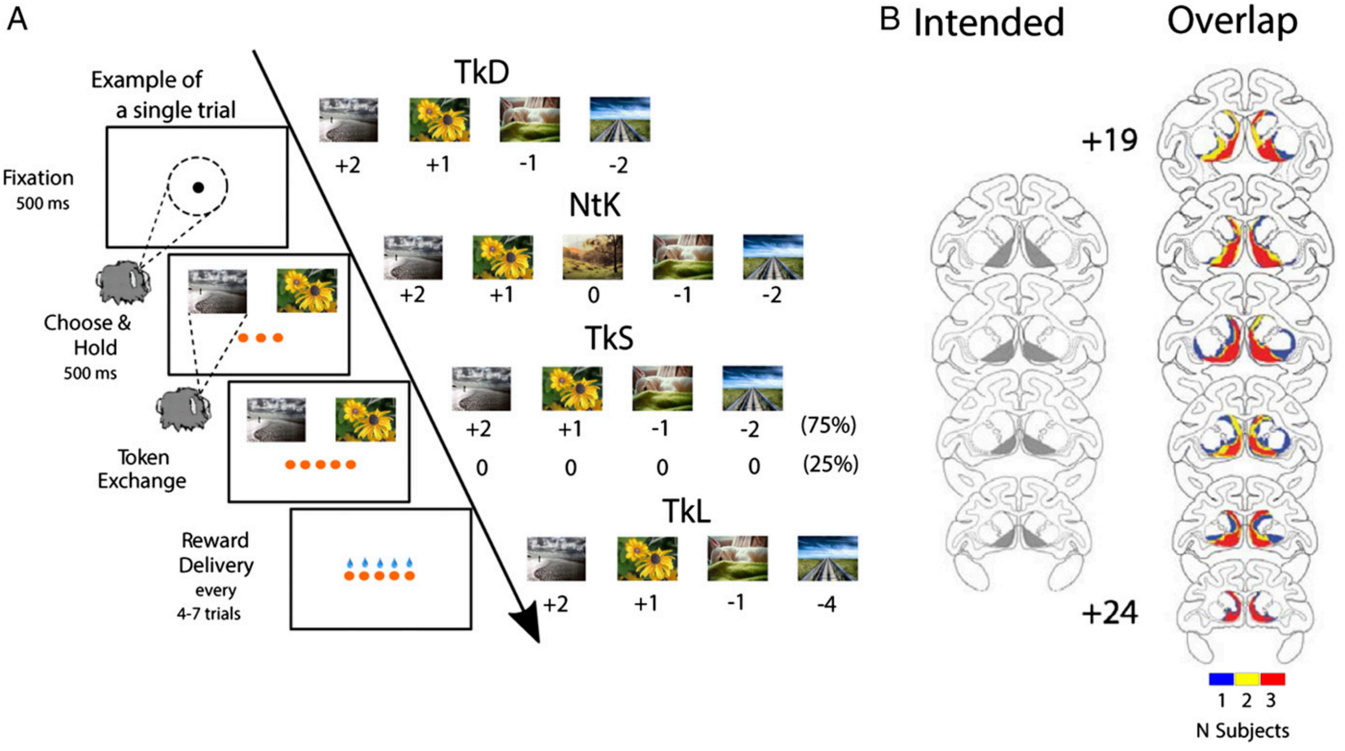

**Fig. 1.** Tasks used and lesion map. (*A*) Diagram of the trial structure used in all tasks. The specific reward magnitudes used in each task are shown. TkD, Task 1 in which deterministic reward magnitudes were +2, +1, −1, and −2. NtK, Task 2 in which we include a null token giving deterministic reward magnitudes of +2, +1, 0, −1, and −2. TkS, Task 3 in which feedback was stochastic with magnitudes of +2, +1, −1, and −2. TkL, Task 4 in which deterministic reward magnitudes, including a large loss, were +2, +1, −1, and −4. (*B*) Lesion map of the three animals in the lesion group. Colors indicate number of animals that had lesion of corresponding extent.

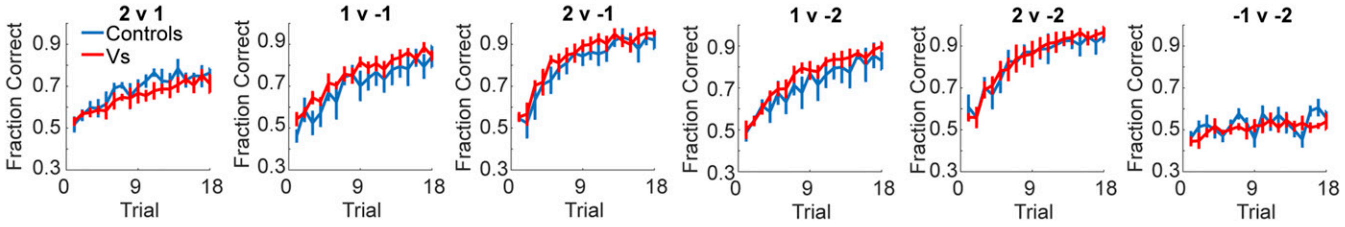

**Fig. 2.** Deterministic RL of stimulus–outcome associations. Task 1 choice behavior. Error bars are ±SEM with $n$ = number of animals. Plots show the fraction of times monkeys chose the higher value option averaged across novel blocks for each group. Numbers at the top of each plot indicate the condition, which corresponds to the cues that were shown in those trials.

The monkeys did not perform well in the −1 vs. −2 condition, although across the groups there was a significant positive correlation between choice accuracy and trial, which indicates learning [$t_{(6)} = 9.1$, $P < 0.001$]. When we examined the groups individually, we found that both groups learned to choose the smaller loss more often with experience [Control: $t_{(3)} = 6.9$, $P = 0.006$; VS: $t_{(2)} = 13.4$, $P = 0.005$].

**Deterministic RL Augmented by a Null Cue.** In task 2, we used five cues in each block with cue–outcome mappings of +2, +1, 0, −1, and −2. This resulted in 10 pairs of cues and therefore 10 conditions. In task 2, both the novel (Fig. 3) and familiar (*SI Appendix*, Fig. S2) blocks were composed of 120 trials, 12 per condition. Therefore, in the novel blocks, the monkeys saw each pair of cues 12 times. Inclusion of the null cue allowed us to test two specific hypotheses. First, does the absolute difference between the value of the cues drive performance independent of the reward value associated with the cues? Second, can animals learn to select the null cue when it is paired with a loss cue?

In the novel blocks (Fig. 3), there was again a difference in performance across conditions [Condition; $F_{(9,19)} = 54.7$, $P < 0.001$] and also a difference in performance across trials in the different conditions [Condition × Trial; $F_{(99,398)} = 8.2$, $P < 0.001$]. There were no differences between groups [Group; $F_{(1,9)} = 0.1$, $P = 0.778$] and no differences by condition [Group × Condition; $F_{(9,14)} = 0.6$, $P = 0.793$]. There was also no difference between groups when we examined only the 2 vs. 1 condition [Group; $F_{(1,5)} = 3.6$, $P = 0.117$].

Similar to task 1, when we grouped all of the animals together, there was a significant correlation between trial and performance when the animals had to choose between the two loss cues [$t_{(6)} = 3.3$, $P = 0.016$]. However, when we separated the groups, neither group reached significance alone [Controls: $t_{(3)} = 2.3$, $P = 0.103$; VS: $t_{(2)} = 3.1$, $P = 0.092$]. There was also significant learning when the animals had to choose between the 0 and −1 cue across groups, but not in either group individually [$t_{(6)} = 3.9$, $P = 0.007$; Controls: $t_{(3)} = 2.9$, $P = 0.062$; VS: $t_{(2)} = 2.1$, $P = 0.164$]. When the animals had to choose between the 0 and −2 cues there was learning across groups [$t_{(6)} = 5.7$, $P = 0.001$]. However, when we examined the groups separately, we found that only the controls performed significantly better than chance [Controls: $t_{(3)} = 3.5$, $P = 0.037$; VS: $t_{(2)} = 4.1$, $P = 0.053$].

**Stochastic RL.** In task 3, we introduced four cues with cue–outcome associations of +2, +1, −1, and −2. However, the cue–outcome associations were stochastic. Therefore, when the animals chose one of the options, they received the outcome associated with that option in 75% of the trials, and no outcome (i.e., no change in tokens) in 25% of the trials. We introduced this task because we have previously seen that monkeys with VS lesions learn poorly under stochastic schedules (7), and the VS may be more important for slow learning, which is more affected by trial-by-trial stochasticity (17). Both novel (Fig. 4) and familiar (*SI Appendix*, Fig. S3) blocks were 108 trials.

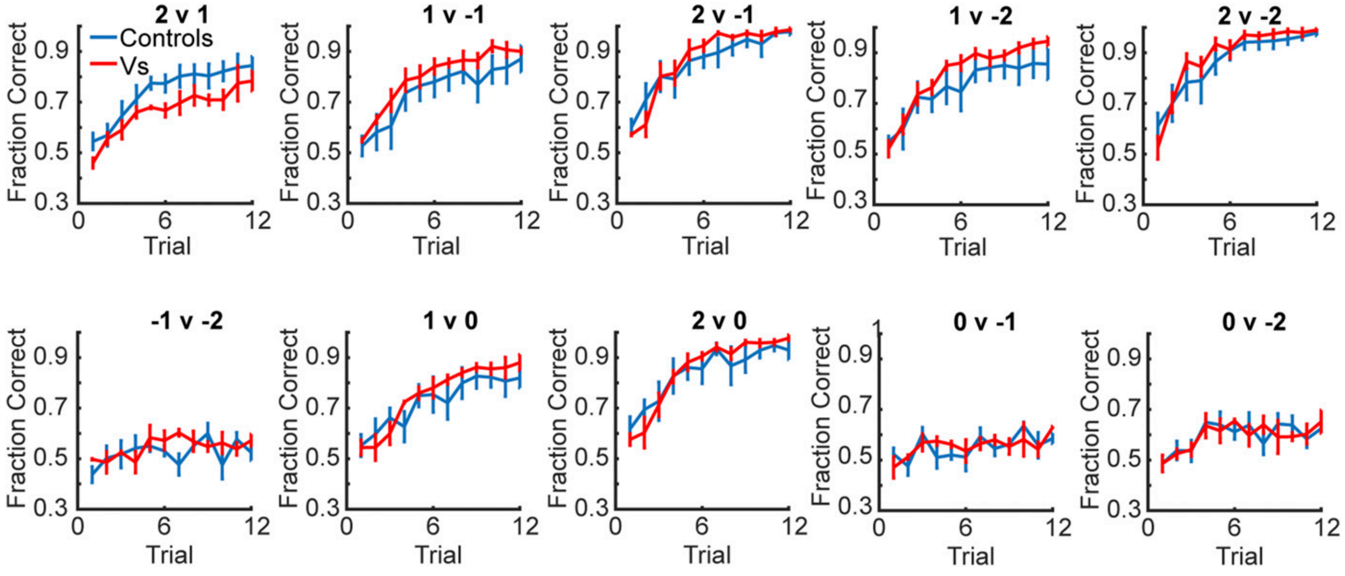

**Fig. 3.** Deterministic RL augmented by a null cue. Task 2 choice behavior. Error bars are ±SEM with $n$ = number of animals. Plots show the fraction of times monkeys chose the higher value option averaged across novel blocks for each group. Numbers at the top of each plot indicate the condition, which corresponds to the cues that were shown in those trials.

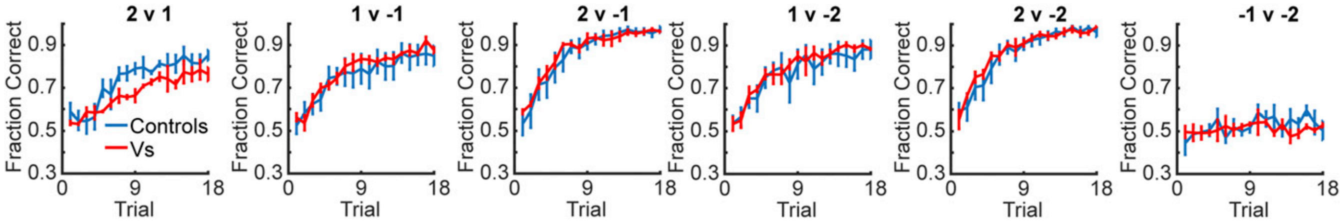

**Fig. 4.** Stochastic RL. Task 3 choice behavior. Error bars are ±SEM with $n$ = number of animals. Plots shows the fraction of times monkeys chose the higher value option averaged across novel blocks for each group. Numbers at the top of each plot indicate the condition, which corresponds to the cues that were shown in those trials.

Performance was consistent with the previous tasks (Fig. 4). In the novel blocks, there was a difference in performance across conditions [Condition; $F_{(5,38)} = 315.0$, $P < 0.001$] and learning also differed across trials in the different conditions [Condition × Trial; $F_{(85,135)} = 7.4$, $P < 0.001$]. In addition to these effects, and unlike the case for the tasks with deterministic outcomes, there was an overall effect of group [Group; $F_{(1,70)} = 15.2$, $P < 0.001$]. When we examined differences between groups in each condition, we found that the 2 vs. 1 condition approached significance, but this did not survive correction for six comparisons [Group; $F_{(1,5)} = 6.73$, $P = 0.049$]. In this task, monkeys showed learning when choosing between the −1 and −2 cues [All animals: $t_{(6)} = 2.7$, $P = 0.037$]. When we looked at the groups separately we found that only controls learned to choose the smaller loss, [Control: $t_{(3)} = 3.3$, $P = 0.045$; VS: $t_{(2)} = 0.7$, $P = 0.523$].

**Deterministic RL with a Large Loss.** In task 4, we introduced four cues with cue–outcome associations of +2, +1, −1, and −4. We added the larger loss cue to see whether animals would learn to pick the smaller loss cue more effectively, when the difference between the two loss cues was larger. We also gave the monkeys an endowment of four tokens on the first trial after each cash-out. We did this to ensure that the animals had sufficient tokens to experience the large loss and to maintain motivation. Novel (Fig. 5) and familiar (*SI Appendix*, Fig. S4) blocks were both composed of 108 trials, with 18 trials per condition.

Performance in novel blocks again showed a difference in performance across conditions [Fig. 5; Condition; $F_{(5,43)} = 231.0$, $P < 0.001$] and a difference in performance across trials in different conditions [Condition × Trial; $F_{(85,313)} = 4.6$, $P < 0.001$]. There was also a main effect of group [Group; $F_{(1,31)} = 30.7$, $P < 0.001$]. None of the group effects in individual conditions survived multiple-comparisons corrections. The monkeys were able to learn to choose the smaller of the two losses [$t_{(5)} = 10.8$, $P < 0.001$]. In addition, when we examined each group separately, we found that both groups were able to learn to choose the smaller of the two losses [Control: $t_{(2)} = 5.2$, $P = 0.035$; VS: $t_{(2)} = 14.6$, $P = 0.004$].

**RL Models.** Next, we fit RL models to the data in the novel blocks in all tasks. We fit two models that varied in the number of free parameters used to model the choice behavior. One model used one parameter for positive cues and one parameter for negative cues (VALENCE model), thus allowing learning rates to vary for positive vs. negative outcomes. The second model used one parameter for each cue (CUE model), allowing for different learning rates for each outcome. The VALENCE model fit behavior well in most conditions, particularly in the control group (*SI Appendix*, Fig. S5). However, the VALENCE model overpredicted performance in the 2 vs. 1 condition. This could be seen, for example, in task 3 which had stochastic outcomes, and in which there was a large discrepancy between behavior and model predictions in the 2 vs. 1 condition (*SI Appendix*, Fig. S5 *B* and *D*). This effect was strongest for the VS animals (*SI Appendix*, Fig. S5*D*) but could also be seen for the control animals (*SI Appendix*, Fig. S5*B*). In the other conditions, however, the VALENCE model fit well. The CUE model did not show biases in any condition in either group (*SI Appendix*, Fig. S5 *A* and *C*).

Averaged across tasks, the VALENCE model overpredicted performance in the 2 vs. 1 condition in both groups [Fig. 6 *A* and *B*; VALENCE model vs. Behavior, $F_{(1,5)} = 167.9$, $P < 0.001$]. The VALENCE model overpredicted behavior more for the VS group than the controls [Group × VALENCE model vs. Behavior, $F_{(1,5)} = 19.4$, $P = 0.007$]. The CUE model, on the other hand, did not differ from behavior in the 2 vs. 1 condition, across tasks [CUE model vs. behavior, $F_{(1,5)} = 0.2$, $P = 0.582$], although the fit did differ by group [CUE model vs. Behavior × Group, $F_{(1,5)} = 14.2$, $P = 0.016$], with a closer fit between behavior and model in the VS group. We also used the Bayesian information criterion to assess which model fit best in each session for each animal and task. In all animals in both groups, averaged across tasks, the CUE model was more frequently the best model than the VALENCE model. Across groups, there was a preference for the CUE model over the VALENCE model [$t_{(6)} = 2.84$, $P < 0.030$; 57% of sessions best fit by CUE model]. This preference was not significant individually in the control [$t_{(3)} = 1.6$, $P = 0.210$; 57%] or VS animals [$t_{(2)} = 3.03$, $P = 0.094$; 57%].

Next, we compared the learning rate parameters between groups from the CUE model (Fig. 6*C*). We found that the parameters

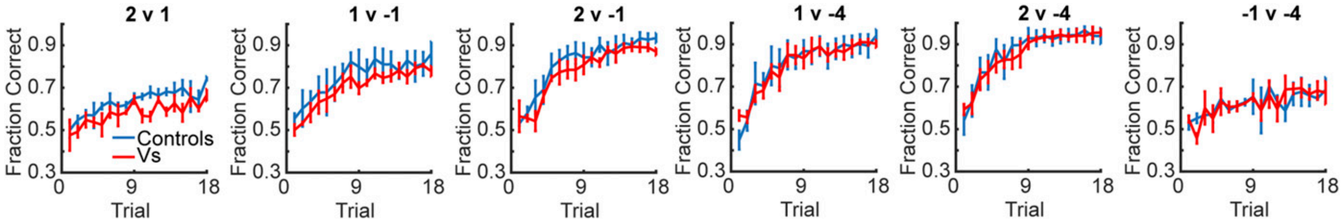

**Fig. 5.** Deterministic RL with a large loss. Task 4 choice behavior. Error bars are ±SEM with $n$ = number of animals. Plots shows the fraction of times monkeys chose the higher value option averaged across novel blocks for each group. Numbers at the top of each plot indicate the condition, which corresponds to the cues that were shown in those trials.

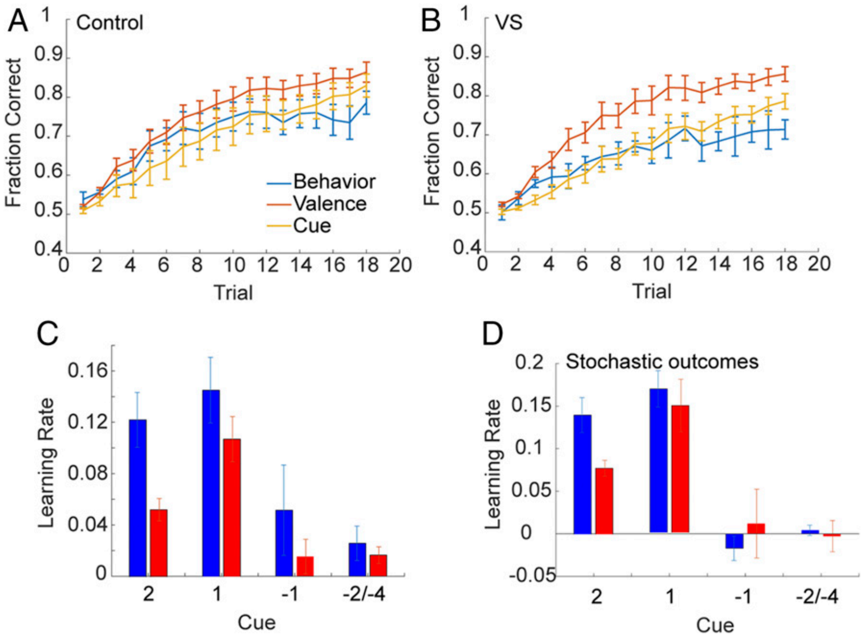

**Fig. 6.** Best-fitting RL models. (*A*) Overlay of behavior and predicted performance, averaged across experiments, for the 2 vs. 1 condition for the control animals. (*B*) Same as *D* for the VS animals. (*C*) Learning-rate parameters averaged across tasks, extracted from the RL CUE model. Error bars are SEM across monkeys in each group. (*D*) Average learning rates for the CUE model, cue parameters, in task 3, with stochastic feedback.

varied across cues [Fig. 6*C*; $F_{(3,15)} = 45.6$, $P < 0.001$] and tasks [$F_{(3,14)} = 4.0$, $P = 0.030$]. Learning rates were lower in the VS group [Group; $F_{(1,5)} = 9.75$, $P = 0.026$]. The groups did not differ across cues [Group × Cue, $F_{(3,15)} = 2.9$, $P = 0.070$] or tasks [Group × Task, $F_{(3,14)} = 1.8$, $P = 0.193$]. We also examined effects within the gain cues and within the loss cues, separately. There were group differences within the gain cues [Group, $F_{(1,5)} = 10.35$, $P = 0.024$], and these effects differed marginally across the two cues [Group × Cue, $F_{(1,5)} = 9.3$, $P = 0.029$]. There were no group differences in the loss cues [Group, $F_{(1,5)} = 2.1$, $P = 0.206$], but the groups did differ by cue [Group × Cue, $F_{(1,5)} = 9.2$, $P = 0.029$]. When we examined group differences in the individual tasks, we found a difference in groups across cues for task 3 with stochastic outcomes [Fig. 6*D*; Group × Cue, $F_{(3,15)} = 5.0$, $P = 0.014$] but no other group differences in the other tasks ($P > 0.05$). We also examined the inverse temperatures from the model fits. These differed by experiment [$F_{(3,14)} = 4.4$, $P = 0.021$]. However, there were no differences across groups [$F_{(1,5)} = 0.0$, $P = 0.887$] or by group across experiments [$F_{(3,14)} = 1.5$, $P = 0.267$]. None of the choice autocorrelation parameters (*Methods*) varied by group ($P > 0.05$).

Therefore, across tasks, the animals with VS lesions consistently had deficits in learning to discriminate between the two gain cues in the 2 vs. 1 condition, and this manifested as a significantly larger deficit relative to the VALENCE model predictions in the 2 vs. 1 condition and a significant reduction in learning rates relative to controls specifically for the gain cues.

**Aborted Trials and Reaction Times.** We also examined aborted trials and reaction times across tasks (Fig. 7). It was sometimes the case that, when the two images were presented, the animals broke fixation and did not select either image in the novel (Fig. 7*A*) and familiar conditions (Fig. 7*B*). The monkeys broke fixation more frequently when they had to choose between the −1 and −2 cues. This was true even though following an error we repeated the same condition. This differed by condition [Novel: $F_{(24,116)} = 17.9$, $P < 0.001$; Familiar; $F_{(24,116)} = 9.92$, $P < 0.001$] and task [Novel: $F_{(3,23)} = 4{,}653.0$, $P < 0.001$; $F_{(3,16)} = 155.9$, $P < 0.001$]. There were, however, no interactions of task with other variables, and there was no effect of group [Novel: $F_{(1,5)} = 2.26$, $P = 0.189$; Familiar: $F_{(1,5)} = 0.85$, $P = 0.400$]. In novel and familiar blocks, both groups aborted more trials in the −1 vs. −2 condition than any of the other conditions, when compared pairwise (Controls: $P < 0.001$; VS: $P < 0.001$ for all pairs). The number of aborted trials was also larger in the familiar than novel tasks in the −1 vs. −2 condition [Novel vs. Familiar, $F_{(1,5)} = 8.95$, $P = 0.031$].

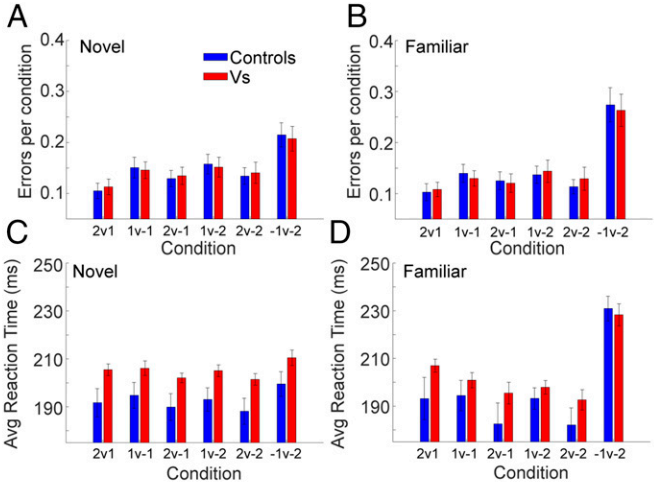

**Fig. 7.** Aborted trials and reaction times averaged across tasks. Note that the data from task 2 are averaged here except the conditions that included a Null cue (i.e., 0/1, 0/−1, etc.). See *SI Appendix*, Fig. S6 for all conditions of task 2. In addition, the ANOVA model included all conditions, as they were nested under Task. (*A*) Aborted trials in the novel conditions. Errors indicate the fraction of trials where the animals held initial fixation, but then failed to select one of the choice options. (*B*) Aborted trials in the familiar condition. (*C*) Reaction times in novel conditions. (*D*) Reaction times in familiar conditions.

Next, we examined reaction times. In both novel (Fig. 7*C*) and familiar (Fig. 7*D*) blocks, there were differences in reaction times across conditions [Novel: $F_{(24,116)} = 11.4$, $P < 0.001$; Familiar: $F_{(24,115)} = 15.2$, $P < 0.001$] and marginal differences across tasks [Novel: $F_{(3,14)} = 3.4$, $P = 0.046$; Familiar: $F_{(3,14)} = 4.09$, $P = 0.029$]. However, there were no differences between groups [Novel: $F_{(1,5)} = 0.45$, $P = 0.534$; Familiar: $F_{(1,5)} = 0.25$, $P = 0.636$] and no higher-order interactions ($P > 0.05$). In both the novel and familiar blocks, animals in both groups were slowest when choosing between the two loss cues relative to all other conditions (Controls: $P < 0.001$; VS: $P < 0.001$ for all pairs).

## Discussion

We carried out four tasks in which we examined learning from gains and losses, using tokens as secondary reinforcers. We found that monkeys learned to make choices that increased their tokens and to avoid choices that decreased their tokens. When we examined group differences in learning novel cues, monkeys with VS lesions were impaired when the feedback was stochastic, and when the large loss choice had a value of −4. We also fit RL models to behavior and found a preference for the CUE model, which had a separate learning rate for each of the cues, relative to the VALENCE model, which had one learning rate for positive outcomes and one for negative outcomes. When we compared learning rates from this model between groups, we found that animals with VS lesions had significantly reduced learning rates specifically for the gain cues. Furthermore, when we examined behavior relative to the VALENCE model, to see where it failed to account well for choices, we found that it specifically overpredicted performance in the 2 vs. 1 condition, and this overprediction was larger in the VS animals than the control animals. This was after optimizing learning rates in this model. Therefore, animals with VS lesions show specific deficits in learning to choose between secondarily reinforced rewarding options, with no apparent deficits in learning to choose between gain and loss cues, or between two loss cues.

Token-based reward mechanisms have been used previously to motivate behavior in macaques (11). We found that monkeys learned effectively to choose options that increased their tokens and avoid options that decreased their tokens. In addition, aversive stimuli can affect behavior in multiple ways (1). Consistent with this, we found that when monkeys had to choose between two loss options, they learned to choose the option leading to the smaller loss. They also aborted significantly more trials and had the longest reaction times when they had to choose between two losses. By these measures, our monkeys found losing tokens to be aversive. We found effects of VS lesions on choice behavior, but not reaction times or aborted trial behavior. Therefore, these behaviors may be mediated by different systems, or the VS may contribute more to choice behavior than speed of response and avoidance.

**Distinct Circuitry Underlying Appetitive and Aversive Learning.** Recent work has attempted to delineate separable neural circuits underlying appetitive and aversive learning. For example, Lammel et al. (16) suggested that a circuit from the lateral habenula, through a subset of dopamine neurons that responded to aversive stimuli, to the IL/PL region of medial prefrontal cortex, was important for aversive learning. Distinct from this, another circuit from the rostral-medial tegmental nucleus, through a subset of dopamine neurons that responded to appetitive stimuli, to the VS, was important for appetitive learning. However, subsequent anatomical work has not supported the suggestion that dopamine neurons with different projection targets have different inputs (18, 19). Other work has suggested that basolateral amygdala neurons that project to the VS are important for appetitive learning, and basolateral amygdala neurons that project to the central nucleus of the amygdala are important for aversive learning (13, 20). Circuitry connecting the amygdala to the dorsal anterior cingulate cortex has also been implicated in aversive learning (21). Inactivation of both ventrolateral prefrontal cortex, and orbitofrontal cortex has also been shown to increase sensitivity to punishment (22). In addition, there is extensive work supporting the amygdala's role in aversive learning (23).

In addition to the circuit work in rodents and nonhuman primates, other work has directly examined learning in the context of winning money, or not losing money, which has similarities to our use of tokens. This work has also suggested that the VS, and dopamine modulation of VS activity, is important for learning to choose rewarding options (5). The same study suggested that the insular cortex was important for learning to avoid losing money, a finding supported by work in patients with insular cortex lesions (24). Notably, avoiding monetary losses has consistently been shown to be independent of dopamine (5, 25). Additional work has shown that aversive pruning, which is the process of eliminating choices that lead to future situations in which large punishments might be experienced, engages the subgenual cingulate cortex (26). It has also been shown that microstimulation in a related subgenual cingulate region can bias choices away from aversive options, although not in the context of learning (27). Thus, the VS has often been implicated in appetitive learning. Aversive learning, on the other hand, has been linked to dorsal cingulate cortex, subgenual cingulate cortex, insular cortex, and the central nucleus of the amygdala. Our data are consistent with the hypothesis that the VS plays a specific role in learning about gains, without contributing to learning about or choosing when losses are involved. In our behavioral data, the deficits in learning about gains were specific to choosing between pairs of gains and did not manifest when a gain was paired with a loss. However, the RL model showed that learning rates were overall lower for gain cues.

The differences between the circuit work in rodents and the systems work in humans and monkeys, which have identified different systems for aversive learning, may in part be due to differences in the appetitive and aversive modalities used. Appetitive and aversive stimuli come in many forms, and these are processed in separable systems, at some level (1, 28). For example, nociceptive information relayed via the dorsal horn of the spinal cord is distinct from information about threats from conspecifics, which may arrive via the auditory or visual system, depending on the nature of the threats. In addition, the processing of token losses would presumably involve different neural circuits from conditioned defensive responses to shock or loud noise. It is currently not clear where information about appetitive or aversive outcomes arising from different modalities is integrated, or if it is ever integrated. Thus, there may be no simple circuit that processes all appetitive or aversive information, independent of modality.

Although the VS is not often implicated in aversive learning, studies have shown that VS neurons respond to both rewarding and aversive stimuli (29). In addition, VS neurons respond to rewarding stimuli that have been subsequently negatively conditioned using injections of lithium chloride (30). Other work has shown that cues that have been negatively conditioned can lead to increased dopamine release in the VS shell, but decreased dopamine release in the VS core (31). In contrast to this, however, tail pinch has been shown to increase dopamine release in the dorsal striatum and the core of the VS (32, 33). Removal of tail pinch also leads to increased dopamine release in the VS shell (32), consistent with the hypothesis that pain relief can be rewarding (34). Further work has shown that oral infusions of quinine, which is aversive, leads to decreased dopamine concentration in the VS core, whereas oral infusion of sucrose leads to increased dopamine concentration (35). Therefore, the relationship between single-neuron responses, dopamine concentration, and appetitive and aversive stimuli in the VS core and shell is complex and depends on the modality of the stimulus and perhaps anesthesia state.

**Learning Deficits in VS-Lesioned Animals.** In our study in the novel condition, we found differences between monkeys with VS lesions and unoperated controls in both task 3, in which outcomes were stochastic, and task 4, in which we used a large loss. We have previously found that animals with VS lesions learn more effectively when outcomes are deterministic, and have substantial deficits when outcomes are stochastic (7, 36, 37). The deficits are consistently largest when the monkeys with VS lesions have to learn to choose between two options that have the same reward magnitude, but differ in reward probability. This may be consistent with work showing that lesions of the VS affect dopamine coding of prediction errors for reward delays, but not reward magnitudes (38). Reward rate estimation, which is required for learning values in tasks with stochastic outcomes, requires estimates of time between rewards.

In the current study, the two options always had different reward magnitudes but the same reward probabilities. While the monkeys with VS lesions had deficits in these tasks, the effect was smaller than we observed in a series of tasks with stochastic outcomes. We have suggested that the amygdala, and also cortical systems (8, 39), learn in parallel with the VS. The amygdala, however, learns with a higher learning rate than the VS (7, 17). Therefore, in monkeys with VS lesions, the amygdala and anatomically related cortical systems may play a larger role in learning than in intact monkeys. The higher learning rate amygdala system is more susceptible to noise, because it rapidly updates value estimates following a nonrewarded choice, and values therefore tend to oscillate when feedback is stochastic (17). The VS updates values with a slower learning rate than the amygdala. When the VS is intact, learning is less affected by stochastic outcomes because the VS value estimates are updated less, after individual outcomes, thereby offsetting the rapid updating carried out by the amygdala. Thus, lesions of the VS lead to larger deficits when feedback is stochastic, and particularly large deficits when only reward probability, and not reward magnitude, can be used to optimize choices. It is likely that the cortex, and mediodorsal thalamus, also contribute to learning in these tasks (39–41). However, how this monosynaptically connected circuit works together to mediate learning is a topic for future research.

We also found group differences in the familiar condition in all tasks. Except in task 1, however, the behavioral differences tended to be rather subtle. One possible explanation for the finding of subtle yet significant differences in the familiar conditions is that the controls often had near-perfect performance in some conditions. Because accuracy is a bounded variable (and despite the fact that we used a transform to normality before running the ANOVAs), this near-perfect performance leads to very small variance, which leads to significant differences. Therefore, the subtle differences in choice accuracy were significant. In most conditions, performance was very high in the conditions that had at least one gain cue. Performance in the −1 vs. −2 condition, or in the 0 vs. loss conditions in task 2, never reached high levels, even after extensive experience.

In previous tasks, we also found that monkeys with VS lesions responded faster than controls (7, 36). In the current task, there were no group differences in reaction times, and there was a trend for the monkeys with VS lesions to respond more slowly than the controls. Thus, the presence of loss cues in the token tasks slowed the reaction times of the VS animals. Previously, we also found that much of the deficit in the VS animals, relative to controls, could be accounted for if reaction times were matched between groups (7). This followed because there was a speed–accuracy trade-off, such that responding quickly led to less consistent choice of the best cue. Thus, the slowed reaction times in the current task may partially explain the accurate performance of the VS-lesioned animals in several conditions.

**RL Model.** An RL model with a separate learning rate for each cue (CUE model) best fit the data for both the control and VS groups in all tasks. For most of our tasks, the VALENCE model is the same as a model that would fit one learning rate for positive RPEs and one for negative RPEs, because values start out at 0 and outcomes are deterministic. When we examined learning rates across experiments, the monkeys with VS lesions had reduced learning rates specifically for gain cues. When we examined performance of the models in each condition, to see where the VALENCE model failed to account for behavior, we found that it overpredicted performance in the 2 vs. 1 condition in both groups, but that this effect was larger in the monkeys with VS lesions relative to controls. Therefore, analysis of learning across experiments showed specific deficits in the animals with VS lesions in learning the values of gain cues, with no overall deficits in learning the values of loss cues.

As a final point, the monkeys in both groups also appeared to learn poorly in the −1 vs. −2 condition, although they did show statistically significant learning in all tasks. We did not find, however, that allowing for a different choice consistency parameter (i.e., inverse temperature) for loss choices improved the fit of the RL model. Both the CUE and VALENCE models used different learning rates for gain and loss cues and learning was slower in the loss conditions. In addition to the smaller learning rates for the loss cues, however, these cues were also chosen less often, and therefore their values were less frequently updated. For example, the +2 cue was frequently chosen in every pair it was part of, whereas the −2 cue was rarely chosen. Value updates only happen in the RL model when an option is chosen and the outcome is experienced. Therefore, the decreased learning in the −1 vs. −2 condition follows both from decreased learning rates and less experience with the outcomes associated with those options.

**Conclusion.** We compared learning from gains and losses in animals with VS lesions and an unoperated control group. We found behavioral deficits in monkeys with VS lesions in two of the four tasks, when comparing choice accuracy. These deficits were consistently driven by trials in which animals had to choose between two cues that differed in positive reward magnitude. There were no deficits when animals had to choose between options, one of which was associated with a loss. We also fit RL models to the data and found that learning rates were lower for gain cues in the VS animals relative to controls. Thus, lesions of the VS, in this task, specifically affected learning to choose between rewarding options and had no effect on learning to avoid losses.

## Methods

**Subjects.** The subjects included six male and one female rhesus macaques with weights ranging from 6 to 11 kg. Three of the male monkeys received bilateral excitotoxic lesions of the VS. The remaining four monkeys served as unoperated controls (three males and one female). One of the male control animals was not able to complete all four tasks, and therefore task 4 only has three controls. For the duration of the study, monkeys were placed on water control. On testing days, monkeys earned their fluid from their performance on the task. Experimental procedures for all monkeys were performed in accordance with *Guide for the Care and Use of Laboratory Animals* (42) and were approved by the National Institute of Mental Health Animal Care and Use Committee.

**Surgery.** Three monkeys received two separate stereotaxic surgeries, one for each hemisphere, which targeted the VS using quinolinic acid. After both lesion surgeries, each monkey received a cranial implant of a titanium head post to facilitate head restraint. Unoperated controls received the same cranial implant. Behavioral testing for all monkeys began after they had recovered from the implant surgery. Lesioned animals were used in three previous studies (7, 36, 37).

**Lesion Assessment.** Lesions of the VS were assessed from postoperative MRI scans. We evaluated the extent of the damage with T2-weighted scans taken after the initial surgeries. For the lesioned monkeys, MR scan slices were matched to drawings of coronal sections from a standard rhesus monkey brain at 1-mm intervals. We then plotted the lesions onto standard sections.

**Task and Apparatus.** We tested rhesus macaques on three deterministic and one stochastic two-arm bandit learning task. We conditioned tokens as reinforcers, which allowed us to assess learning from both gains and losses within the same dimension. All animals completed the four tasks in the same order. Each experimental session was composed of nine novel and three familiar blocks that were randomly interleaved. In each novel block, we introduced images the animal had never seen before and they had to learn the cue–outcome associations. The images in the familiar blocks were kept constant for the duration of a task. We completed testing on each task before beginning the next.

During the experiment, the animals were seated in a primate chair facing a computer screen. Eye movements were used as behavioral readouts. In each single trial, the animals first acquired fixation (Fig. 1*A*). After a fixation hold period, we presented two images, left and right of fixation. The animals made an eye movement to one of the images to indicate their choice. They were allowed to make an eye movement as soon as the targets appeared. After a hold period, the number of tokens associated with their choice was added or subtracted from their accumulated tokens. Every four to seven trials, with the interval randomly selected, the animals received one drop of juice for each token they had at the time of cash-out. When each drop of juice was delivered, one of the tokens disappeared from the screen.

**TkD: Token Task 1 (Deterministic Learning).** In the first task (TkD), novel blocks consisted of 108 trials and familiar blocks of 36 trials. Novel blocks consisted of four images the animals had never seen before. Associated with each image was a value (+2, +1, −1, −2), such that if that image was chosen, the animal gained or lost the corresponding number of tokens. On each trial, monkeys had to acquire and hold central fixation for 500 ms. After monkeys held central fixation, two of the images would appear to the left and the right of the fixation point. The animal chose one by making a saccade to the image and holding for 500 ms. The number of tokens associated with the image was then added or subtracted from their total count, represented by circles at the bottom of the screen. The animals could not have less than zero tokens, however. Therefore, if they had one token and they chose a −2 image, they were reduced to 0 tokens. Every four to seven trials, their tokens were cashed out. At cash-out, the animals were given one drop of juice for each token. When each drop of juice was delivered, one token was removed from the screen. There were six individual conditions in this task, defined by the possible pairs of images. The conditions within a block of 108 trials were presented pseudorandomly. The animals saw each condition twice, once on the left and once on the right, every 12 trials before seeing any condition a third time. At the end of each 108-trial block, we introduced four new images and the animals began the learning from scratch.

**TkN: Token Task 2 (Deterministic Learning).** In the second task, we included an image in the set that, if chosen, led to no change in the number of tokens. Thus, at the beginning of each novel block, we introduced five new images. The images had associated token outcomes of +2, +1, 0, −1, and −2. There were, therefore, 10 different pairs of objects, which we refer to as conditions. These were administered in blocks of 120 trials. Each pair of images was seen twice every 20 trials, with each image presented once on the left and once on the right. As before, the conditions were randomly interleaved every 12 trials.

**TkS: Token Task 3 (Stochastic Learning).** In the third task, we examined performance when feedback was stochastic. In this task, at the beginning of each block, we introduced four new images with associated reward magnitudes of +2, +1, −1, and −2. The design was otherwise the same as the first token task. Except, in this task, in 75% of the trials the number of tokens was adjusted by the magnitude associated with the chosen option, but in 25% of the trials there was no change in the number of tokens. This makes learning more difficult, and information has to be integrated across a larger number of trials to learn the correct choice.

**TkL: Token Task 4 (Deterministic Learning).** In the final task, we again used deterministic feedback to examined performance. This version of the task was similar to task 1 (TkD) with two differences. First, we changed the value of the −2 cue to −4. So the cues for this task were +2, +1, −1, and −4. Second, we gave the animals an endowment of four tokens after every cash-out to maintain motivation, and to increase the number of trials on which they would experience the actual four-token loss.

In some trials, the animals had zero tokens and chose a loss cue. In this case, they had no change in tokens. Therefore, the animals would know that they had not chosen a gain token, but they would not know the magnitude of the loss. In task 1, this happened 12.5% of the time; in task 2, 17.5% of the time; in task 3, 13.2% of the time; and in task 4, 2.6% of the time.

**Images and Eye Tracking.** Images provided as choice options were normalized for luminance and spatial frequency using the SHINE toolbox for MATLAB (43). All images were converted to grayscale and subjected to a 2D fast Fourier transform to control spatial frequency. To obtain a goal amplitude spectrum, the amplitude at each spatial frequency was summed across the two image dimensions and then averaged across images. Next, all images were normalized to have this amplitude spectrum. Using luminance histogram matching, we normalized the luminance histogram of each color channel in each image so it matched the mean luminance histogram of the corresponding color channel, averaged across all images. Spatial frequency normalization always preceded the luminance histogram matching. Each day before the monkeys began the task, we manually screened each image to verify its integrity. Any image that was unrecognizable after processing was replaced with an image that remained recognizable. Eye movements were monitored and the image presentation was controlled by PC computers running the Monkeylogic (version 1.1) toolbox for MATLAB (44) and Arrington Viewpoint eye-tracking system (Arrington Research).

**Reinforcement Learning Models.** We fit a large set of models that varied in the number of parameters they used to model the conditions. In the results, we focus on two models that most often accounted for behavior. All models were built around a Rescorla–Wagner, or stateless RL value update equation given by the following:

$$v_i(t+1) = v_i(t) + \alpha_c(R - v_i(t)). \qquad [1]$$

These values were then passed through a soft-max function to give choice probabilities for the pair presented in each trial:

$$d_j(t) = \left(1 + e^{\beta_k\left(v_i(t) - v_j(t) + h_i(t) - h_j(t)\right)}\right)^{-1}, \quad d_i(k) = 1 - d_j(k). \qquad [2]$$

The variable $v_i$ is the value estimate for option *i*, *R* is the change in the number of tokens that followed the choice in trial *t*, and $\alpha_c$ is the condition-dependent learning rate parameter, for condition *c*. In addition, we also used, for some models, condition-dependent values of the choice consistency or inverse temperature parameter, $\beta_k$. The variable $h_i(t)$ implemented a choice autocorrelation function (45), which increased the value of a cue that had occurred in the same location, recently. The autocorrelation function was defined as follows:

$$h_i(t) = \kappa e^{-\lambda\left(t - t_{l(i)}\right)}, \qquad [3]$$

where the variables $\kappa$ and $\lambda$ were free parameters scaling the size of the effect and the decay rate, respectively. The variable $t_{l(i)}$ indicates the last trial on which a given cue, *i*, was chosen in a given location. There were eight separate values for $t_{l(i)}$ as it tracked the four cues across locations, except for task 2 (TkN), which had 10 values.

We then maximized the likelihood of the animal's choices, *D*, given the parameters present in the model under consideration, using as a cost function:

$$f(D|\alpha_j, \beta_k, \kappa, \lambda) = \prod_t [d_1(k)c_1(k) + d_2(k)c_2(k)], \qquad [4]$$

where $c_1(k)$ was an indicator variable that took on a value of 1 if option 1 was chosen and zero otherwise, and $c_2(k)$ took on a value of 1 if option 2 was chosen and 0 otherwise.

The VALENCE model had one inverse temperature, and two learning rates, one for positive cues and one for negative cues. The CUE model had one inverse temperature, and one learning rate for each cue. Note that the null cue in task 2 would always have a 0 RPE, because the reward associated with this cue was 0, and its values started at 0. Therefore, it does not need a learning rate.

We also explored additional models that had the following: (*i*) one inverse temperature and one learning rate; (*ii*) two inverse temperatures, one for the loss–loss condition and one for the rest of the conditions, and two learning rates, one for positive outcomes and one for negative outcomes; (*iii*) two inverse temperatures, one for the 2 vs. 1 condition, and one for the rest of the conditions, and two learning rates, one for positive feedback and one for negative feedback. None of these models predicted behavior well, however, so to simplify presentation we do not show their results.

**Statistics.** To quantify differences between choice behavior in each group, we performed an arcsine transformation on the choice accuracy values from each session, as this transformation normalizes the data (46). We then carried out an *N*-way ANOVA (ANOVAN). Monkey and session were included as random effects with session nested under monkey. All other factors were fixed effects. The ANOVA on learning-rate parameters across experiments was also done as a mixed-effects ANOVA with session and monkey as random effects and experiment and cue as fixed effects.

All within-group post hoc analysis of aborted trials and reaction time was done using the multcompare function in MATLAB, specifically using the Bonferroni method. Unless otherwise stated, multcompare within group stats will only be reported for the condition that is the least significant.

**ACKNOWLEDGMENTS.** This research was supported by the Intramural Research Program of the National Institute of Mental Health [Grants ZIA MH002928 (to B.B.A.) and MH002887 (to E.A.M.)].

1. Jean-Richard-Dit-Bressel P, Killcross S, McNally GP (2018) Behavioral and neurobiological mechanisms of punishment: Implications for psychiatric disorders. *Neuropsychopharmacology* 43:1639–1650.
2. Kubanek J, Snyder LH, Abrams RA (2015) Reward and punishment act as distinct factors in guiding behavior. *Cognition* 139:154–167.
3. Rasmussen EB, Newland MC (2008) Asymmetry of reinforcement and punishment in human choice. *J Exp Anal Behav* 89:157–167.
4. Farashahi S, Azab H, Hayden B, Soltani A (2018) On the flexibility of basic risk attitudes in monkeys. *J Neurosci* 38:4383–4398.
5. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442: 1042–1045.
6. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441:876–879.
7. Costa VD, Dal Monte O, Lucas DR, Murray EA, Averbeck BB (2016) Amygdala and ventral striatum make distinct contributions to reinforcement learning. *Neuron* 92: 505–517.
8. Averbeck BB, Costa VD (2017) Motivational neural circuits underlying reinforcement learning. *Nat Neurosci* 20:505–512.
9. Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47:129–141.
10. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
11. Seo H, Lee D (2009) Behavioral and neural changes after gains and losses of conditioned reinforcers. *J Neurosci* 29:3627–3641.
12. Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* 4:e233.
13. Namburi P, et al. (2015) A circuit mechanism for differentiating positive and negative associations. *Nature* 520:675–678.
14. Ambroggi F, Ishikawa A, Fields HL, Nicola SM (2008) Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. *Neuron* 59:648–661.
15. Everitt BJ, Morris KA, O'Brien A, Robbins TW (1991) The basolateral amygdala-ventral striatal system and conditioned place preference: Further evidence of limbic-striatal interactions underlying reward-related processes. *Neuroscience* 42:1–18.
16. Lammel S, et al. (2012) Input-specific control of reward and aversion in the ventral tegmental area. *Nature* 491:212–217.
17. Averbeck BB (2017) Amygdala and ventral striatum population codes implement multiple learning rates for reinforcement learning. *IEEE Symposium Series on Computational Intelligence* (IEEE, New York).
18. Beier KT, et al. (2015) Circuit architecture of VTA dopamine neurons revealed by systematic input-output mapping. *Cell* 162:622–634.
19. Menegas W, et al. (2015) Dopamine neurons projecting to the posterior striatum form an anatomically distinct subclass. *eLife* 4:e10032.
20. Beyeler A, et al. (2016) Divergent routing of positive and negative information from the amygdala during memory retrieval. *Neuron* 90:348–361.
21. Taub AH, Perets R, Kahana E, Paz R (2018) Oscillations synchronize amygdala-to-prefrontal primate circuits during aversive learning. *Neuron* 97:291–298.e3.
22. Clarke HF, Horst NK, Roberts AC (2015) Regional inactivations of primate ventral prefrontal cortex reveal two distinct mechanisms underlying negative bias in decision making. *Proc Natl Acad Sci USA* 112:4176–4181.
23. LeDoux JE (2000) Emotion circuits in the brain. *Annu Rev Neurosci* 23:155–184.
24. Palminteri S, et al. (2012) Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. *Neuron* 76:998–1009.
25. Eisenegger C, et al. (2014) Role of dopamine D2 receptors in human reinforcement learning. *Neuropsychopharmacology* 39:2366–2375.
26. Lally N, et al. (2017) The neural basis of aversive Pavlovian guidance during planning. *J Neurosci* 37:10215–10229.
27. Amemori K, Graybiel AM (2012) Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nat Neurosci* 15:776–785.
28. Gross CT, Canteras NS (2012) The many paths to fear. *Nat Rev Neurosci* 13:651–658.
29. Roitman MF, Wheeler RA, Carelli RM (2005) Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron* 45:587–597.
30. Roitman MF, Wheeler RA, Tiesinga PH, Roitman JD, Carelli RM (2010) Hedonic and nucleus accumbens neural responses to a natural reward are regulated by aversive conditioning. *Learn Mem* 17:539–546.
31. Badrinarayan A, et al. (2012) Aversive stimuli differentially modulate real-time dopamine transmission dynamics within the nucleus accumbens core and shell. *J Neurosci* 32:15779–15790.
32. Budygin EA, et al. (2012) Aversive stimulus differentially triggers subsecond dopamine release in reward regions. *Neuroscience* 201:331–337.
33. Park J, Bucher ES, Budygin EA, Wightman RM (2015) Norepinephrine and dopamine transmission in 2 limbic regions differentially respond to acute noxious stimulation. *Pain* 156:318–327.
34. Navratilova E, Atcherley CW, Porreca F (2015) Brain circuits encoding reward from pain relief. *Trends Neurosci* 38:741–750.
35. Roitman MF, Wheeler RA, Wightman RM, Carelli RM (2008) Real-time chemical responses in the nucleus accumbens differentiate rewarding and aversive stimuli. *Nat Neurosci* 11:1376–1377.
36. Rothenhoefer KM, et al. (2017) Effects of ventral striatum lesions on stimulus-based versus action-based reinforcement learning. *J Neurosci* 37:6902–6914.
37. Vicario-Feliciano R, Murray EA, Averbeck BB (2017) Ventral striatum lesions do not affect reinforcement learning with deterministic outcomes on slow time scales. *Behav Neurosci* 131:385–391.
38. Takahashi YK, Langdon AJ, Niv Y, 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.
39. Rudebeck PH, Saunders RC, Lundgren DA, Murray EA (2017) Specialized representations of value in the orbital and ventrolateral prefrontal cortex: Desirability versus availability of outcomes. *Neuron* 95:1208–1220.e5.
40. Chakraborty S, Kolling N, Walton ME, Mitchell AS (2016) Critical role for the mediodorsal thalamus in permitting rapid reward-guided updating in stochastic reward environments. *eLife* 5:e13588.
41. Murray EA, Rudebeck PH (2018) Specializations for reward-guided decision-making in the primate ventral prefrontal cortex. *Nat Rev Neurosci* 19:404–417.
42. National Research Council (2011) *Guide for the Care and Use of Laboratory Animals* (National Academies Press, Washington, DC), 8th Ed.
43. Willenbockel V, et al. (2010) Controlling low-level image properties: The SHINE toolbox. *Behav Res Methods* 42:671–684.
44. Asaad WF, Eskandar EN (2008) A flexible software tool for temporally-precise behavioral control in Matlab. *J Neurosci Methods* 174:245–258.
45. Gershman SJ, Pesaran B, Daw ND (2009) Human reinforcement learning subdivides structured action spaces by learning effector-specific values. *J Neurosci* 29: 13524–13531.
46. Zar JH (1999) *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, NJ).