Systems/Circuits

Effects of Ventral Striatum Lesions on Stimulus-Based versus Action-Based Reinforcement Learning

Kathryn M. Rothenhoefer, Vincent D. Costa, Ramón Bartolo, Raquel Vicario-Feliciano, Elisabeth A. Murray, and Bruno B. Averbeck

Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892

Learning the values of actions versus stimuli may depend on separable neural circuits. In the current study, we evaluated the performance of rhesus macaques with ventral striatum (VS) lesions on a two-arm bandit task that had randomly interleaved blocks of stimulus-based and action-based reinforcement learning (RL). Compared with controls, monkeys with VS lesions had deficits in learning to select rewarding images but not rewarding actions. We used a RL model to quantify learning and choice consistency and found that, in stimulus-based RL, the VS lesion monkeys were more influenced by negative feedback and had lower choice consistency than controls. Using a Bayesian model to parse the groups' learning strategies, we also found that VS lesion monkeys defaulted to an action-based choice strategy. Therefore, the VS is involved specifically in learning the value of stimuli, not actions.

Key words: **reinforcement learning; ventral striatum**

Significance Statement

Reinforcement learning models of the ventral striatum (VS) often assume that it maintains an estimate of state value. This suggests that it plays a general role in learning whether rewards are assigned based on a chosen action or stimulus. In the present experiment, we examined the effects of VS lesions on monkeys' ability to learn that choosing a particular action or stimulus was more likely to lead to reward. We found that VS lesions caused a specific deficit in the monkeys' ability to discriminate between images with different values, whereas their ability to discriminate between actions with different values remained intact. Our results therefore suggest that the VS plays a specific role in learning to select rewarded stimuli.

Introduction

Learning comes in many forms and separable circuits underlie different types of learning (McDonald and White, 1993; Knowlton et al., 1996). Reinforcement learning (RL), or more specifically learning to make choices that yield reward, is frequently attributed to the striatum (Houk et al., 1995; Frank, 2005). Previous studies have shown that the ventral striatum (VS) may function as a critic, assessing whether outcomes after choices are better or worse than expected and generating a prediction error (O'Doherty et al., 2004; Frank, 2005). Complementing this, the dorsal striatum (DS) maintains a choice policy that learns to select rewarded actions on the basis of prediction errors generated by the critic. Previous studies on this problem have used tasks in which instrumental actions are the selection of one of two visual stimuli and are independent of the motor response required to choose the stimulus (O'Doherty et al., 2004). Therefore, it is unclear whether the fMRI correlates seen in these tasks are specific to learning the values of visual stimuli or if they would generalize to instrumental scenarios in which the choice was between two motor actions. Previous work in macaques (Costa et al., 2016) and marmosets (Clarke et al., 2008) has shown that lesions to the VS can affect stimulus-based RL. Related work has shown that injections of dopamine antagonists in the DS can affect choice consistency in the context of RL when monkeys learn to perform specific sequences of actions (Lee et al., 2015). These experiments are consistent with separable systems underlying stimulus and action-based RL, but none of them examined both types of learning simultaneously.

Along with the work on the role of the VS in RL, other studies have shown that the VS can function as a limbic–motor interface (Mogenson et al., 1980; Shiflett and Balleine, 2010). For example, basolateral amygdala inputs to the VS can drive specific Pavlovian-to-instrumental-transfer (PIT), in which the presence of a Pavlovian cue increases responding on a lever that delivers the reward

Received March 8, 2017; revised May 5, 2017; accepted June 6, 2017.

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

This work was supported by the Intramural Research Program of the National Institute of Mental Health–National Institutes of Health, ZIA MH002928-01.

The authors declare no competing financial interests.

Correspondence should be addressed to Bruno B. Averbeck, Ph.D., Laboratory of Neuropsychology, NIMH/NIH, Building 49, Room 1B80, 49 Convent Drive, MSC 4415, Bethesda, MD 20892-4415. E-mail: bruno.averbeck@nih.gov.

DOI:10.1523/JNEUROSCI.0631-17.2017

Copyright © 2017 the authors 0270-6474/17/376902-13$15.00/0

previously associated with the cue (Corbit and Balleine, 2005; Shiflett and Balleine, 2010). In addition, whereas dopaminergic innervation of the VS mediates conditioned reinforcement (Taylor and Robbins, 1984; Cador et al., 1991), the VS was shown to be unnecessary for learning about cues that serve as conditioned reinforcers (Parkinson et al., 1999). Although this work shows that learned cue–outcome associations can lead to potentiation of motor behavior, it does not suggest a specific role for the VS either in learning to associate cues and outcomes or in learning to select actions that produce reward.

In the present study, we investigated whether the VS makes a causal contribution to action-based and stimulus-based RL in the context of a two-armed bandit reversal learning task. If the VS functions as a critic, then it might facilitate learning to select actions defined by both a specific motor response and selection of a specific visual stimulus. Although prior work suggests that action selection may depend more on the DS (Seo et al., 2012; Lee et al., 2015; Parker et al., 2016), it is not known if the VS is similarly involved in action selection. Therefore, we developed a bandit task in which monkeys could learn stochastic reward discriminations by selecting one of two images regardless of the saccade direction (What blocks) or one of two saccade directions regardless of the image (Where blocks). As expected, based on our previous work (Costa et al., 2016), stimulus-based RL was compromised in the VS lesion group. In contrast, action-based RL in the VS lesion group was unimpaired. These results indicate that the VS is not necessary for action-based RL, but is necessary for stimulus-based RL, possibly to drive stimulus selection.

Materials and Methods

Subjects. We studied eight male rhesus monkeys (*Macaca mulatta*), with weights ranging from 6.5–11 kg. Three monkeys received bilateral excitotoxic lesions of the VS and five were used as unoperated controls. For the duration of the study, the monkeys were placed on water control and earned their fluid through their performance on the task on testing days. Experimental procedures for all monkeys 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. Three monkeys received two separate stereotaxic surgeries, one for each hemisphere, which targeted the VS using quinolinic acid (for details, see Costa et al., 2016). 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. Further, the lesion and control animals used in the study reported here were used previously in a study in which they learned only stimulus-based reward associations in a different two-arm bandit task (Costa et al., 2016). After completion of that study, we began training animals on the What–Where two-arm bandit task reported in the current study.

Lesion assessment. Lesions of the VS were assessed from postoperative MRI scans. We evaluated the extent of 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. The extent and location of the ventral striatum lesions are shown in Figure 1*C*.

Experimental setup. The eight monkeys completed an average of 29.13 sessions (SD = 3.83), with an average of 19.51 blocks per session (SD = 4.74). Each block consisted of 80 trials and one reversal of the stimulus-based or action-based reward contingencies (Fig. 1*A*). On each trial, monkeys had to acquire and hold a central fixation point for a random interval (400–600 ms). After the monkeys acquired and held central fixation, two images appeared, one each to the left and right (6° visual angle from fixation) of the central fixation point. The presentation of the two images signaled to the monkeys to make their choice. The monkeys reported their choices by making a saccade to their selection, which could be based on the image or the direction of their saccade. After holding their choice for 500 ms, a reward was stochastically delivered according to the current reward schedule: 80%/20%, 70%/30%, or 60%/40%. The reward schedule for each block was randomly assigned at the start of the block and remained constant throughout the block. If the monkeys failed to acquire central fixation within 5 s, hold central fixation for the required time, or make a choice within 1 s, the trial was aborted and then repeated.

Each block used two novel images that were randomly assigned to the left or right of the fixation point for every trial. The images were changed across blocks but remained constant within a block. What and Where blocks were randomly interleaved throughout the session and block type was not indicated to the monkey. For What blocks, reward probabilities were assigned to each image independently of the saccade direction necessary to select an image. Conversely, for Where blocks, reward probabilities were assigned to each saccade direction independently of the particular images presented on either side of central fixation. The block type (What or Where) was held constant for each 80-trial block. One of the images or one of the saccade directions had a lower probability of being rewarded and the other had a higher probability. The probabilities were determined by which probabilistic schedule (80%/20%, 70%/30%, or 60%/40%) was assigned to that specific block. The trial in which the reward mapping reversed in each block was randomly selected from a uniform distribution from trial 30 to 50, inclusive. The reversal trial was independent of the monkey's performance and was not signaled to the monkey. At the reversal in a What block, the less frequently rewarded image became the more frequently rewarded image and vice versa. At the reversal in Where blocks, the less frequently rewarded saccade direction became the more frequently rewarded saccade direction and vice versa.

Images provided as choice options were normalized for luminance and spatial frequency using the SHINE toolbox for MATLAB (Willenbockel et al., 2010). All images were converted to grayscale and subjected to a 2D FFT 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 (Asaad and Eskandar, 2008) and Arrington Viewpoint eye-tracking system (Arrington Research).

Task training. All experiments reported here were performed after data collection for a previous study had been completed (Costa et al., 2016). Therefore, before we began training on the What–Where task, the animals had extensive experience on learning stimulus-based reward associations in a different two-arm bandit task in which stimulus location was irrelevant to learning reward associations (i.e., it lacked a Where condition). In the current study, to facilitate learning of What and Where reward mappings, all monkeys were trained with a deterministic schedule (100%/0%) in both conditions. They were first introduced to one block type, either What or Where, with block type randomly assigned. Once the monkeys could successfully perform 15–24 blocks per session, we introduced the other block type by itself; then, upon stabilized performance in that block type, we mixed the two block types into one session. Once the monkeys reached stable performance in the deterministic setting, we gradually introduced probabilistic outcomes. Then, probabilities were lowered until the final probabilistic schedules of 80%/20%, 70%/30%, and 60%/40% were reached. Data from the training sessions were not included in any of the analyses.

Saccadic reaction times (RTs). Choice RTs were computed on a trial-by-trial basis and were defined as the time between the onset of the two visual stimuli and the initiation of a saccade that targeted one of the two

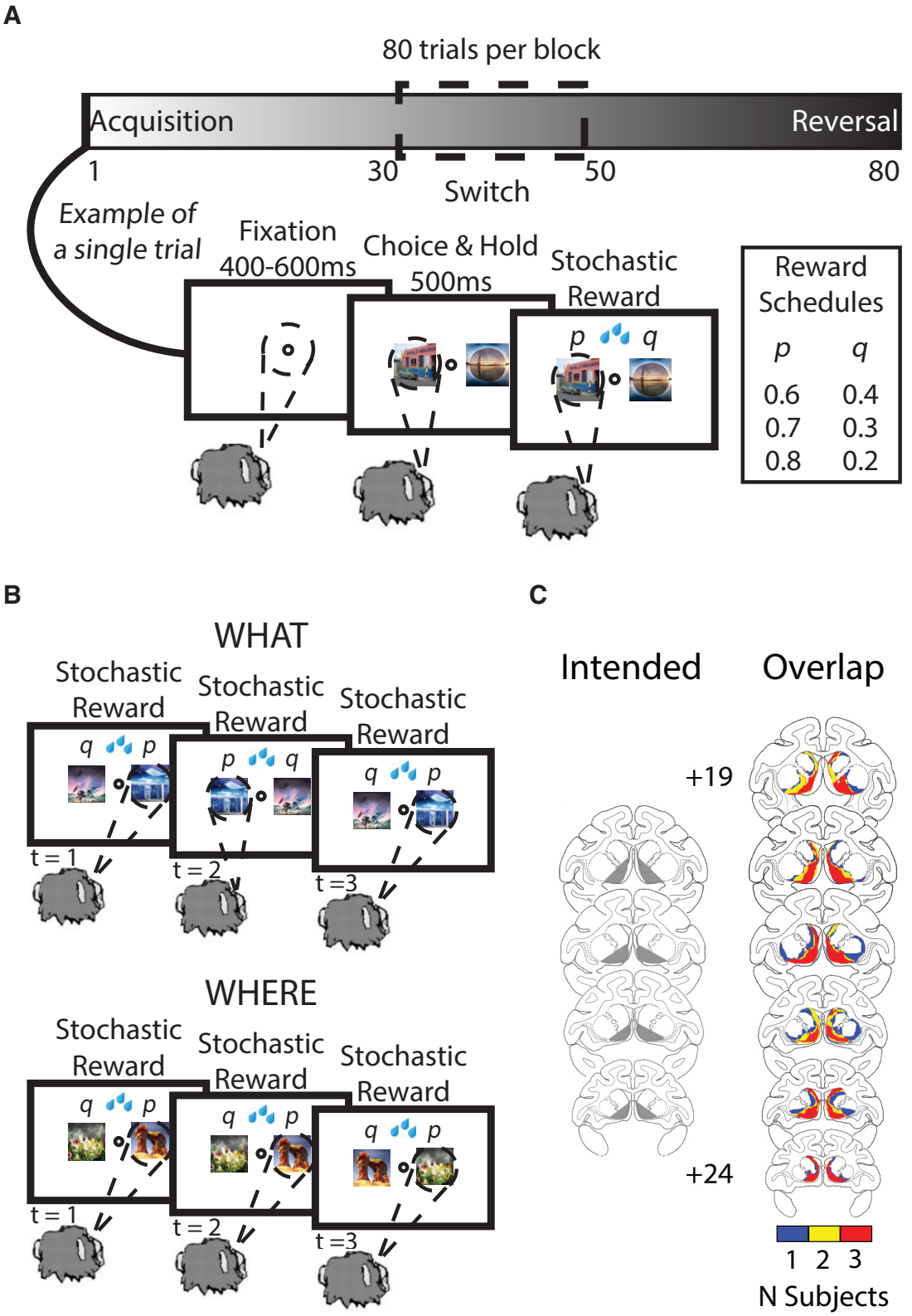

Figure 1. Two-arm bandit task, different block types, and lesion maps. ***A***, The task was presented to the monkeys in blocks of 80 trials. At the start of each block, two novel visual stimuli were used. Dependent upon the block type, either one of the two visual stimuli would be probabilistically more rewarded than the other (What blocks) or one of the two saccade directions would be probabilistically more rewarded than the other (Where blocks). On a randomly selected trial between 30 and 50, the reward mappings were reversed. After the 80 trials of a block were completed, a new block began and two novel stimuli were introduced. For each trial, the monkey had to make an initial fixation, after which two visual stimuli would be presented, one to the left and one to the right of central fixation. The monkeys indicated their choice by making a saccade to one of the two stimuli based on its identity or location. They were then either rewarded or not rewarded depending upon the current reward schedule and which option they chose. ***B***, Three example trials in a What and Where block. In the What block type, the monkey repeatedly chooses a specific visual stimulus that is assigned a higher reward probability. Conversely, in the Where block type, the monkey repeatedly chooses to saccade to one side that is assigned a higher reward probability regardless of what visual stimulus is on that side. ***C***, Lesion extent mapped for the monkeys with bilateral excitotoxic VS lesions (reproduced with permission from Costa et al., 2016).

options. We constructed RT probability functions by binning RTs in 20 ms bins.

Speed–accuracy tradeoff. To investigate the speed–accuracy tradeoff, we examined the groups' choice accuracy on a trial-by-trial basis as a function of their choice RTs. Next, we used a Gaussian kernel (20 ms) to smooth each groups' fraction correct versus choice RT curves. This was sampled in evenly spaced 20 ms bins from 0 to 400 ms. We then averaged the fraction correct as a function of choice RT across schedules for each block type in each group.

RL model. Using the expected reversal trial calculated with the Bayesian model (see "Bayesian model" section), we split each block into an acquisition and reversal phase. We then fit separate RL models to each phase (i.e., acquisition and reversal) of each of the six schedule and block type combinations (i.e., What 80%/20%, 70%/30%, and 60%/40%; Where 80%/20%, 70%/30%, and 60%/40%). Using a standard RL model, we estimated learning rates from positive and negative feedback, as well as the inverse temperature. Value updates were given by the following:

$$v_i(k+1) = v_i(k) + \alpha_f(R - v_i(k)) \quad (1)$$

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

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

The likelihood is then given by the following:

$$f(x, y|\beta, \alpha_{pos}, \alpha_{neg}) = \prod_k [d_1(k)c_1(k) + d_2(k)c_2(k)] \quad (3)$$

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. Conversely, $c_1(k)$ had a value of 0 if option 2 was chosen and $c_2(k)$ had a value of 0 if option 1 was chosen for trial k. We used standard function optimization methods to maximize the log of the likelihood of the data given the parameters. Because estimation can settle in local minima, we used 10 initial values for the parameters. We then used the maximum of the final log-likelihood across fits.

Bayesian model of reversal learning. We fit a Bayesian model to estimate probability distributions over several features of the animals' behavior as well as ideal observer estimates over these features (Costa et al., 2015; Jang et al., 2015; Costa et al., 2016). The Bayesian ideal observer model inverts the causal model for the task, so it is the optimal model. For the current study, we extracted probability distributions over the reversal point as well as the block type. With the Bayesian estimate of the reversal point, we were able to split each block into an acquisition and reversal phase. Particularly for the difficult schedules (i.e., 70%/30% and 60%/40%) this is a better estimate of the information the animal had about reversals than the actual programmed reversal point.

To estimate the Bayesian model we fit a likelihood function given by the following:

$$f(x, y|r, p, h, b) = \prod_{k=1}^{T} q(k) \quad (4)$$

Where r is the trial on which the reward mapping is reversed ($r \in 0–81$), p is the probability of reward of the high reward option. The variable h encodes whether option 1 or option 2 is the high reward option at the start of the block ($h \in 1, 2$) and b encodes the block type ($b \in 1, 2$ – What or Where). The variable k indexes trial number in the block and T is the current trial. The variable k indexes over the trials up to the current trial so, for example, if $T = 10$, then $k = 1, 2, 3, \ldots 10$. The variable r ranges from 0 to 81 because we allow the model to assume that a reversal may not have happened within the block and that the reversal occurred before the block started or after it ended. In either scenario in which the model assumes the reversal occurs before or after the block, the posterior probability of reversal would be equally weighted for $r = 0$ or 81. The choice data are given in terms of x and y, where elements of x are the rewards ($x_i \in 0, 1$) and elements of y are the choices ($y_i \in 1, 2$) in trial i. The variable p was varied from 0.51 to 0.99 in steps of 0.01. It can also be indexed over just the exact reward schedules (i.e., 0.8, 0.7. and 0.6), although this makes little difference because we marginalize over p for all analyses.

For the ideal observer model used to estimate the reversal trial and the "ideal" curve in the Bayesian analysis, we estimated the probability that a reversal happened at the current trial T based on the outcomes from the previous trials. Therefore, the estimate is based on the information that the monkey had when it made its choice in the current trial. For each schedule, the following mappings from choices to outcomes gave us $q(k)$. For estimates of What ($b = 1$), targets 1 and 2 refer to the individual images and saccade direction is ignored; whereas for Where ($b = 2$), targets 1 and 2 refer to the saccade direction and the image is ignored. For $k < r$ and $h = 1$ (when target 1 is the high probability target and the trial is before the reversal) choose 1 and get rewarded $q(k) = p$, choose 1 and receive no reward $q(k) = 1 - p$, choose 2 and get rewarded $q(k) = 1 - p$, choose 2 and have no reward $q(k) = p$. For $k \geq r$, these probabilities are flipped. For $k < r$ and $h = 2$, the probabilities are complementary to the values where $k < r$ and $h = 1$. To estimate reversal, all values were filled in up to the current trial T.

For the animal's choice behavior, used to estimate the posterior over b for each group, the model is similar except the inference is only over the animal's choices, not whether it was rewarded. We assumed that the animal had a stable choice preference which switched at some point in the block from one higher rewarded choice (a saccade direction or an image) to the other. Given the choice preference, the animals chose the less rewarded option at some lapse rate $1 - p$. Therefore, for $k < r$ and $h = 1$, choosing option 1: $q(k) = p$, choosing option 2: $q(k) = 1 - p$. For $k \geq r$ and $h = 1$, choosing option 1: $q(k) = 1 - p$, choosing option 2: $q(k) = p$. Correspondingly, for $k < r$ and $h = 2$, choosing option 1: $q(k) = 1 - p$, choosing option 2: $q(k) = p$. For $k \geq r$ and $h = 2$, choosing option 1: $q(k) = p$, choosing option 2: $q(k) = 1 - p$.

Using these mappings for $q(k)$, we then calculated the likelihood as a function of r, p, h, and b for each block of trials. The posterior is given by the following:

$$p(r, p, h, b|x, y) = f(x, y|r, p, h, b)\,p(r)\,p(p, h, b)/p(x, y) \quad (5)$$

For p, h, and b, the priors were flat. The prior on r, $p(r)$, was for $r < 30$ or $r > 50$, $p(r) = 0$ and for $r > 29$ and $r < 51$, $p(r) = 1/21$. With this prior, there is general agreement between the ideal observer estimate of the reversal point and the actual programmed reversal point (Costa et al., 2015, 2016).

With these priors, we calculated the posterior over the reversal trial by marginalizing over p, h, and b as follows:

$$p(r|x, y) = \sum_{p,h,b} p(r, p, h, b|x, y) \quad (6)$$

The posterior over block type could correspondingly be calculated by marginalizing over r, p, and h.

Before calculating a point estimate of the reversal using the ideal observer, we calculated the posterior evidence that a reversal had occurred before trial k, specifically, the following:

$$p(r < k|x, y) = \sum_{i=1}^{k-1} p(r = i|x, y) \quad (7)$$

To define the reversal trial, we compared this evidence to a threshold, $p(r < k|x, y) > \theta$, and the first trial to exceed the threshold was defined as the reversal point. We assumed a distribution over thresholds to compute a point estimate of the reversal trial, uniform on 0.51–0.99, and computed an expectation over this distribution as follows:

$$< r = k > := < [\min(k)|p(r < k|x, y) > \theta] >_{p(\theta)} \quad (8)$$

Classical statistics. We entered each dependent variable into full factorial, mixed-effects ANOVA models implemented in MATLAB. Dependent variables that we analyzed were fraction correct, learning rate

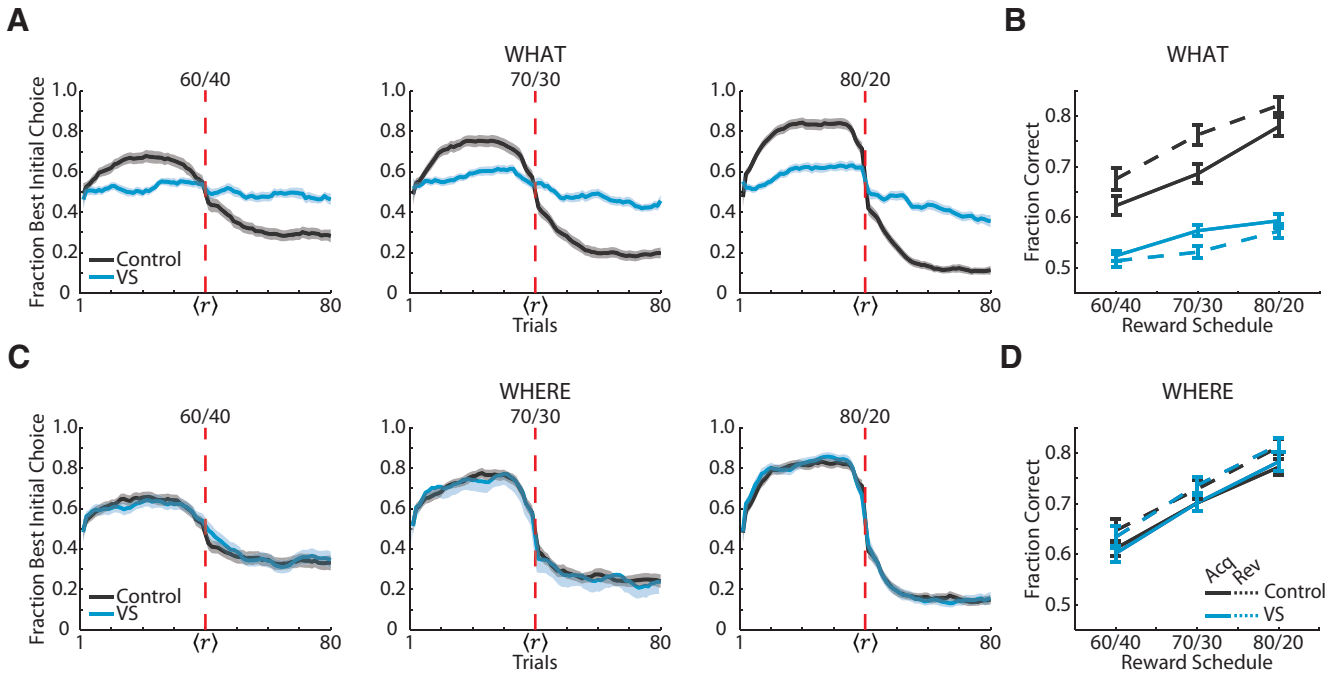

Figure 2. Choice behavior for the lesion and control groups. ***A***, Fraction of times that the monkeys chose the initially higher rewarded visual stimulus option in What blocks averaged by control and VS lesion group and separated into the three different reward schedules. Solid lines show the means, and the shaded region shows ± 1 SEM, which were computed across sessions. The vertical red dashed line shows the reversal point. Because the reversal point varied from block to block, we normalized trial counts in the acquisition and reversal phases for each block. After normalization, we used a moving average window of six trials to smooth the choice curves in the acquisition and reversal phases. ***B***, Overall fraction correct for the control group and VS lesion group for each probabilistic reward schedule and learning phase in What blocks. ***C***, ***D***, The same procedures used in ***A*** and ***B*** were used to plot choice behavior in Where blocks.

(positive and negative), and inverse temperature. When appropriate, group, block type, schedule, feedback type, and learning phase were specified as fixed effects, session as a random factor nested under monkey, and monkey nested under group. In our analyses comparing monkeys' choice behavior in the current task with that in a previous bandit task (Costa et al., 2016), we specified group, task, learning phase, and schedule as fixed effects, with session as a random effect nested under both group and task. For our *post hoc* tests of significant interactions, we computed univariate ANOVAs for component effects and corrected for multiple comparisons.

Results

We tested eight rhesus macaques on a two-armed bandit reversal learning task with a stochastic reward schedule (Fig. 1*A*). The task featured two types of learning blocks: stimulus based (What) and action based (Where) (Fig. 1*B*). The subjects included three monkeys with bilateral excitotoxic lesions of the VS (Fig. 1*C*) and five unoperated controls. The monkeys were tested on multiple, randomly interleaved blocks of 80 trials in each session. Each block was either a What block or a Where block. In addition, the options were stochastically rewarded according to one of three reward schedules: 80%/20%, 70%/30%, or 60%/40%, which was held constant throughout a block. At the beginning of each block, the monkeys were presented with two novel images as choice options. The monkeys were allowed to select one option per trial by making a saccade and fixating on their choice. The images were randomly assigned to the left or right of fixation on each trial. In What blocks, the higher-probability choice was one of the two images independent of the saccade direction needed to select it. In Where blocks, the higher probability choice was one of the two saccade directions independent of the image. There was no cue to indicate which type of block was in force; the monkeys determined block type through inference over choices and feedback. In each block, on a randomly selected trial from 30 to 50, the reward mappings were reversed, making the previously less rewarded option the more rewarded option and vice versa. In addition, the occurrence of a reversal was not signified to the monkeys in any way. After the 80 trials had been completed, a new block began and two novel images were introduced. The monkeys then had to learn again via trial and error whether the reward mapping was based on action type (left or right saccade) or image identity.

Choice behavior

We visualized the monkeys' choice behavior by aligning each block around a reversal point determined by a Bayesian change-point analysis (Fig. 2). The VS lesion group performed substantially worse than controls in What blocks (Fig. 2*A*). In contrast, the VS lesion group performed as well as controls in Where blocks (Fig. 2*C*). To summarize the monkeys' accuracy, we calculated the average fraction correct for each group separated by each schedule and learning phase in both block types (Fig. 2 *B*,*D*).

In What blocks, the controls performed better than the VS lesion group (Fig. 2*B*; group: $F_{(1,224)} = 1005$, $p < 0.001$). In addition, the VS lesion group's performance decreased in the reversal phase from the acquisition phase (phase: $F_{(1,87)} = 14.81$, $p < 0.001$), whereas the control group's performance improved in the reversal phase from the acquisition phase (phase: $F_{(1,136)} = 47.70$, $p < 0.001$). While the VS lesion group performed worse than the control group in What blocks, both groups' performance increased in the richer reward schedules (schedule: $F_{(2,446)} = 146.98$, $p < 0.001$). In addition, this effect of schedule was consistent when we analyzed both groups' individually (VS lesioned: $F_{(2,167)} = 41.85$, $p < 0.001$; control: $F_{(2,267)} = 151.39$, $p < 0.001$). In Where blocks, however, the VS lesion group performed as well as the control group (Fig. 2*D*; group: $F_{(1,221)} = 0.09$, $p = 0.76$). In addition, both groups' performance improved in the richer reward

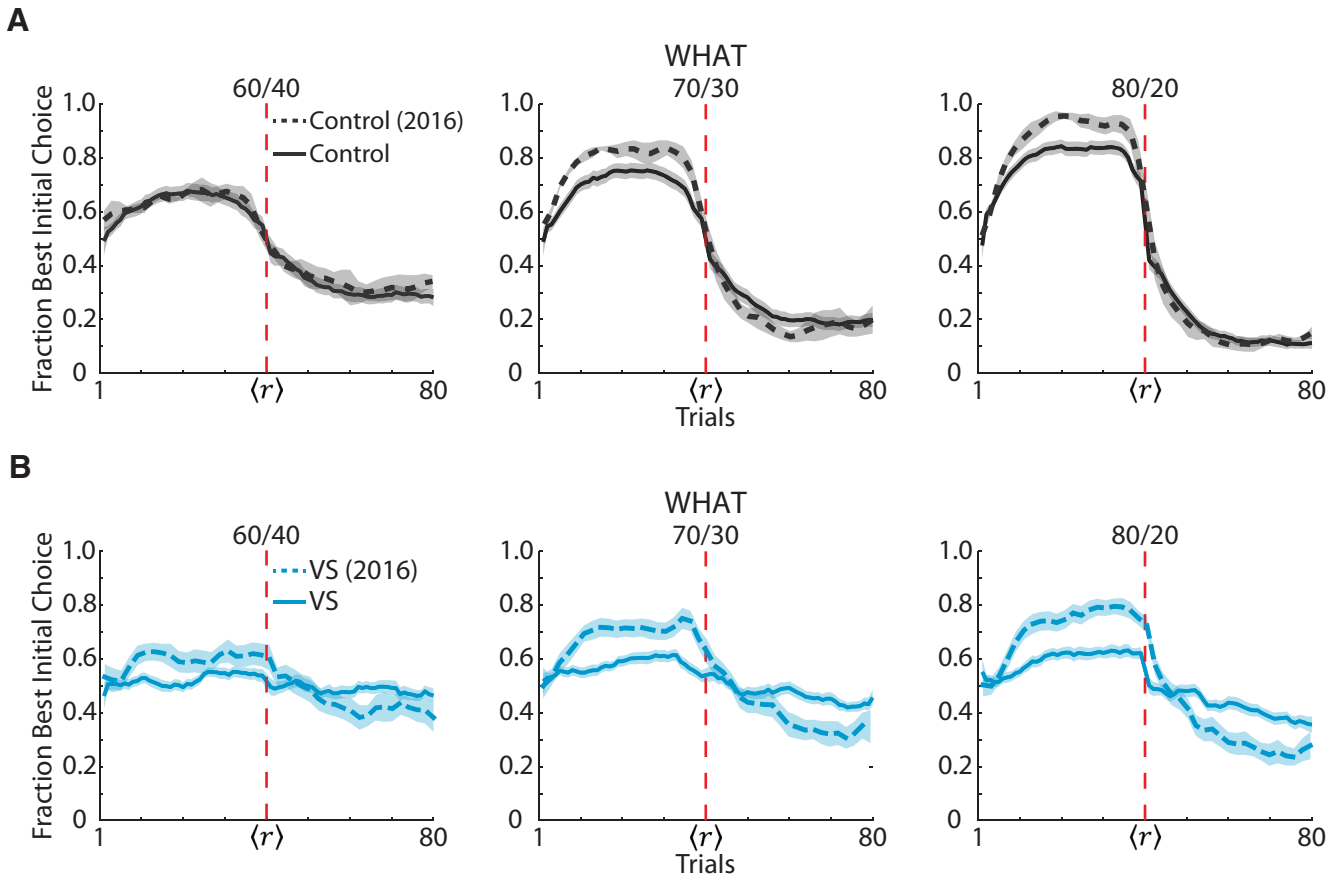

Figure 3. Comparison of stimulus-based RL in bandit tasks with single versus multiple states. ***A***, Fraction of times that the monkeys in the control group chose the visual stimulus that was initially more rewarding in What blocks during the current study and in a previous study (Costa et al., 2016). Although both studies assessed stimulus-based RL, in the previous study, rewards were exclusively associated with choosing a stimulus (one reward state), whereas, in the current study, rewards were associated with either a stimulus or an action (two reward states). Choice behavior in each task is broken out by reward schedule. Solid lines indicate data from the current task and dotted lines indicate data from the previous task. Shaded regions indicate ± 1 SEM computed across sessions. ***B***, Same procedure in ***A*** but for the VS lesion group.

schedules (schedule: $F_{(2,441)} = 353.34$, $p < 0.001$) and in the reversal phase (phase: $F_{(1,228)} = 30.91$, $p < 0.001$). Furthermore, we found that there was no effect of previous block type on performance in the current block, whether it was a What or Where block, for both groups (group × block type × previous block type: $F_{(1,632)} = 0.020$, $p = 0.887$).

The VS lesion animals had previously performed a related version of the current task (Costa et al., 2016). Similar to the current task, the previous task was a two-arm bandit task that had probabilistic, 80-trial blocks with one random reversal between trials 30 and 50. However, the previous task only had stimulus-based reward mapping blocks and no action-based reward mapping blocks. We plotted the choice data from the control and VS lesion groups in What blocks in the current study and from the previous task (Fig. 3). It was clear that having to discern whether actions or stimuli were being rewarded impaired performance (task: $F_{(1,9519)} = 355$, $p < 0.001$) and this was consistent when we analyzed within groups (VS lesioned, task: $F_{(1,2632)} = 322.82$, $p < 0.001$; control, task: $F_{(1,4721)} = 44.83$, $p < 0.001$). However, in the VS lesion group, there was a larger decrease in performance in the current versus the previous task compared with controls (group × task: $F_{(1,4721)} = 44.83$, $p < 0.001$).

RT effects on choice accuracy

Next, we analyzed the RTs in both What and Where blocks, as well as the speed–accuracy tradeoff for each group (Fig. 4). We found that, in both the What and Where blocks, the VS lesion group's RTs were faster than the control group's (KS test, $p < 0.001$). In What blocks (Fig. 4*A*), the VS lesion group's average RT was 159.2 ms and the control group's average RT was 211.4 ms. The control group's accuracy peaks near the mode of their RT distribution. The VS lesion group's accuracy peaks at a point similar to the control group. Because the VS lesion group responded more quickly, however, the peak of their RT distribution is to the left of the peak in their accuracy curve. In Where blocks (Fig. 4*B*), the VS lesion group's average RT was 156 ms and the control group's average RT was 203.2 ms. Unlike in What blocks, both groups showed the highest accuracy at the lowest RTs and then their accuracy decreased as their RTs increased. In addition, the control group's RT distribution peaked as their accuracy began to fall off.

RL model analyses

We fit an RL model to further analyze the monkeys' choice behavior in terms of choice consistency and feedback-dependent learning. We first plotted the RL model predictions over the actual choice behavior separated by group, schedule, and block type (Fig. 5).

Choice consistency

We examined choice consistency using the inverse temperature parameter of the RL model for both groups, block types, phases, and all three stochastic schedules (Fig. 6*A*). The inverse temper-

ature parameter quantifies how consistently the monkeys chose the higher value option. A low inverse temperature indicates noisy choice behavior, whereas a high inverse temperature indicates that the monkeys more frequently chose the better option.

In What blocks, the inverse temperatures were generally higher in the control group compared with the VS lesion group (group: $F_{(1,232)} = 58.14$, $p < 0.001$). In addition, the two groups differed in how consistently they chose the higher value option in the reversal phase (group × phase: $F_{(1,232)} = 8.62$, $p = 0.004$). Among controls, choice consistency increased in the reversal phase compared with the acquisition phase (phase: $F_{(1,110)} = 27.12$, $p < 0.001$). The opposite was observed in the VS lesion group because choice consistency decreased after reversal of the reward contingencies (phase: $F_{(1,58)} = 23.87$, $p < 0.001$). We did not find group differences in the inverse temperature parameter in Where blocks (group: $F_{(1,233)} = 0.63$, $p = 0.428$).

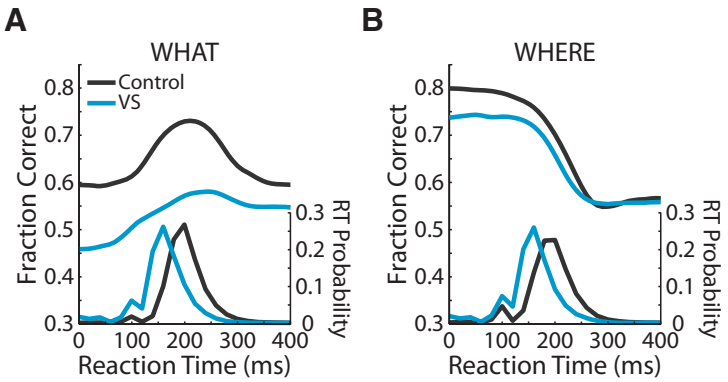

Figure 4. Speed–accuracy tradeoff. ***A***, On the left axis, the fraction of correct choices is plotted as a function of choice RT for both the VS lesion and the control groups in What blocks. On the right axis, choice RT distributions for both groups in What blocks. ***B***, Same as ***A***, but for Where blocks.

Feedback-dependent learning

Next, we investigated the feedback-dependent learning rate parameters, which characterize how quickly the monkeys were able to update the expected value of each of the two choice options in a block. We quantified the monkeys' ability to update value from both negative feedback (no reward) and positive feedback (reward), reflected in negative and positive learning rates, respectively (Fig. 6*B*).

In What blocks, learning rates fit to the behavior of the VS lesion and control groups indicated they differed in their sensitivity to negative and not positive feedback (group × feedback type: $F_{(1,699)} = 14.77$, $p < 0.001$). Learning rates in the control group indicated they were more sensitive to positive versus negative feedback (feedback type: $F_{(1,472)} = 96.49$, $p < 0.001$). However, in the VS lesion group, learning rates indicated equivalent sensitivity to positive and negative feedback (feedback type: $F_{(1,335)} = 3.72$, $p = 0.055$). Moreover, direct comparisons of negative learning rates in the two groups indicated heightened sensitivity of the VS lesion group to negative feedback relative to controls (group: $F_{(1,235)} = 55.79$, $p < 0.001$), whereas positive learning rates did not differ between the groups (group: $F_{(1,239)} = 1.78$, $p = 0.183$). Overweighting negative feedback is maladaptive in a stochastic learning environment and likely prevented the monkeys with VS lesions from accurately learning and updating the values of the two images in What blocks.

We found similar effects when we examined learning rates in Where blocks. The lesion and control groups, again, specifically differed in their sensitivity to negative versus positive feedback (group × feedback: $F_{(1,440)} = 6.77$, $p = 0.009$). In Where blocks, negative learning rates were higher in the VS lesion group compared with controls (group: $F_{(1,228)} = 42.26$, $p < 0.001$), whereas positive learning rates were equivalent in the two groups (group: $F_{(1,227)} = 3.41$, $p = 0.066$). However, unlike in What blocks, both the VS lesion (feedback type: $F_{(1,88)} = 86.95$, $p < 0.001$) and control (feedback type: $F_{(1,130)} = 277.63$, $p < 0.001$) groups exhibited higher learning rates for positive versus negative feedback (feedback type: $F_{(1,796)} = 325.80$, $p < 0.001$). This reflects the improved performance of the VS lesion group in Where compared with What blocks.

Bayesian analysis of reversal learning and posterior probabilities

We used a Bayesian model to predict, for each block, whether the monkeys' choices were consistent with choosing one of the images (What block) or one of the saccade directions (Where block) because this would presumably reflect their inference over block type. We also compared both groups with an ideal observer. For the ideal observer, as the block continued and the observer gained more information about the outcomes associated with each choice, the posterior probability over block type reflected the actual reward mapping in that block. Unsurprisingly, for both What and Where blocks, the ideal posterior probabilities increased throughout the block, reflecting the ideal observer's inference of the correct block type (Fig. 7, pink). In addition, the ideal observer's posterior probabilities were higher in easier reward schedules, reflecting that, with more consistent feedback, the ideal observer can more accurately infer the block type.

In What blocks, the posterior over block type of the control group increased as they completed more trials and thus had more information about their choice options. The control group's increasing posterior probabilities in What blocks show that their choice behavior was consistent with the stimulus-based reward mapping of those blocks (Fig. 7*A*, black). In addition, the control group's posterior probabilities were higher in easier reward schedules, showing that their choices were more consistently reflecting the stimulus-based reward mapping when they had more consistent evidence. Notably, the VS lesion group's posterior probability throughout What blocks was < 0.5. This indicated that their choice behavior was more consistent with the reward mapping of the action-based Where blocks (because the posteriors over block type sum to 1) even when a block's reward mapping is based on the images (Fig. 7*A*, blue; difference between Groups $p < 0.001$, bootstrap analysis across all schedules in What blocks). Much like in What blocks, the control group showed increasing posterior probability over block type throughout the Where blocks as the monkeys gained more information about the reward mapping. In Where blocks (Fig. 7*B*), the VS lesion group show a posterior probability that was either above or aligned with the ideal observer's posterior probability and was consistently above the control group's posterior probabilities (Fig. 7*B*; difference between groups $p < 0.001$, bootstrap analysis across all schedules in Where blocks). Therefore, monkeys with a VS lesion adopted a Where strategy in both block types.

To better illustrate why we believe that the posterior probabilities over block type reflect that the VS lesion monkeys were choosing in accordance to a Where strategy even in What blocks,

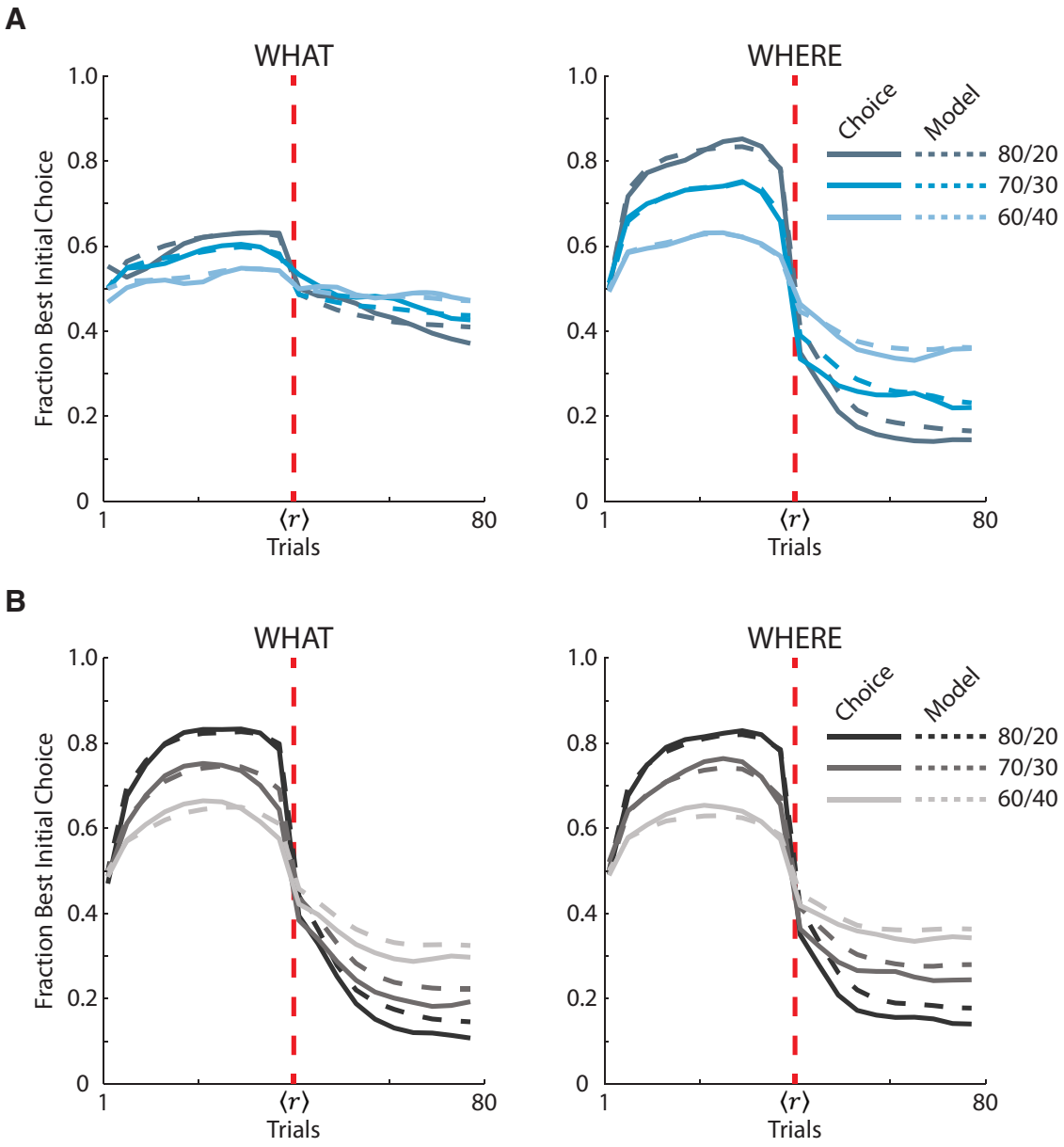

Figure 5. RL model choice behavior predictions compared with actual choice behavior of both lesion and control groups. ***A***, Averaged fraction of times the VS lesion monkeys chose the initially higher rewarded option and the RL model predictions of the VS lesion monkeys' fraction correct separated into the three probabilistic reward schedules for both block types. We separated each block into acquisition and reversal phase and then normalized the trials because of the varying reversal point. The solid lines are the means from the VS lesion monkeys' choice behavior. The dashed lines are the RL models predictions of the VS lesion monkeys' choice behavior. The vertical red dashed line is the reversal point. ***B***, Same as ***A***, but with the control group's choice data and model predictions of their data.

we plotted example blocks from a VS lesion monkey and a control monkey in the 80%/20% What condition (Fig. 8). When we mapped the VS lesion monkey's choices based on the location chosen, it was clear that the monkey was choosing based on location because there were groupings of choices on either side (Fig. 8*A*). In addition, the monkey's posterior probability over Where block type was increasing, whereas the ideal observer's posterior probability was decreasing. This indicated that, based on the feedback, the best strategy would be to choose based on image and not saccade direction (Fig. 8*A*). Conversely, when we mapped the VS lesion monkey's choices based on the image chosen in the same block, there was no consistency in image choice because the choices oscillate from image 1 to image 2 (Fig. 8*B*). Therefore, the VS lesion monkey was not consistently choosing based on image, but rather on location. When we mapped choices from a control monkey based on location chosen in an 80%/20% What block, the monkey's choices oscillated between the two locations, as they should in a What block (Fig. 8*C*). When we mapped the control monkey's choices based on image, the monkey consistently chose image 1, then following the reversal switched and began to consistently choose image 2 (Fig. 8*D*). This was reflected in both the ideal observer's and control monkey's posterior probabilities for the What block type. They both increased and reached asymptote, indicating that the monkey was choosing in accordance with an image-based choice strategy, as would be ideal for the block type.

To analyze the monkeys' posterior updating on a trial-by-trial basis within each block type, we examined the trial-by-trial

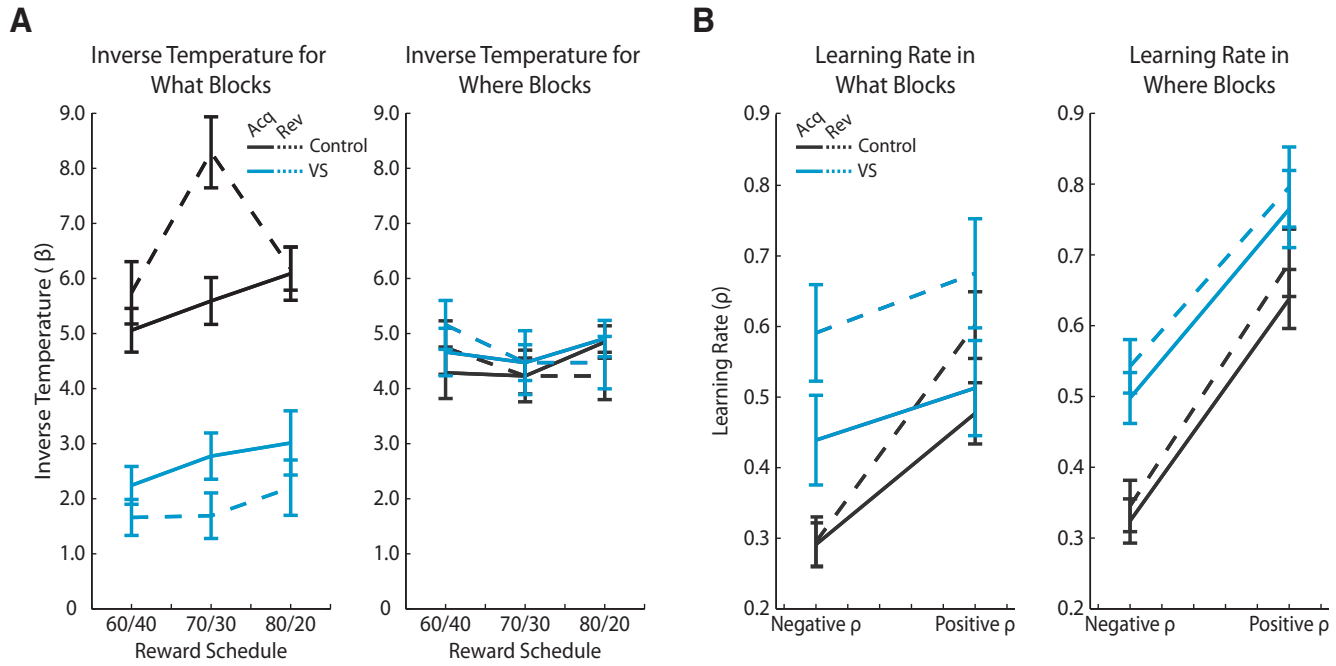

Figure 6. RL model inverse temperature and learning rates. ***A***, RL model inverse temperature parameter separated by groups, phases, and probabilistic reward schedule for each block type. We computed the mean ± 1 SEM across sessions. The black lines represent the control group and the blue lines represent the VS lesion group. The solid lines indicate acquisition phase and the dashed lines indicate reversal phase. ***B***, RL model learning rates separated by groups, phases, and feedback type, for both block types. We computed the mean ± 1 SEM across sessions. The black lines represent the control group and the blue lines represent the VS lesion group. The solid lines indicate acquisition phase and the dashed lines indicate reversal phase.

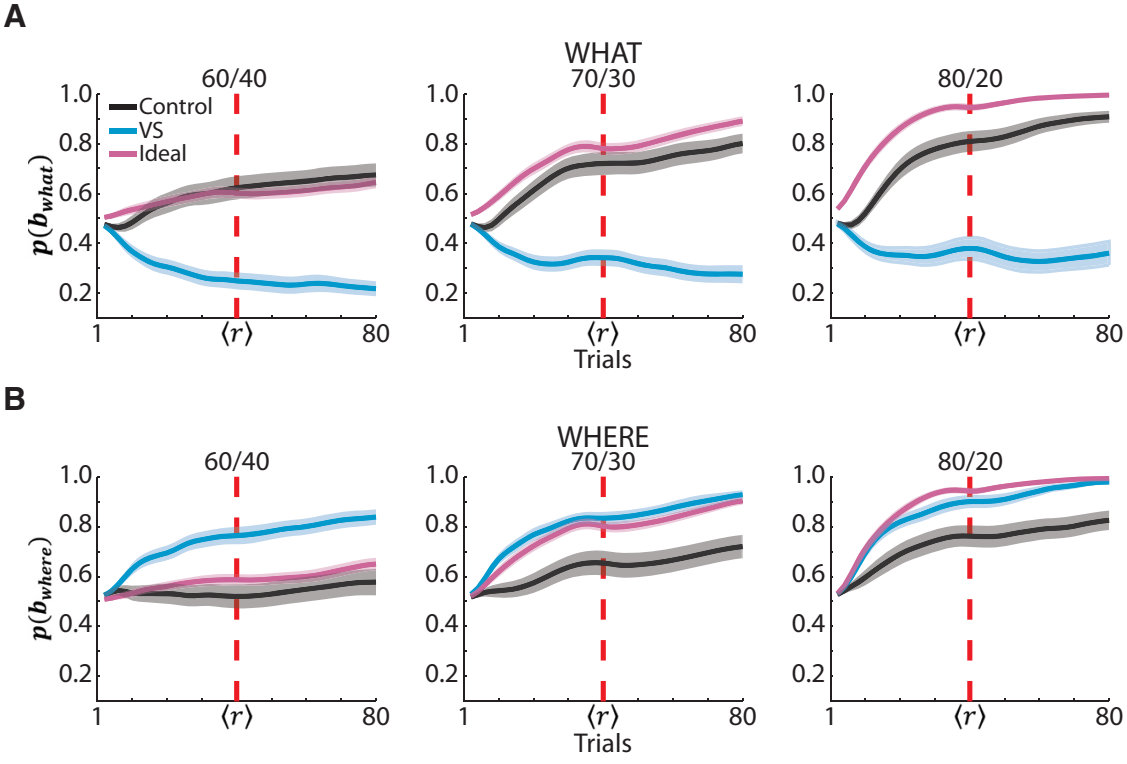

Figure 7. Bayesian posterior probabilities. ***A***, Averaged posterior probability parameter from the Bayesian model in What blocks for the control group, VS lesion group, and the ideal observer (black, blue, and pink, respectively). The dark lines show the means and the shaded region shows ± 1 SEM, computed across sessions. The vertical red dashed line shows the reversal point. Because of the varying reversal point, we normalized the acquisition and reversal phases for each block. After normalization, we used a Gaussian smoothing kernel to smooth the posterior probabilities. ***B***, Same as ***A***, but for Where blocks.

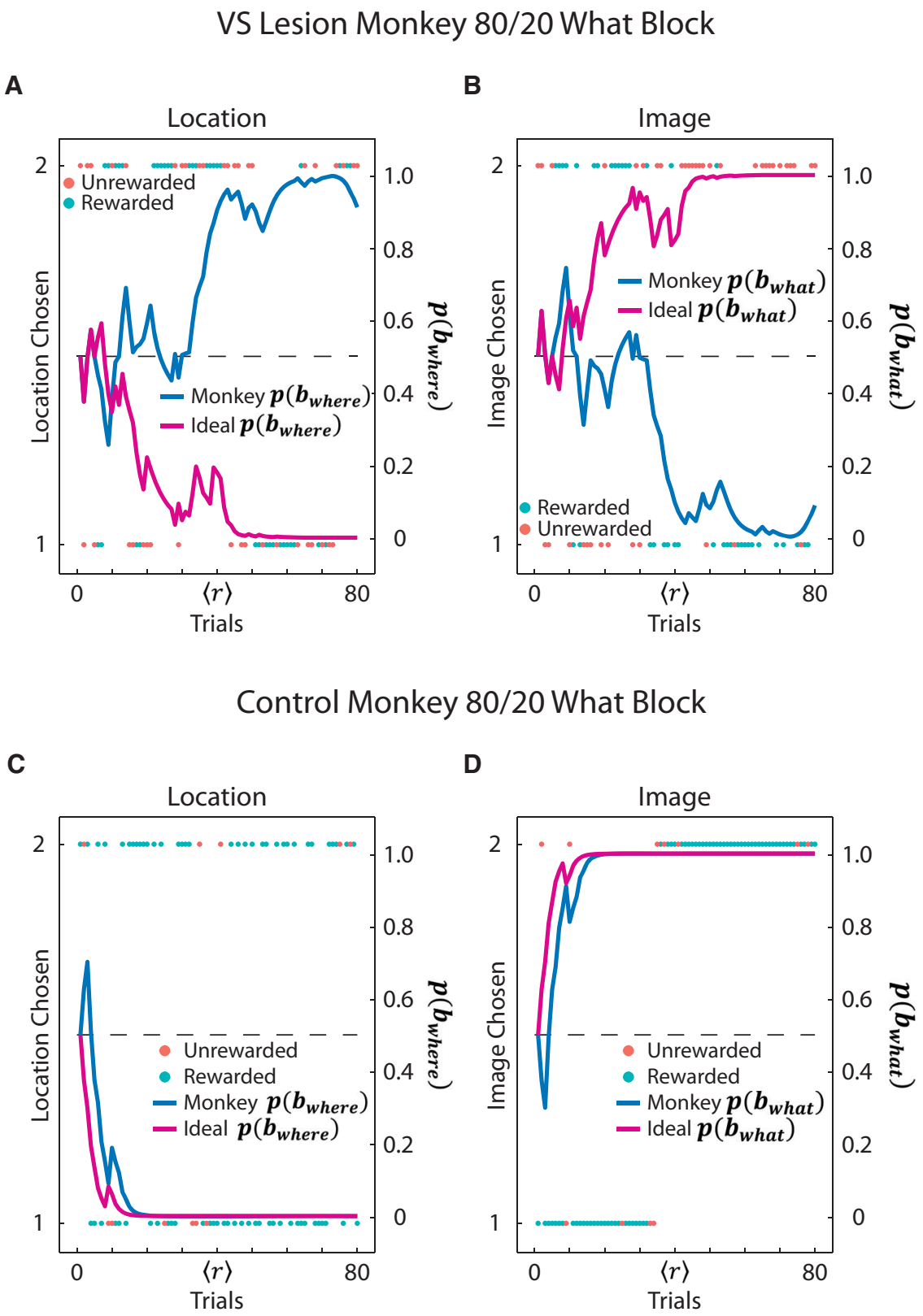

Figure 8. Example Bayesian posterior probabilities. ***A***, ***B***, Example of one VS lesion monkey's choices in one 80%/20% What block separated by the location chosen (***A***) and the image chosen (***B***) and plotted according to whether the choice was rewarded. Overlaid on top are the posterior probabilities for the monkey and for the Bayesian ideal (blue and pink, respectively). For ***A***, the posterior probabilities for the Where block type are shown; for ***B***, the posterior probabilities for the What block type are shown. ***C***, ***D***, Example of one control monkey's choices in one 80%/20% What block, separated by the location chosen (***C***) and the image chosen (***D***) and plotted according to whether the choice was rewarded. Overlaid on top are the posterior probabilities for the monkey and for the Bayesian ideal (blue and pink, respectively). For ***C***, the posterior probabilities for the Where block type are shown; for ***D***, the posterior probabilities for the What block type are shown.

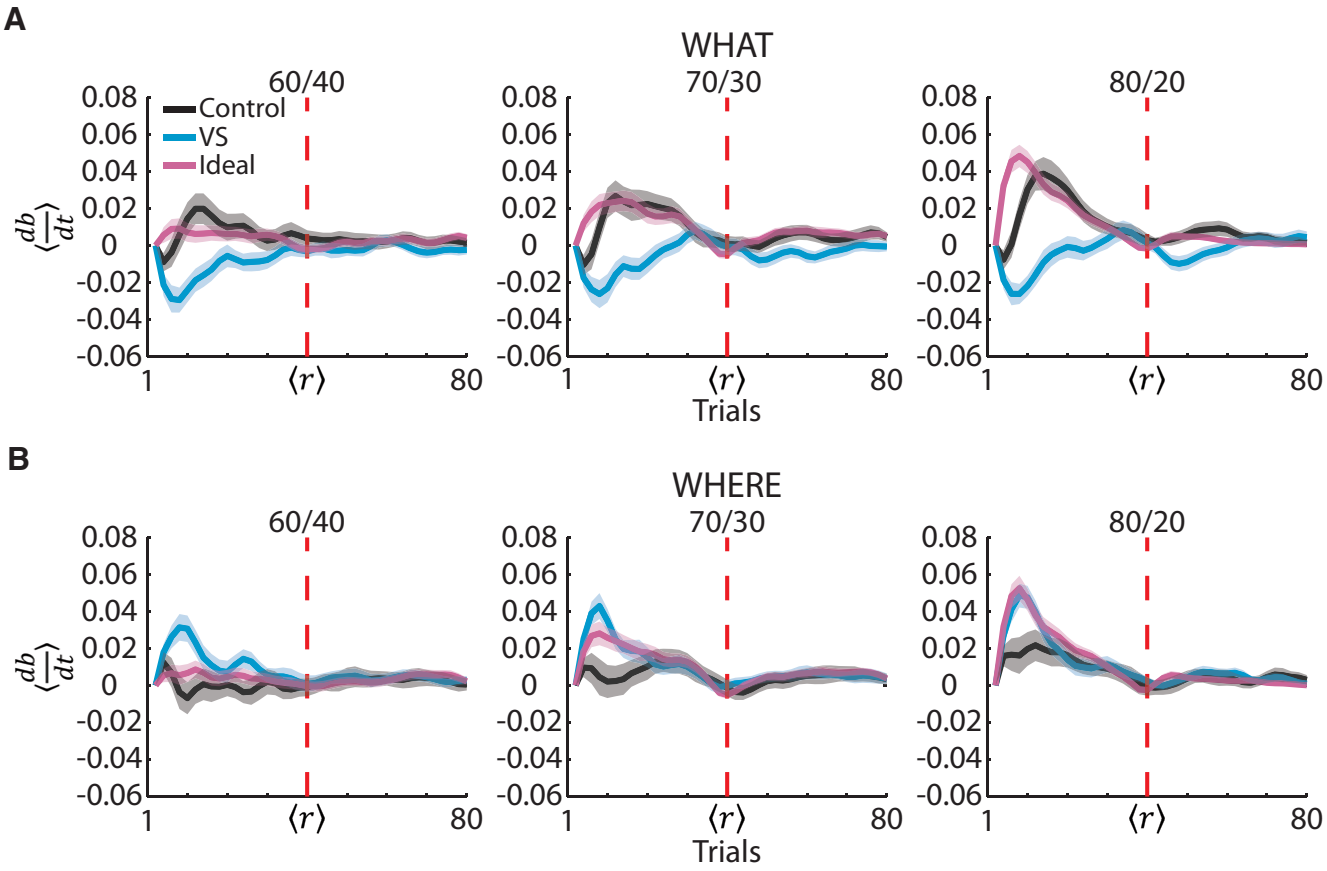

Figure 9. Derivative of the Bayesian posterior probabilities. ***A***, Averaged derivative of the posterior probability parameter from the Bayesian model in What blocks for the control group, VS lesion group, and the ideal observer (black, blue, and pink, respectively). The dark lines show the means and the shaded region shows ± 1 SEM, computed across sessions. The vertical red dashed line shows the reversal point. Because of the varying reversal point, we normalized the acquisition and reversal phases for each block. After normalization, we used a Gaussian smoothing kernel to smooth the derivatives. ***B***, Same as ***A***, but for Where blocks.

derivatives of the posterior probabilities (Fig. 9). A derivative of zero shows that a group's posterior probability was stable. A negative derivative shows that a group's posterior probability was decreasing, changing to choose more consistently with the reward mapping of the wrong block type. A positive derivative shows that a group's posterior probability was increasing and choosing more consistently with the reward mapping of the correct block type. It can be seen that, by the reversal phase, both groups and the ideal observer's derivatives came back to zero because, at that point, choices were consistent with either a stimulus-based or an action-based RL block. In both the What and Where blocks, the control animals tested both stimulus-based and action-based decisions before consistently adopting one of these choice strategies. In contrast, the VS lesion group's behavior reflects that, no matter what type of block they were in, they more consistently chose a saccade direction and not an image. The tendency of the VS lesion group to choose according to saccade direction is shown in Where blocks by the derivatives only being positive (Fig. 9*B*). By only having positive derivatives in Where blocks, this shows that the VS lesion group never considered consistently choosing an image. In What blocks (Fig. 9*A*), this was shown by the VS lesion group's derivatives only being negative. Just like in Where blocks, the VS monkeys' derivatives of the posterior probability over block type reflects that they were making choices based on saccade direction, not on image identity, even though these choices were not being rewarded at a high probability.

Discussion

In the present study, we found that lesions of the VS affected learning to select rewarding stimuli, but not rewarding actions. When we used a RL model to examine the inverse temperature and learning rates, we found that the inverse temperature was lower in What blocks for the VS lesion group, as expected based on their performance. There were no group differences in inverse temperature in Where blocks. The negative learning rate was also higher for the VS lesion group in both block types. In What blocks, elevated negative learning rates reflected the fact that the VS lesion monkeys often responded to the negative feedback by switching to the nonoptimal choice. However, in Where blocks, the high learning rate for negative feedback was balanced by a higher learning rate for positive feedback, which reflected accurate performance.

We used a Bayesian model to infer the choice strategy and found that monkeys with VS lesions were using an action-based strategy in both block types. Interestingly, because the VS lesion group always used a Where strategy, in the Where condition, their inference over block type, shown by the Bayesian posterior probability over block type, was above or aligned with an ideal observer. The control group was somewhat slower to adopt the correct strategy because they entertained both possibilities, stimulus-based and action-based reward mapping.

Distinct lines of research have attributed related but different functions to the VS. Experiments that developed from behavioral learning theory paradigms show that the VS is important for

increasing response rate in tasks in which cues have been associated with rewarding outcomes and then presented during instrumental behavior (Cardinal et al., 2002). For example, in PIT, when a cue has been associated with a rewarding outcome and a lever has been associated with the same rewarding outcome, rats will respond more on that lever in the presence of the cue. Lesions of the VS decrease the response enhancement (Corbit and Balleine, 2005). Similarly, in conditioned reinforcement paradigms, rats will lever press to obtain a cue that has previously been associated with a reward and injection of amphetamines into the VS potentiates this effect (Burns et al., 1993). In both cases, it has been thought that the amygdala underlies the stimulus–outcome association and that amygdala input to the VS drives enhanced responding. These findings have led to the suggestion that the VS functions as a limbic–motor interface (Mogenson et al., 1980; Shiflett and Balleine, 2010).

Aspects of our data were consistent with these findings. Specifically, we have shown the VS is important for learning to select or direct motor responses toward visual cues that were more frequently rewarded. In this respect, our task likely engaged a related function; specifically, the ability of visual cues to drive motor behavior. In our task, however, the cue was driving a choice between options, whereas in previous studies, the cue drove an increase in response rate. An increase in response rate could be conceptualized as a choice between action and inaction, but in any case, the choice is not directed at the cue. Rather, the cue drives an action that leads to the same outcome (PIT) or that delivers the cue (conditioned reinforcement). In this sense, our results were perhaps closest to conditioned reinforcement because choice of a cue in our task is mediated by fixation of the cue. However, the results from neither of these tasks can account directly for our findings.

In a complementary line of research, fMRI has been used in human subjects to study responses in bandit choice tasks like the task in the current study (O'Doherty et al., 2004; Pessiglione et al., 2006). In this work, the BOLD signal in the VS consistently correlates with reward prediction error (RPE) or perhaps with reward during learning. RPEs are the difference between the reward that is expected and the reward that is received. This finding, along with the substantial dopamine innervation of the VS and the finding that the dopamine neurons code RPEs, has led to the assertion that the VS represents the critic in actor–critic RL models (Houk et al., 1995; Collins and Frank, 2014). The critic learns state value representations that provide prediction errors for action learning systems. Our data are consistent with these previous studies in that we found the VS to be important for learning to select the best option of two presented images. The actor–critic RL models, however, frame the behavioral problem as learning to select the correct action and refer to this as instrumental learning. There is no prediction regarding stimulus-based choices. In addition, previous studies have not explicitly separated the motor response from the stimulus chosen; rather, they have examined only stimulus selection independently of the action required to select the stimulus. Although, in both cases (i.e., learning to select an image or learning to perform a specific action), an action is required, our data suggest that learning to select actions differs from learning to select images. Therefore, the single term instrumental is not appropriate in this case.

As mentioned previously, the VS lesion monkeys had faster RTs than control monkeys. This suggests that the VS may be important for delaying responses in some cases to obtain more rewards. The VS has been shown to play an important role in temporal-discounting tasks, in which choices are between immediate small rewards and delayed larger rewards (Bezzina et al., 2007; da Costa Araújo et al., 2009; Valencia-Torres et al., 2012). Although it is unclear whether responding quickly is the same as choosing to get rewarded faster, it is possible that these findings are linked. In addition, it has been shown recently that VS lesions affect RPE responses in dopamine neurons to delayed rewards, but not to reward magnitude (Takahashi et al., 2016). Therefore, this structure appears to play an important role in choices related to reward timing, as well as withholding choices to increase the probability of getting a reward.

In rats, there are anatomical differences between the NAc core and shell (Záborszky et al., 1985). Previous research has shown that shell lesions affect probabilistic response reversal learning, whereas core lesions do not, and core lesions affect switching of response strategies, whereas shell lesions do not (Floresco et al., 2006; Dalton et al., 2014). Although it has been shown that there are structural separations between the NAc core and shell in monkeys, there are no studies identifying functional differences of the same structures in monkeys (Meredith et al., 1996; Friedman et al., 2002) and our lesions covered both.

Although our results clearly implicate the VS in stimulus-based RL, a remaining question is which striatal circuits are critical for action-based RL. Our previous work (Seo et al., 2012; Lee et al., 2015) and that of others (Samejima et al., 2005; Lau and Glimcher, 2008; Parker et al., 2016) suggests a role for the DS in action selection. Specifically, DS neurons code the selection of actions in action-based RL (Seo et al., 2012). Further, the DS has an enhanced representation of action value relative to the lateral prefrontal cortex. Interestingly, this enhanced value representation was found both in a condition in which the monkeys could infer action value using information available within the current trial and when action value had to be learned across trials. Therefore, the DS action value representation was not specific to learned values. To test the causal contribution of the DS to action value learning, we examined the effect of injection of D2 antagonists into the DS and found that inverse temperature and consistency in choosing the best option were affected, but not learning rates (Lee et al., 2015). Future studies comparing the role of the dorsal versus ventral striatum directly in stimulus-based and action-based RL can clarify whether the DS plays a generic or specific role in signaling state value.

We found previously that VS lesions caused learning deficits that were affected by the reward schedule used (Costa et al., 2016). When deterministic stimulus-based mappings were used, deficits were statistically insignificant when the faster RTs of the VS monkeys and their effect on the speed–accuracy tradeoff were controlled for. However, in stochastic schedules such as those used in the current study, the monkeys with VS lesions show deficits that cannot be accounted for by changes in RTs. Interestingly, when we compared the differences in the VS lesion groups' performance in the What blocks from the current task with their performance on the previous task (which had only the What condition), the difference in performance was larger in the lesioned group than in controls between tasks (Costa et al., 2016). It seems that, when action-based RL is an option, it interferes with stimulus-based RL more in monkeys with VS lesions. The Bayesian analysis supports the hypothesis that the VS lesion animals were consistently choosing a side, so their behavior was not random in the What blocks. Therefore, some of the deficit in the current task is attributable to the lesioned animals defaulting to an action-based strategy, or an inability to switch between strategies and inhibit use of an action-based strategy (Floresco et al., 2006).

Conclusion

The current data suggest a specific role for the VS in learning to select a more frequently rewarded image over another that is less frequently rewarded. It is not currently evident whether these results follow from the formation within the VS of stimulus-based reward associations or if the information about stimuli and reward is represented in areas that project to the VS, including the amygdala (Averbeck and Costa, 2017), and orbital and medial prefrontal cortical areas (Haber et al., 2006). However, it is clear that the VS contributes to learning to select stochastically rewarded stimuli but not actions. Therefore, it does not play a generic, state-dependent role as a critic in RL.

References

Asaad WF, Eskandar EN (2008) A flexible software tool for temporally-precise behavioral control in Matlab. J Neurosci Methods 174:245–258. CrossRef Medline

Averbeck BB, Costa VD (2017) Motivational neural circuits underlying reinforcement learning. Nat Neurosci 20:505–512. CrossRef Medline

Bezzina G, Cheung TH, Asgari K, Hampson CL, Body S, Bradshaw CM, Szabadi E, Deakin JF, Anderson IM (2007) Effects of quinolinic acid-induced lesions of the nucleus accumbens core on inter-temporal choice: a quantitative analysis. Psychopharmacology (Berl) 195:71–84. CrossRef Medline

Burns LH, Robbins TW, Everitt BJ (1993) Differential effects of excitotoxic lesions of the basolateral amygdala, ventral subiculum and medial prefrontal cortex on responding with conditioned reinforcement and locomotor activity potentiated by intra-accumbens infusions of D-amphetamine. Behav Brain Res 55:167–183. CrossRef Medline

Cador M, Taylor JR, Robbins TW (1991) Potentiation of the effects of reward-related stimuli by dopaminergic-dependent mechanisms in the nucleus accumbens. Psychopharmacology (Berl) 104:377–385. CrossRef Medline

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26:321–352. CrossRef Medline

Clarke HF, Robbins TW, Roberts AC (2008) Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. J Neurosci 28:10972–10982. CrossRef Medline

Collins AG, Frank MJ (2014) Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol Rev 121:337–366. CrossRef Medline

Corbit LH, Balleine BW (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. CrossRef Medline

Costa VD, Tran VL, Turchi J, Averbeck BB (2015) Reversal learning and dopamine: a bayesian perspective. J Neurosci 35:2407–2416. CrossRef Medline

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. CrossRef Medline

da Costa Araújo S, Body S, Hampson CL, Langley RW, Deakin JF, Anderson IM, Bradshaw CM, Szabadi E (2009) Effects of lesions of the nucleus accumbens core on inter-temporal choice: further observations with an adjusting-delay procedure. Behav Brain Res 202:272–277. CrossRef Medline

Dalton GL, Phillips AG, Floresco SB (2014) Preferential involvement by nucleus accumbens shell in mediating probabilistic learning and reversal shifts. J Neurosci 34:4618–4626. CrossRef Medline

Floresco SB, Ghods-Sharifi S, Vexelman C, Magyar O (2006) Dissociable roles for the nucleus accumbens core and shell in regulating set shifting. J Neurosci 26:2449–2457. CrossRef Medline

Frank MJ (2005) Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cogn Neurosci 17:51–72. CrossRef Medline

Friedman DP, Aggleton JP, Saunders RC (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. CrossRef Medline

Haber SN, Kim KS, Mailly P, Calzavara R (2006) Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. J Neurosci 26:8368–8376. CrossRef Medline

Houk JC, Adamas JL, Barto AG (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 (Houk JC, Davis JL, Beiser DG, eds), pp 249–274. Cambridge, MA: MIT.

Jang AI, Costa VD, Rudebeck PH, Chudasama Y, Murray EA, Averbeck BB (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. CrossRef Medline

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

Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451–463. CrossRef Medline

Lee E, Seo M, Dal Monte O, Averbeck BB (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. CrossRef Medline

McDonald RJ, White NM (1993) A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. Behav Neurosci 107:3–22. CrossRef Medline

Meredith GE, Pattiselanno A, Groenewegen HJ, Haber SN (1996) Shell and core in monkey and human nucleus accumbens identified with antibodies to calbindin-D28k. J Comp Neurol 365:628–639. CrossRef Medline

Mogenson GJ, Jones DL, Yim CY (1980) From motivation to action: functional interface between the limbic system and the motor system. Prog Neurobiol 14:69–97. CrossRef Medline

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452–454. CrossRef Medline

Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, Daw ND, Witten IB (2016) Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat Neurosci 19:845–854. CrossRef Medline

Parkinson JA, Olmstead MC, Burns LH, Robbins TW, Everitt BJ (1999) Dissociation in effects of lesions of the nucleus accumbens core and shell on appetitive Pavlovian approach behavior and the potentiation of conditioned reinforcement and locomotor activity by D-amphetamine. J Neurosci 19:2401–2411. Medline

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. CrossRef Medline

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340. CrossRef Medline

Seo M, Lee E, Averbeck BB (2012) Action selection and action value in frontal-striatal circuits. Neuron 74:947–960. CrossRef Medline

Shiflett MW, Balleine BW (2010) At the limbic-motor interface: disconnection of basolateral amygdala from nucleus accumbens core and shell reveals dissociable components of incentive motivation. Eur J Neurosci 32:1735–1743. CrossRef Medline

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. CrossRef Medline

Taylor JR, Robbins TW (1984) Enhanced behavioural control by conditioned reinforcers following microinjections of d-amphetamine into the nucleus accumbens. Psychopharmacology (Berl) 84:405–412. CrossRef Medline

Valencia-Torres L, Olarte-Sanchez CM, da Costa Araújo S, Body S, Bradshaw CM, Szabadi E (2012) Nucleus accumbens and delay discounting in rats: evidence from a new quantitative protocol for analysing inter-temporal choice. Psychopharmacology (Berl) 219:271–283. CrossRef Medline

Willenbockel V, Sadr J, Fiset D, Horne GO, Gosselin F, Tanaka JW (2010) Controlling low-level image properties: the SHINE toolbox. Behav Res Methods 42:671–684. CrossRef Medline

Záborszky L, Alheid GF, Beinfeld MC, Eiden LE, Heimer L, Palkovits M (1985) Cholecystokinin innervation of the ventral striatum: a morphological and radioimmunological study. Neuroscience 14:427–453. CrossRef Medline