Behavioral/Cognitive

Injection of a Dopamine Type 2 Receptor Antagonist into the Dorsal Striatum Disrupts Choices Driven by Previous Outcomes, But Not Perceptual Inference

Eunjeong Lee, Moonsang Seo, Olga Dal Monte, and Bruno B. Averbeck
Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland 20892-4415

Decisions are often driven by a combination of immediate perception and previous experience. In this study, we investigated how these two sources of information are integrated and the neural systems that mediate this process. Specifically, we injected a dopamine type 1 antagonist (D1A; SCH23390) or a dopamine type 2 antagonist (D2A; eticlopride) into the dorsal striatum while macaques performed a task in which their choices were driven by perceptual inference and/or reinforcement of past choices. We found that the D2A affected choices based on previous outcomes. However, there were no effects of the D2A on choices driven by perceptual inference. We found that the D1A did not affect perceptual inference or reinforcement learning. Finally, a Bayesian model applied to the results suggested that the D2A may be increasing noise in the striatal representation of value, perhaps by disrupting the striatal population that normally represents value.

Key words: **action value; dorsal striatum; neuromodulation; Parkinson's disease; reinforcement learning; sequential decision making**

Introduction

Decisions are often driven by a combination of immediate perceptual evidence and previous experience. Several groups have studied perceptual decision-making tasks, in which stochastic perceptual information presented to subjects drives choices among options (Shadlen and Newsome, 2001; Ratcliff et al., 2003; Fetsch et al., 2013). In these experiments, all of the information required to make a choice is presented within the current trial, and past outcomes cannot be used to better predict the current choice. Perceptual decision making has often been studied in cortical networks (Newsome et al., 1989; Shadlen and Newsome, 2001; Gu et al., 2008; Rorie et al., 2010). However, subcortical areas, including the colliculus and basal ganglia, have been implicated in these tasks (Lovejoy and Krauzlis, 2010; Ding and Gold, 2013). The exact role of various areas is currently unclear, however.

In another line of research, several groups have examined reinforcement learning (Barraclough et al., 2004; Samejima et al., 2005; Pessiglione et al., 2006; Eisenegger et al., 2014) or learned value tasks (Wallis, 2012; Rudebeck et al., 2013). In these tasks, choices are driven by previously experienced associations between rewards and cues, and there is no explicit information available in the current trial to drive the choice. Reinforcement learning or learning from past outcomes has often been attributed to plasticity within the striatum (Graybiel, 2008; Cockburn et al., 2014). Dopamine neurons show phasic responses to reward prediction errors when animals are highly over-trained in pavlovian tasks (Schultz et al., 1997), and the dopamine neurons send a large projection to the striatum (Haber and Fudge, 1997). Recent studies have also implicated dopamine causally in reinforcement learning (RL) (Steinberg et al., 2013). Some studies, however, suggest that learning can take place in the absence of dopamine (Robinson et al., 2005).

In many situations, both immediate perception and past experience can be used to drive choices. Therefore, we wanted to understand how these two sources of information are integrated and the neural systems that mediate this process. In the current study, animals performed a sequence of choices where individual choices were driven by immediately available perceptual information and/or reinforcement of past actions. We previously found that the lateral prefrontal cortex represents choices before the dorsal striatum (dStr) when choices are based on immediately available perceptual information (Seo et al., 2012). The dStr, on the other hand, represents the value of actions, whether value is driven by the reinforcement of previous choices or immediately available perceptual information. To examine the contribution of dopamine and the dStr to the task, we injected dopamine antagonists into the dStr while animals performed the task. We found that a dopamine type 2 receptor antagonist (D2A; eticlopride) injected into the dorsal striatum specifically affected choices based on previous outcomes. There were no effects of the D2A on choices driven by perceptual inference, and there were no effects of a dopamine type 1 receptor antagonist (D1A; SCH23390) on perceptual inference or reinforcement learning. A model applied

Received Nov. 4, 2014; revised Feb. 2, 2015; accepted Feb. 26, 2015.

Author contributions: E.L. and B.B.A. designed research; E.L., M.S., O.D.M., and B.B.A. performed research; E.L. and B.B.A. analyzed data; E.L. and B.B.A. wrote the paper.

This work was supported by the Brain Research Trust, the Wellcome Trust, and the Intramural Research Program of the National Institute of Mental Health.

This article is freely available online through the *J Neurosci* Author Open Choice option.

Correspondence should be addressed to Dr. Bruno B. Averbeck, 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.4561-14.2015

 0270-6474/15/356298-09$15.00/0

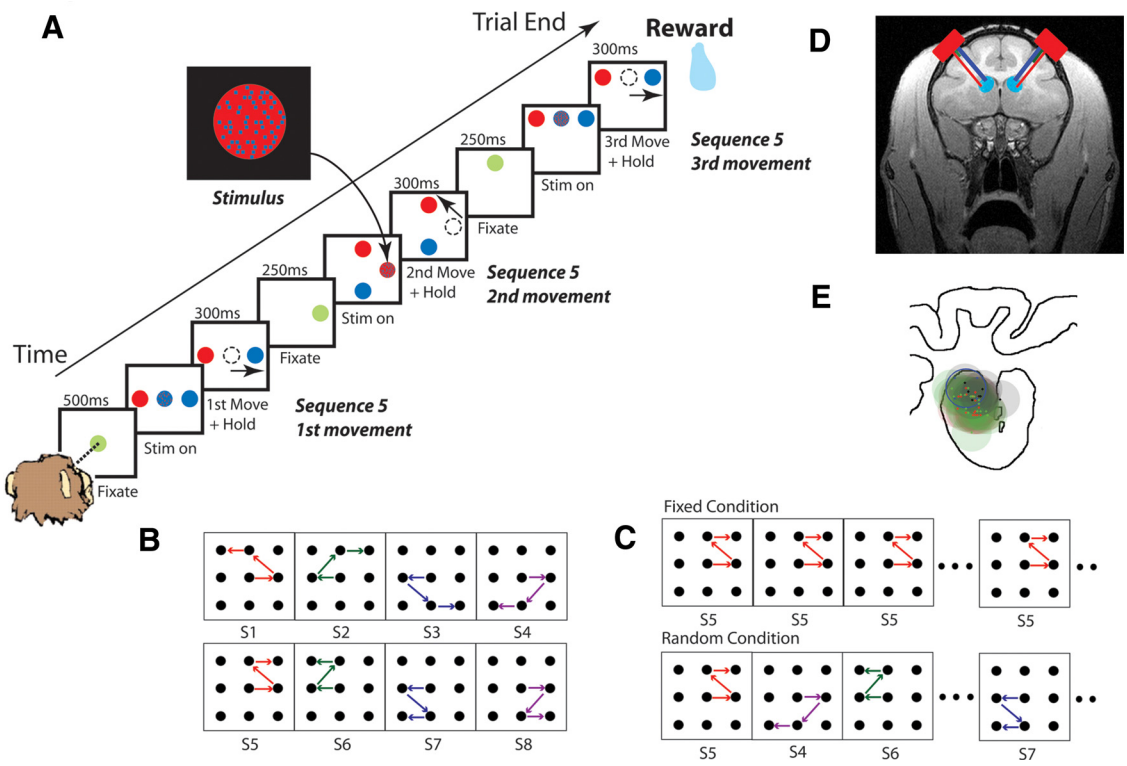

Figure 1. Task. ***A***, Events in a single trial. The beginning of a new trial was indicated by a green dot at the center of the initial fixation frame. A stimulus with blue and red pixels indicated the correct decision to the peripheral target matching the dominant color of the stimulus. The inset shows an example of a single frame from the stimulus. ***B***, A monkey was trained to execute eight possible sequences of saccadic eye movements. Each movement occurred in at least two sequences. S1, S2, etc., indicate sequence 1, sequence 2, etc. ***C***, In the fixed condition (top), the sequence of eye movements was fixed for eight correct trials and then switched to a new sequence and remained fixed again. In the random condition (bottom), the sequence changed every trial. ***D***, MRI of the anterior portion of the macaque brain with approximate bilateral injection areas. The red squares indicate the location of the chambers. Blue, green, and red lines indicate injection cannulae, electrodes in PFC, and electrodes in dStr, respectively. ***E***, Locations of the injection sites for saline (green dots and circles), the D2R antagonist (red dots and circles), and the D1R antagonist (black dots and circles). The dots show the center points of the four cannulae for each session. The shaded circles show approximate injected areas. Injection locations were symmetric in the two hemispheres, so we only show them for one hemisphere. The coronal section is taken from +29 mm anteroposterior (AP) (center of the chamber was +28 AP, 17 medial-lateral). Four cannulae were used for each hemisphere. For one of the sessions, the locations of the four individual cannulae are shown in blue to show the orientation. The cannulae were spaced by ~2.5 mm dorsoventral and 2.7 mm mediolateral.

to the results suggested that the D2A may be increasing noise in the striatal representation of value.

Materials and Methods

General

Three male rhesus macaques (B, C, and I) were used as subjects in this study. Experimental procedures for monkey B were in accordance with the United Kingdom Animals (Scientific Procedures) Act 1986, and the procedures for monkeys C and I were in accordance with the National Institutes of Health *Guide for the Care and Use of Laboratory Animals*. Most procedures were equivalent except that the UK animal received food pellet rewards whereas the U.S. animals received juice reward.

Behavior task

The monkeys were trained on an oculomotor sequential decision-making task (Fig. 1). During the task, the monkeys performed perceptual inference to determine whether a fixation point had more blue or more red pixels, and they made a saccade to the peripheral target that matched the majority color (Fig. 1*A*). After completing the saccade correctly, the saccade target became the new fixation point, and the monkey had to repeat the perceptual inference to select the next target in the sequence. After selecting three targets correctly, the monkey was given a juice reward. After an incorrect decision, the monkey was forced back to the previous fixation point, not the beginning of the sequence, and was allowed to repeat the decision. This operation was performed under two conditions: random and fixed. In the random condition, the correct spatial sequence of eye movements varied from trial to trial, and the animal had to perform a difficult perceptual decision to determine where to saccade (Fig. 1*C*). In the fixed condition, the spatial sequence remained the same for blocks of eight correct trials. Thus, once the animal learned the sequence, it could execute it from memory without using information in the fixation stimulus. Since we used a fixed set of eight sequences (Fig. 1*B*), all of the sequences were well learned. The fixation stimulus was generated by randomly choosing the color of each pixel in the stimulus ($n = 518$ pixels) to be blue (or red) with probability q (Fig. 1*A*, inset). On each screen refresh, the color of 10% of the pixels was updated. The color bias, q, was selected randomly for each movement and could be different for the different movements in a trial: in the fixed condition, q {(0.50, 0.55, 0.60, 0.65); and in the random condition, q} (0.50, 0.55, 0.60, 0.65).

Each day's session was randomly started with either a fixed or a random block. Then the two conditions were interleaved. Each random block was 64 completed trials, where a trial was only counted as completed if the animal made it to the end of the sequence and received a reward. We analyzed only completed trials. Each fixed block was 8 correct trials of each sequence (64 total correct trials, where a correct trial is a completed trial without any incorrect choices).

Surgery and drug injection procedures

Stainless steel chambers (18 mm diameter: two bilaterally in monkeys C and I; one unilaterally in monkey B) were placed over the lateral prefrontal cortex (lPFC) in a sterile surgery using stereotaxic coordinates derived from a structural MRI. In each session, we inserted four injection cannulae bilaterally (in monkey B, two injection cannulae unilaterally), along with electrodes, into the dStr. The electrodes were used to map the depths at which single neuron activity was recorded. This allowed us to deter-

mine the depth for the injection cannulae. After the injection cannulae were lowered, the animals performed two (for monkey I) or four (for monkeys B and C) blocks of the task (one or two blocks of the random condition and one or two blocks of the fixed condition). After completing the baseline blocks (subsequently referred to as the predrug period), we injected a D1A (SCH23390, 2 μg/1 μl for bilateral injection, 10 μg/1 μl for unilateral injection), a D2A (eticlopride, 1.2 μg/1 μl), or saline. In monkey I, we performed nine injections of the D2A and six injections of saline. In monkey C, we performed 3 injections of the D1A, 12 injections of the D2A, and 7 injections of saline. In monkey B, we performed four injections of the D1A and four injections of saline. We never saw any effects of the D1A in single sessions in either monkey. Therefore, we did not continue bilateral injections in monkey I. It took about 11 min to complete the injection at 3 nl/s. The injection volume at each site was ~2 μl. Thus, the bilateral injections (four cannulae per hemisphere) were 8 μl per side, and the unilateral injection (two cannulae) was 4 μl. The volume of 2 μl was selected based on its estimated diffusion area from a previous report that specifically studied the relationship between injection volume and diffusion area (Myers, 1966). After the injection was completed, we waited 9 min to allow for the drugs to diffuse and the brain to recover from the mechanical effects of the injection. We then ran the animals until they stopped working. We only analyzed data from sessions in which the animals completed more than four blocks of the task (two blocks of random condition and two blocks of the fixed condition) after the injection.

Bayesian model of choice. To characterize the effects of the injections, we developed a Bayesian model that integrated information in the fixation stimulus with learned value information to predict choices. We estimated the information available from the fixation stimulus by fitting a model to data from the random condition. We then used a reinforcement learning algorithm to estimate the information available to the animal from previous outcomes in the fixed condition. This model estimated the information available from past reinforcement while controlling for information in the fixation stimulus. We then used Bayes rule to combine information from these two sources, taking into account reaction times, to predict behavior.

Value-related information derived from previous reinforcement. We fit a reinforcement learning model to the data from the fixed blocks to generate value estimates (Seo et al., 2012). The model was fit separately to data from the preinjection and postinjection periods of each session. The value, v_i, of each action, i, was updated after it was selected by the following:

$$v_i(t) = v_i(t-1) + \rho_f(r(t) - v_i(t-1)). \quad (1)$$

Rewards, $r(t)$, for correct actions were 1 and for incorrect actions were 0. This was the case for each movement, not just the movement that led to the juice reward. The variable ρ_f is the learning rate parameter. We used one value of ρ_f for positive feedback (i.e., correct actions) and one value for negative feedback (incorrect actions). When sequences switched across blocks, the action values were reset to 0. To estimate the log-likelihood, we first calculated choice probabilities using the following:

$$d_i(t) = \frac{e^{\beta v_i(t) - \gamma CB}}{\sum_{i=1}^{2} e^{\beta v_i(t) - \gamma CB}}. \quad (2)$$

The sum is over the two actions possible at each point in the sequence. Additionally, the animal's choice accuracy depended not only on past outcomes, but also on the color bias in the fixation stimulus (Seo et al., 2012). Therefore, to estimate the partial effect of the value of past outcomes on choice accuracy, we controlled for the information in the fixation stimulus by subtracting off a term related to the color bias (CB) times a free parameter, γ, which characterized the weight of the fixation stimulus in the choice process, in fixed trials. The inverse temperature, β, was also modeled as a free parameter. If β is small, then the animal is less likely to pick the higher-value target, whereas if β is large, the animal is more likely to pick the higher-value target for a fixed difference in target values. We calculated the log-likelihood (ll) of the animal's decision sequence as follows:

$$ll = -\sum_{t=1}^{T} \log(d_1(t)c_1(t) + (1-d_1(t))(1-c_1(t))). \quad (3)$$

The sum is over all decisions in a period, T. The variable $c_1(t)$ indicates the chosen action and has a value of 1 for action 1 and 0 for action 2. The two learning rate parameters ρ_f, the inverse temperature parameter β, and the color bias weight γ were optimized to minimize the log-likelihood using *fminsearch* in matlab. The minimization was done separately for each period (i.e., before and after injection) of each session. Most importantly for the overall model development, these value estimates do not characterize how performance develops over time. They only characterize the overall fraction correct in each condition.

The reinforcement learning model gave a fixed value estimate for each trial in each condition. However, it does not characterize how this information develops over time, and therefore it does not allow us to take into account the reaction time effects. To estimate the temporal evolution, we assumed the subjects were integrating a noisy internal representation of value. The value was given by a Gaussian distribution with a mean equal to the mean action value, estimated by the RL algorithm fit to the choice data, as outlined above. The variance, σ^2, of the Gaussian was a free parameter. The mean integrated value estimate at time t is given by $V(t) = tV_0$, where V_0 is estimated by the reinforcement learning model (Eq. 1). The variable $V(t)$ is the average sum of t draws from a distribution with a mean of V_0. The variance at time t is given by σ^2. Belief in one of the choices, for a particular value estimate, $V(t)$, is given by $P(Choice \mid V(t)) = \frac{P(V(t) \mid Choice)\,P(Choice)}{P(V(t))}$, where choice is one of the two options at each point in the sequence (e.g., left vs right or up vs down). The variances of the two choice distributions were matched, and the mean of the incorrect choice was set to the negative of the mean of the correct choice, i.e., $V(t) = -tV_0$ for the incorrect choice distribution. This value could also be taken as zero without substantially affecting the model results. We estimated the mean and variance of belief by sampling from the distribution of values (10,000 samples from $P(V(t) \mid \text{Choice})$) rather than computing them numerically directly off the distribution.

Immediately available perceptual information. The pixelating fixation stimulus can be characterized as draws from a binomial distribution. The distribution of the number of blue pixels for a given color bias, q, is given by the following:

$$P(N_{blue} = k \mid q) = \binom{N_{pixel}}{k} q^k (1-q)^{N_{pixel}-k}. \quad (4)$$

By conditioning on N_{blue} instead of q, the same equation gives a distribution over q. These distributions evolve over time because a fraction of the pixels, m, is updated on each frame. Thus, the subject's total number of pixels over j frames is $N_{pixel}(j) = N_{\frac{pixel}{frame}}(1 + (j-1)m)$. The belief that the stimulus is predominantly blue is then as follows:

$$P(blue \mid N_{blue}(t)) = \int_{0.5}^{1} P(N_{blue} = k \mid q)\,dq. \quad (5)$$

This ideal observer model deviated from monkey behavior in two ways. First, the model significantly outperformed the monkeys, likely because the animals cannot precisely count the number of blue and red pixels in each frame. Therefore, we assumed that the animals were effectively using only a subset of the pixels. We parameterized the number used by the animal, $N^{beh}_{pixel/frame}$, to predict the actual behavior of the animals in each color bias condition over time. This optimization was done for the behavioral performance in the random condition only. Data from the fixed condition were not used to determine $N^{beh}_{pixel/frame}$. The model was fit by minimizing the squared deviation between the model-estimated posterior probability, $P(\text{blue} \mid N_{\text{blue}}(t))$, and the animal's choice accuracy versus time in each color bias condition. Note that the animal's performance was at chance before ~200 ms, so we assumed a 200 ms lag time before information accumulation began. The model provided an estimate of the animal's performance based on information in the fixation stimulus versus time for each color bias condition. The behavior was consistent with

the animals extracting information from ~100 pixels per screen refresh. Thus, we used $m = 1$ and $N_{\frac{pixel}{frame}} = 100$ for the model estimates that are presented in Results. The model also tended to overpredict performance on the easy conditions and underpredict in the hard conditions. Therefore, we additionally assumed a sublinear function that mapped actual q into an effective q. It turned out that using $q = [0.5\ 0.55 + (0{:}0.025{:}0.075)]$ gave a good match to data. In other words, the effective q was approximately half the actual q for values above 0.55. Note that this fitting was done in the random condition but the final model was used to predict performance in the fixed condition. Therefore, this fitting was done to an independent data set.

Integration of value and perceptual information. In the fixed condition, information was available from both the fixation stimulus and reinforcement from previous trials. The information from each modality was independent. We assumed they were combined by the animal using Bayes rule, such that:

$$P(Choice \mid V(t), N_{blue}(t)) = \frac{P(V(t) \mid Choice)\ P(N_{blue}(t) \mid Choice)\, P(Choice)}{P(V(t), N_{blue}(t))}. \quad (6)$$

$P(N_{\mathrm{blue}}(t) \mid \mathrm{Choice})$ is given by Equation 5, because the prior is flat. The value of $N_{\mathrm{blue}}(t)$ we used was given by the reaction time of the animal in the corresponding condition. Specifically, the animals usually were faster as they learned the sequence. However, after injection of the D_2 receptor (D2R) antagonist, the animals did not speed up. Therefore, the model took this into account by assuming they were extracting more information from the fixation stimulus. We also optimized σ^2 for the RL model to minimize the squared error between the actual behavioral performance over trials after a new sequence was introduced for each color bias condition and the performance predicted by Equation 6. Thus, the parameters for information extraction versus time in each color bias condition were optimized in the random condition. This information was then combined with the model of value accumulation in the fixed condition, and σ^2 was separately optimized to predict behavioral performance before and after injection of the D2R in the fixed condition. This single parameter was used to characterize the decrease in choice accuracy before and after injection of the D2R in the fixed condition. We used an F test to compare the residual variance of a model that fit separate variances to preinjection and postinjection data with a model that fit a single variance to preinjection and postinjection data. These models differ by 1 df (i.e., two variances or one).

ANOVA analyses. The effects of pharmacological manipulations were analyzed using mixed-effects ANOVAs. The dependent variable was either reaction time or fraction correct. Independent variables included session, drug condition (D_2 vs saline or D_1 vs saline) nested under session, preinjection versus postinjection (period), color bias, and trial after switch (fixed condition only). Session was modeled as a random effect. All other factors were modeled as fixed effects. Except where indicated explicitly, reported results examined an interaction of preinjection versus postinjection and drug condition. This examines whether behavior changed from preinjection to postinjection in a drug-dependent way (i.e., differently for antagonists vs saline). Because preinjection data were included in the ANOVA, main effects of drug do not test for the effect of the drug.

Results

We trained three monkeys on a sequential decision-making task (Fig. 1). In the task, the animals had to acquire fixation and hold it for 500 ms. We then presented a stimulus at fixation that was composed of blue and red pixels. The probability of each pixel in the stimulus being either blue or red, the color bias, was fixed within a given trial and could be between 0.5 and 0.65. The color of a subset of pixels was updated on each screen refresh, according to the current color bias. The animal's task was to determine whether there were more blue or more red pixels and make a saccade to the peripheral target that matched the majority pixel color (Fig. 1*A*). After making three correct decisions in one of eight possible spatial sequences (Fig. 1*B*), the animals received a juice reward. The task was performed under two conditions, which we refer to as the random and fixed conditions (Fig. 1*C*). In the random condition, the spatial sequence of saccades varied from trial to trial. Therefore, the animals had to infer the majority pixel color to make the correct choice at each point in the sequence. In the fixed condition, the spatial sequence remained fixed until the animals executed the sequence without errors eight times. After eight correct trials, we switched to one of the other sequences (Fig. 1*B*). Therefore, in the fixed condition, in addition to the information available in the fixation stimulus, the animals could use outcomes from previous trials to determine which sequence was correct. In each session, the animals performed either one or two blocks of each condition at baseline (before injection). We then paused the task and injected saline, a D2A (eticlopride), or a D1A (SCH23390) into the dorsal caudate nucleus (Fig. 1*D*). After the injection, the animals completed at least two additional blocks of each condition.

The animal's choice accuracy in the random and fixed conditions was driven by the information provided in each condition. When we analyzed performance in the saline sessions combining preinjection and postinjection data, we found that in the random condition, the animal's performance improved with increasing color bias (Fig. 2*A*; $F_{(3,115)} = 374.9$, $p < 0.001$). Thus, as it became easier for the animals to infer the majority pixel color, they more often made the correct choice. In the fixed condition, as they learned from previous trials, their performance improved (Fig. 2*C*; $F_{(9,1144)} = 116.3$, $p < 0.001$), and they often made the correct choice even when there was no information in the fixation stimulus (Fig. 2*A*; CB50, fixed condition). In the fixed condition, there was also an effect of color bias on accuracy ($F_{(3,1144)} = 179.8$, $p < 0.001$) and an interaction between color bias and trials after switch ($F_{(27,1144)} = 4.94$, $p = 0.002$). The effect of color bias was driven primarily by the first few trials after switching sequences. We examine this in more detail below. The animals also made choices more quickly as they learned in the fixed condition (Fig. 2*D*; $F_{(9,1144)} = 17.3$, $p < 0.001$), but this effect did not depend on color bias ($F_{(3,1144)} = 0.2$, $p = 0.895$). Thus, the animals were able to use reinforcement from previous trials in the fixed condition and the information in the fixation stimulus in the random condition to make the correct decision.

Next, we examined performance before and after injection of a D_1 or D_2 antagonist and compared these drug sessions with saline sessions. The saline sessions controlled for time on task, satiation, mechanical displacement of the tissue, and other confounding factors. We found that when we injected the D2A into the dorsal striatum, the animals made more errors in the fixed condition, compared with saline sessions (Fig. 2*E*,*G*; period (preinjection vs postinjection) × drug; $F_{(1,41)} = 4.94$, $p = 0.032$). The difference was larger later in the block after switching to a new sequence (Fig. 2*G*; trials after switch × period × drug; $F_{(9,2467)} = 2.26$, $p = 0.016$). There was, however, no difference in the effect of drug across color bias levels (Fig. 2*E*; color bias × period × drug; $F_{(3,2467)} = 1.77$; $p = 0.151$).

We also found effects of the D_2 antagonist on reaction times in the fixed condition. Before injection, when the animals selected sequences in the fixed condition, their reaction times decreased with learning (Fig. 2*H*). However, after injection of the D_2 antagonist, this decrease in reaction time was smaller (Fig. 2*H*; period × drug; $F_{(1,34.9)} = 5.97$, $p = 0.020$). This effect also depended on trials after switch (trials after switch × period × drug; $F_{(9,2467)}$

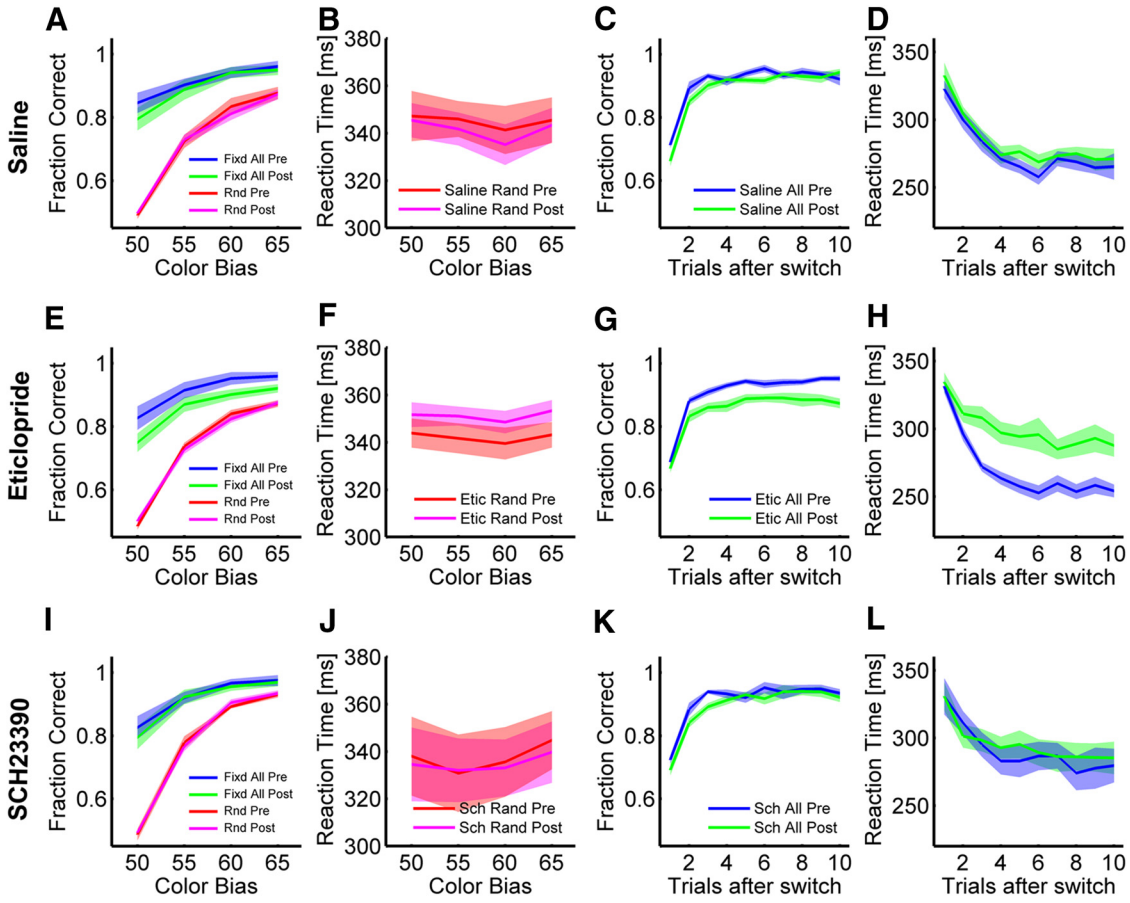

Figure 2. Effect of bilateral injection of saline, D2R antagonist eticlopride, and D1R antagonist SCH23390 in dStr. Heavy lines are means. Each shaded region is 1 SEM with *n* given by the number of sessions from all animals for each condition (see Materials and Methods). ***A***, Fraction of correct decisions as a function color bias for fixed and random conditions for the presaline injection (blue, red) and postsaline injection (green, pink) conditions. ***B***, Reaction times as a function of color bias for the random condition for the presaline injection (red) and postsaline injection (pink) conditions. ***C***, Fraction of correct decisions as a function of trials after switch for fixed conditions. ***D***, Reaction time as a function of trials after switch for fixed conditions. Data shown here are averaged across color bias conditions. ***E–H***, Same as ***A–D*** for the D_2 antagonist. ***I–L***, Same as ***A–D*** for performance in the D_1 antagonist condition.

$= 2.35, p = 0.012$) but not color bias (color bias × period × drug; $F_{(3,2467)} = 0.65, p = 0.583$).

In contrast to the fixed condition, there were no effects on choice accuracy in the random condition when we compared preinjection and postinjection data with saline sessions (Fig. 2*E*; period × drug; $F_{(1,230)} = 0.015, p = 0.903$). There was, however, an overall increase in the reaction time after injection (Fig. 2*F*; period × drug; $F_{(1,230)} = 12.8, p < 0.001$), but this effect did not interact with color bias (color bias × period × drug; $F_{(3,230)} = 0.12, p = 0.946$). The animals were also slower overall in the random condition than the fixed condition (Fig. 2, compare *F*, *H*). Therefore, the fact that animals did not decrease their reaction times with learning after injection of the D2A in the fixed condition, and the fact that there were no effects on choice accuracy in the random condition, suggests that they may have been relying more on information in the fixation stimulus after injection of the D2A in the fixed condition.

We next compared D1A sessions with saline sessions (Fig. 2*I–L*). When compared with saline sessions, there were no statistically significant effects in the fixed condition on choice accuracy or reaction time (period × drug; fraction correct: $F_{(1,1144)} = 0.00$, $p = 0.948$; reaction time: $F_{(1,1144)} = 0.00, p = 0.963$). There were also no significant interactions ($p > 0.234$). Similarly, in the random condition, there was no effect on choice accuracy after injecting the D_1 antagonist compared with injecting saline (Fig. 2*I*; period × drug; $F_{(1,118)} = 0.41, p = 0.521$). There was also no significant interaction with color bias (color bias × period × drug; $F_{(3,118)} = 0.21, p = 0.892$). There were also no effects of the D_1 antagonist on reaction times in the random condition (Fig. 2*J*; period × drug; $F_{(1,118)} = 0.02, p = 0.895$).

D_2 antagonist effect on integration of perceptual and value information

To characterize the effects of the D_2 antagonist injections in more detail, we developed a time-dependent Bayesian model of choice behavior in the fixed condition. The Bayesian model assumed that the animals combined or integrated immediately available information from the fixation stimulus with value estimates driven by the outcomes of previous trials to make their choice. This integration is consistent with previous analyses of behavior and neural data (Seo et al., 2012). Therefore, we had to estimate the information available to the animals in each condition from the perceptual stimulus and from past outcomes. More specifically, when the animals could only use the perceptual stimulus in the random condition, how accurate would they be at each color bias level? This can be modeled directly from the choice data in the random condition. Also, if the animals could only use reinforcement of past outcomes, how accurate would they be? This can be modeled using the data from the fixed condition, taking into account how much information is available in the fixation

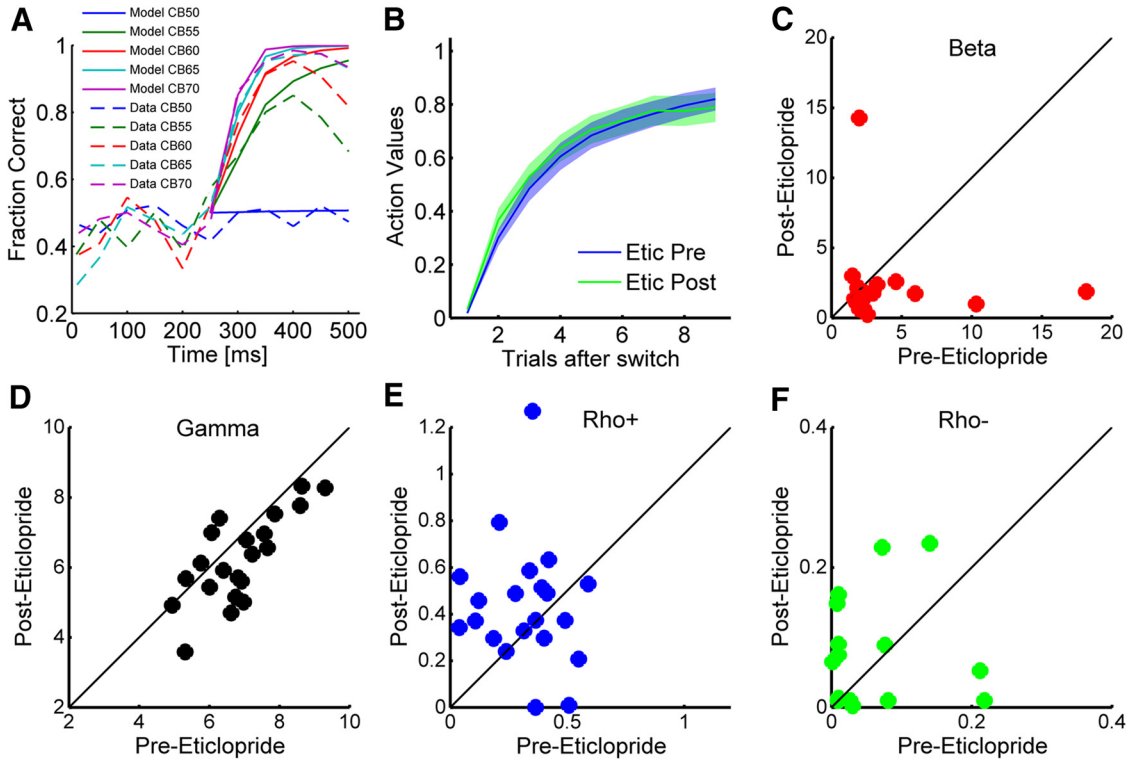

Figure 3. Model fits for perceptual inference and reinforcement learning. ***A***, Fraction correct in the random condition, for each color bias (CB), as a function of time after stimulus presentation (time 0). Bin size, 50 ms. Dotted lines are data, and solid lines are model fits. ***B***, Before (Etic Pre) and after (Etic Post) action values in the fixed condition as a function of trials after switch. ***C***, Inverse temperature (β parameter) before and after injection for the D_2 antagonist. One point was excluded for plotting (before, 2.21; after, 175.33). ***D***, Gamma parameter representing weight of color bias in the fixed condition. ***E***, Learning rate parameter for positive feedback. ***F***, Learning rate parameter for negative feedback.

stimulus and how much information is available from past outcomes. Once these were estimated, we could use Bayes rule to integrate them and generate an estimate of choice accuracy as a function of time.

We first estimated the information available to the animals from the stimulus as a function of time. In the random condition, performance improved with time after ~200 ms (Fig. 3*A*) and it also improved with increasing color bias. When the animals made their choice before ~200 ms, they were at chance. After 200 ms, there was a rapid increase in performance that peaked at ~400 ms. After 400 ms, performance began to decrease, probably because of lapses in attention (Drugowitsch et al., 2012). These accuracy versus reaction time curves were similar before and after injection in the D2A sessions for reaction times <500 ms (data not shown). The differences in reaction time in Figure 2*B* were driven by an increase in relatively long reaction times (>500 ms). We fit a model to this reaction time data from the random condition. This characterized the amount of information that the animals extracted in each color bias condition, as a function of time (Fig. 3*A*).

Next, we characterized the information available to the animals from the outcomes of previous trials using a RL model fit to the choice data in the fixed condition. Although the task in the fixed condition is, in principle, deterministic, the animal's choice behavior was not. Therefore, their choice accuracy can be modeled using a δ learning rule RL algorithm. Interpreted another way, the algorithm fits a lag-1 autoregressive logistic regression model to the behavioral data that predicts choice accuracy, and it is this prediction that we used in the Bayesian model. The model was fit separately to each session and separately within each session to the preinjection and postinjection data. We examined the parameters of the RL model fits to see if they varied with injection of the D_2 antagonist. We first examined the inverse temperature, which characterizes how consistently the animals chose the correct option after learning. We found that the inverse temperature was larger before than after the injection of the D2A (Fig. 3*C*; Mann–Whitney *U* test, $p = 0.011$). We also examined the weight given to the color bias in the fixed condition (Fig. 3*D*), before and after drug, and found no significant effect ($p > 0.05$). When we examined the changes in the learning rate parameters, there were no differences before and after injection of the D2A (Fig. 3*E*,*F*; $p > 0.05$). Because the learning rates were consistent before and after injection, the value estimates were also consistent before and after injection (Fig. 3*B*; $p > 0.05$). Therefore, the change in choice accuracy could be characterized as increased noise in the choice mechanism that converted value estimates into choices. We used the Bayesian model to further characterize this, while also taking into account the reaction times, which differed before and after injection, and performance in the random condition. This was necessary because the animals slowed down in the fixed condition after injection of the D2A, and this slowing may have reflected an increased reliance on the fixation stimulus. Thus, their choice accuracy, on average, dropped less than expected because they were able to compensate with perceptual information, when it was available.

In the fixed condition, we assumed the animals integrated a noisy representation of the static value estimate, using a drift diffusion-like mechanism (Ratcliff, 1978). A drift diffusion mechanism computes the sum of random samples from a distribution with a fixed mean and SD. In each time step, a sample is drawn from the distribution and added to the sum from the previous time step. Normally in choice tasks, there are two pos-

sible distributions with means that are symmetric around zero, and the goal is to figure out which distribution is being sampled from. Over time, the sums will diverge, and the rate at which they diverge depends on the means (i.e., how far each distribution is from zero) and the SD. Interpreted in neural terms, the idea is that one is integrating (i.e., summing) the output of an upstream population code that is a noisy representation. In our example, this would be a noisy representation of the value of the two saccades, at each point in time. Therefore, in our model, the mean value that was integrated (i.e., the mean of the distribution from which samples were drawn) came from the reinforcement learning algorithm (Fig. 3*B*). The variance of this value estimate in the drift diffusion process (i.e., the variance of the distribution from which samples were drawn) was the single free parameter used to fit the model to the choice data in the fixed condition. The variance of the value estimate was parameterized, and the model was fit to the choice behavior separately before and after injection of the D2R antagonist. We found that the choice behavior was characterized effectively by the model (Fig. 4; Spearman correlation, $p < 0.001$ for both preinjection and postinjection fits). Furthermore, we found before injection, the estimated variance of the noise on value integration was 1.81, whereas after injection, it was 5.48 (Fig. 4*A*, inset). Therefore, the value integration process was noisier after injection of a D2A. To test this statistically, we compared a model that fit separate noise estimates to predata and postdata with a model that had a single noise term (single noise term of 3.12) and found that the model with separate noise terms for predata and postdata fit significantly better ($F_{(1,80)} = 56.1$, $p < 0.001$). Thus, the model suggests that injection of the D2A into the dStr increased noise in the value representation that was integrated by choice mechanisms.

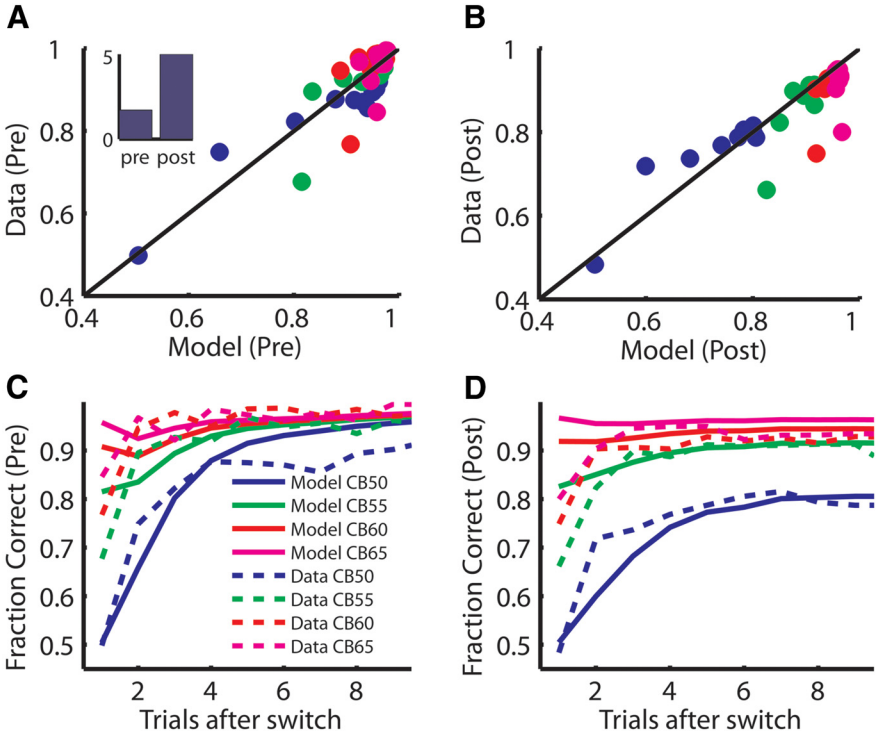

Figure 4. Behavioral performance and model prediction. Data plotted are observed and predicted fraction corrected across trials after switch. Color indicates the color bias level, as in ***C***. All data are from D2A sessions. Pre–post injections are indicated on the corresponding plots. ***A***, Scatter plot of choice accuracy before injection for the model and data. The inset shows the estimated variance of the noise on value integration. ***B***, Scatter plot of choice accuracy after injection for the model and data. ***C***, ***D***, Same data as plotted in ***A*** and ***B***, plotted as a function of trials after switch, broken out by the color bias level.

Discussion

We found that neither D_1 nor D_2 antagonists injected into the dorsal striatum affected choice accuracy in the random condition. However, injection of the D_2 antagonist decreased choice accuracy in the fixed condition. In addition, animals normally responded more quickly in the fixed condition as they learned the correct sequence. However, after injection of the D_2 antagonist, there was less decrease in reaction time, consistent with hypotheses that the basal ganglia are important for response vigor (Turner and Desmurget, 2010). These results suggest that the animals were able to use the information in the fixation stimulus to drive their decisions in the fixed block, but they were less able to use past reinforcement. We examined this by fitting a Bayesian model, which characterized the integration of immediately available perceptual information and value-related information. When we fit the model to the data in the D_2 sessions, we found that after injection of the D_2 antagonist, there was more noise in the value representation. This suggests that the D_2 antagonist injections are affecting the population code for value in the dorsal striatum.

Our data are consistent with previous studies that have shown a role for D_2 receptors in choices driven by past reinforcement. Studies in healthy human subjects have shown that systemic injections of D_2 receptor antagonists (Pessiglione et al., 2006; Eisenegger et al., 2014) can affect choice accuracy. The study by Pessiglione et al. (2006) found that subjects who received L-dopa earned more money than subjects that received haloperidol. There were no effects, however, in a loss condition and no effects on learning. Using a similar task, Eisenegger et al., 2014 found that subjects receiving sulpiride, a selective D_2/D_3 antagonist, who also had a genetically characterized decreased density of D_2 receptors, had decreased choice accuracy when choices were driven by rewards. Again, there were no effects on choices driven by loss and no effects on learning. Rather, the manipulation affected the consistency with which subjects chose the option that had been rewarded more often, after learning had plateaued. Furthermore, studies in macaques have shown that the availability of dopamine D_2 receptors in the dorsal striatum affects the consistency with which subjects choose an option after it has just been rewarded, without affecting choices after negative feedback (Groman et al., 2011). Studies in mice lacking either D_1 or D_2 receptors have also shown effects of D_2 but not D_1 receptors on learning (Kwak et al., 2014). Thus, these studies consistently found that lower D_2 receptor function decreases the consistency with which subjects choose the correct option, when correct choices were driven by previous rewards.

Our study extends these results in several ways. First, we found effects with local, causal manipulations of D_2 receptor function in the caudate. Pessiglione et al. (2006) and Eisenegger et al. (2014)

manipulated dopamine systemically, and Groman et al. (2011) used PET imaging, which is correlative and cannot dissociate receptor availability from dopamine concentration. Second, we found no effects of D_1 or D_2 antagonists on selecting actions on the basis of perceptual inference, and therefore the effects of the D_2 antagonist were specific to learning from past reinforcement. This suggests a limited role for striatal dopamine in perceptual inference tasks. Correlates of perceptual inference tasks have been found in the caudate, and micro-stimulation in the caudate can affect choice performance (Ding and Gold, 2013). However, micro-stimulation in the striatum has complex effects on the circuitry, leading, for example, to substantial dopamine release (Schluter et al., 2014). Another hypothesis suggests that the basal ganglia/colliculus circuit may be important for setting a threshold on evidence integration, with threshold crossing implying choice commitment (Lo and Wang, 2006). Correlates of a decision to commit to a choice have also been seen in cortical networks, studied with fMRI (Furl and Averbeck, 2011; Costa and Averbeck, 2015), which suggests that the basal ganglia/colliculus circuit may play a specific role in only eye-movement tasks with fast reaction times, if it plays a role at all. In addition, the striatum does not appear to reflect a fixed bound in perceptual decision-making tasks (Ding and Gold, 2010, 2013), and we have found that choice information is coded in lPFC before it is coded in the dStr in our task (Seo et al., 2012). Therefore, these studies have not identified a clear role for the striatum in perceptual inference.

Our study also extends previous results by implicating the D_2 system in the dorsal striatum in associating rewards with actions, whereas previous studies have examined the association of rewards and visual cues. The ventral striatum may be more important for associating objects with rewards, whereas the dorsal striatum may be more important for associating actions with rewards. This hypothesis is consistent with the finding that when images cue go or no-go actions, the dorsal striatum is more involved (Guitart-Masip et al., 2012, 2014). One the other hand, reward prediction errors correlate with activation in the ventral striatum (Pessiglione et al., 2006) when rewards are being associated with objects. Also consistent with this, the dorsal striatum receives a strong anatomical input from the frontal eye field (FEF) and caudal area 46 (Haber et al., 2006; Averbeck et al., 2014), both of which are rich in eye movements signals, whereas the ventral striatum receives more inputs from areas that process visual information, including orbital frontal cortex (Haber and Knutson, 2010). Our current results are also consistent with our previous finding that the dorsal striatum represented the value of actions earlier and more strongly than the lPFC (Seo et al., 2012). If one assumes that this representation is integrated or read out in some way to generate a choice, then degrading this value representation should affect choice accuracy because the population code for value will be noisier (Averbeck et al., 2006). There are other possible interpretations for our results, however. It is possible that the animals have intact value representations but that they fail to convert these representations to correct choices. The D_2 antagonists may make the animals more distractible, or they may be less motivated and therefore more prone to errors.

Theories of the basal ganglia (Frank, 2005) have predicted a more important role for D_1 receptors than D_2 receptors in learning to associate actions with rewards, because D_1 receptors are more common in the direct pathway and because the direct pathway is thought to be more important for action and the indirect pathway for withholding action. Our data and the other studies cited above do not appear to directly support this. However, we do not see any effects on learning, only on making choices that are consistent with learned values. Additionally, most of the learning in our task is driven by negative outcomes, and the model described by Frank (2005) suggests that learning from negative feedback should be driven by the D_2 system if one is learning to not make an action (Piray, 2011). In this sense, our results are consistent with the prediction of the model of Frank (2005). Furthermore, the components of the striatal microcircuit that are affected by our D_2 antagonist injections are not yet clear. D_2 receptors in the striatum are located preferentially on medium spiny neurons in the indirect pathway and also on cholinergic interneurons (Yan et al., 1997; Alcantara et al., 2003). Thus, the effects of the D_2 antagonists may be mediated directly by effects on medium spiny neurons, indirectly by effects on cholinergic interneurons, or by a combination of these. Finally, although we did not find effects of D_1 antagonists in our study, previous studies have shown that activating D_1-containing medium spiny neurons in the striatum when a lever is pressed can lead to preference for that lever, whereas activating D_2-containing medium spiny neurons when a lever is pressed leads to avoidance of that lever (Kravitz et al., 2012). However, it should be noted that in this study, systemic D_1 and D_2 antagonists had no effect on the results, whereas we find results of local injections of D_2 antagonists. Therefore, the behavior studied by Kravitz et al. (2012) appears to be interacting differently with the striatal system.

Conclusion

We have found that blocking D_2 receptors locally in the dorsal striatum affects choices driven by past reinforcement, without affecting learning. There were no effects of the D_2 antagonist on choice accuracy in the random condition and no effects of D_1 receptor antagonists in either condition. Furthermore, when we fit a model to behavioral performance before and after injection of the D_2 antagonist in the fixed condition, the model suggested that the decreased performance is driven by increased noise in the population representation of value that is being integrated by the choice mechanism. Therefore, we believe that manipulation of the D_2 system affects the population code for value in the dorsal striatum.

References

Alcantara AA, Chen V, Herring BE, Mendenhall JM, Berlanga ML (2003) Localization of dopamine D2 receptors on cholinergic interneurons of the dorsal striatum and nucleus accumbens of the rat. Brain Res 986:22–29. CrossRef Medline

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358–366. CrossRef Medline

Averbeck BB, Lehman J, Jacobson M, Haber SN (2014) Estimates of projection overlap and zones of convergence within frontal-striatal circuits. J Neurosci 34:9497–9505. CrossRef Medline

Barraclough DJ, Conroy ML, Lee D (2004) Prefrontal cortex and decision making in a mixed-strategy game. Nat Neurosci 7:404–410. CrossRef Medline

Cockburn J, Collins AG, Frank MJ (2014) A reinforcement learning mechanism responsible for the valuation of free choice. Neuron 83:551–557. CrossRef Medline

Costa VD, Averbeck BB (2015) Frontal-parietal and limbic-striatal activity underlies information sampling in the best choice problem. Cereb Cortex 25:972–982. CrossRef Medline

Ding L, Gold JI (2010) Caudate encodes multiple computations for perceptual decisions. J Neurosci 30:15747–15759. CrossRef Medline

Ding L, Gold JI (2013) The basal ganglia's contributions to perceptual decision making. Neuron 79:640–649. CrossRef Medline

Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A (2012) The cost of accumulating evidence in perceptual decision making. J Neurosci 32:3612–3628. CrossRef Medline

Eisenegger C, Naef M, Linssen A, Clark L, Gandamaneni PK, Muller U, Robbins TW (2014) Role of dopamine D2 receptors in human reinforcement learning. Neuropsychopharmacology 39:2366–2375. CrossRef Medline

Fetsch CR, DeAngelis GC, Angelaki DE (2013) Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. Nat Rev Neurosci 14:429–442. 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

Furl N, Averbeck BB (2011) Parietal cortex and insula relate to evidence seeking relevant to reward-related decisions. J Neurosci 31:17572–17582. CrossRef Medline

Graybiel AM (2008) Habits, rituals, and the evaluative brain. Annu Rev Neurosci 31:359–387. CrossRef Medline

Groman SM, Lee B, London ED, Mandelkern MA, James AS, Feiler K, Rivera R, Dahlbom M, Sossi V, Vandervoort E, Jentsch JD (2011) Dorsal striatal D2-like receptor availability covaries with sensitivity to positive reinforcement during discrimination learning. J Neurosci 31:7291–7299. CrossRef Medline

Gu Y, Angelaki DE, Deangelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. Nat Neurosci 11:1201–1210. CrossRef Medline

Guitart-Masip M, Chowdhury R, Sharot T, Dayan P, Duzel E, Dolan RJ (2012) Action controls dopaminergic enhancement of reward representations. Proc Natl Acad Sci U S A 109:7511–7516. CrossRef Medline

Guitart-Masip M, Duzel E, Dolan R, Dayan P (2014) Action versus valence in decision making. Trends Cogn Sci 18:194–202. CrossRef Medline

Haber SN, Fudge JL (1997) The primate substantia nigra and VTA: integrative circuitry and function. Crit Rev Neurobiol 11:323–342. CrossRef Medline

Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35:4–26. 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

Kravitz AV, Tye LD, Kreitzer AC (2012) Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat Neurosci 15:816–818. CrossRef Medline

Kwak S, Huh N, Seo JS, Lee JE, Han PL, Jung MW (2014) Role of dopamine D2 receptors in optimizing choice strategy in a dynamic and uncertain environment. Front Behav Neurosci 8:368. CrossRef Medline

Lo CC, Wang XJ (2006) Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nat Neurosci 9:956–963. CrossRef Medline

Lovejoy LP, Krauzlis RJ (2010) Inactivation of primate superior colliculus impairs covert selection of signals for perceptual judgments. Nat Neurosci 13:261–266. CrossRef Medline

Myers RD (1966) Injection of solutions into cerebral tissue: relation between volume and diffusion. Physiol Behav 1:171–174. CrossRef

Newsome WT, Britten KH, Movshon JA (1989) Neuronal correlates of a perceptual decision. Nature 341:52–54. CrossRef 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

Piray P (2011) The role of dorsal striatal D2-like receptors in reversal learning: a reinforcement learning viewpoint. J Neurosci 31:14049–14050. CrossRef Medline

Ratcliff R (1978) A theory of memory retrieval. Psychol Rev 85:59–108. CrossRef

Ratcliff R, Cherian A, Segraves M (2003) A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. J Neurophysiol 90:1392–1407. CrossRef Medline

Robinson S, Sandstrom SM, Denenberg VH, Palmiter RD (2005) Distinguishing whether dopamine regulates liking, wanting, and/or learning about rewards. Behav Neurosci 119:5–15. CrossRef Medline

Rorie AE, Gao J, McClelland JL, Newsome WT (2010) Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. PLoS One 5:e9308. CrossRef Medline

Rudebeck PH, Mitz AR, Chacko RV, Murray EA (2013) Effects of amygdala lesions on reward-value coding in orbital and medial prefrontal cortex. Neuron 80:1519–1531. 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

Schluter EW, Mitz AR, Cheer JF, Averbeck BB (2014) Real-time dopamine measurement in awake monkeys. PLoS One 9:e98692. CrossRef Medline

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599. CrossRef Medline

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

Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916–1936. Medline

Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16:966–973. CrossRef Medline

Turner RS, Desmurget M (2010) Basal ganglia contributions to motor control: a vigorous tutor. Curr Opin Neurobiol 20:704–716. CrossRef Medline

Wallis JD (2012) Cross-species studies of orbitofrontal cortex and value-based decision-making. Nat Neurosci 15:13–19. CrossRef Medline

Yan Z, Song WJ, Surmeier J (1997) D2 dopamine receptors reduce N-type Ca2+ currents in rat neostriatal cholinergic interneurons through a membrane-delimited, protein-kinase-C-insensitive pathway. J Neurophysiol 77:1003–1015. Medline