

The Dorsomedial Striatum Reflects Response Bias during Learning

Eyal Y. Kimchi^{1,2} and Mark Laubach^{1,3}

¹The John B. Pierce Laboratory, New Haven, Connecticut 06519, and ²Interdepartmental Neuroscience Program and ³Department of Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06520

Previous studies have established that neurons in the dorsomedial striatum track the behavioral significance of external stimuli, are sensitive to contingencies between actions and outcomes, and show rapid flexibility in representing task-related information. Here, we describe how neural activity in the dorsomedial striatum changes during the initial acquisition of a Go/NoGo task and during an initial reversal of stimulus-response contingencies. Rats made nosepoke responses over delay periods and then received one of two acoustic stimuli. Liquid rewards were delivered after one stimulus (S+) if the rats made a Go response (entering a reward port on the opposite wall of the chamber). If a Go response was made to other stimulus (S−), rats experienced a timeout. On 10% of trials, no stimulus was presented. These trials were used to assess response bias, the animals' tendency to collect reward independent of the stimulus. Response bias increased during the reversal, corresponding to the animals' uncertainty about the stimulus-response contingencies. Most task-modulated neurons fired during the response at the end of the delay period. The fraction of response-modulated neurons was correlated with response bias and neural activity was sensitive to the behavioral response made on the previous trial. During initial task acquisition and initial reversal learning, there was a remarkable change in the percentages of neurons that fired in relation to the task events, especially during withdrawal from the nosepoke aperture. These results suggest that changes in task-related activity in the dorsomedial striatum during learning are driven by the animal's bias to collect rewards.

Introduction

Neurons in the dorsomedial striatum (or caudate nucleus) represent contingencies between stimuli, actions, and outcomes in a highly flexible manner (Balleine et al., 2007) and are thought to be crucial for rapidly switching plans of actions based on current task conditions (Kawagoe et al., 1998; Itoh et al., 2003; Pasupathy and Miller, 2005; Samejima et al., 2005; Watanabe and Hikosaka, 2005; Williams and Eskandar, 2006; Ding and Hikosaka, 2006; Kimchi and Laubach, 2009). Previous studies of the dorsomedial striatum were done in well trained subjects that had extensive experience with changes in task contingencies. While several studies have investigated changes in striatal activity during the initial learning [T-maze: Jog et al. (1999); skill learning: Costa et al. (2004); operant tasks: Carelli et al. (1997), Tang et al. (2007), Teagarden and Rebec (2007), Kimchi et al. (2009)], no study has examined neural correlates of initial Go/NoGo learning (e.g., the first time an animal learns to collect reward following one stimulus and not another) or initial reversal learning (e.g., the first

time that an animal experiences a reversal in task contingencies). As a result, it is unclear whether the same types of rapid changes in task-related neural activity are found in naive and experienced subjects. Resolving this issue is important to understand the role of the striatum in normal daily activities, especially those that occur without extensive prior experience.

Lesions of the dorsomedial striatum impair initial reversal learning, but not discrimination learning (Kirkby, 1969; Pisa and Cyr, 1990; Adams et al., 2001; Ragozzino et al., 2002; Palencia and Ragozzino, 2004; Featherstone and McDonald, 2005; Broadbent et al., 2007; Ragozzino, 2007). Based on these studies, we hypothesized that neurons in the dorsomedial striatum would show major changes in task-related activity (such as selective firing during Go or NoGo responding, as in the study by Kimchi and Laubach, 2009) during reversal learning, but not during Go/NoGo learning. To examine this issue, a stepwise approach was used for behavioral training that allowed us to identify changes in neural activity associated with sustained responding over a delay period, with responding to a rewarded stimulus, with learning to selectively respond to one stimulus and not another, and to reversing stimulus-response contingencies. We trained rats to perform a simple reaction time task and then implanted arrays of electrodes into the dorsomedial striatum. Neural activity was recorded in one session with the simple reaction time task to assess baseline levels of neural activity related to the basic task procedure. Next, rats were trained to discriminate between two stimuli, one from the simple reaction time task that was associated with reward and a new stimulus that was associated with nonreward. Last, and most importantly, we studied how dorsomedial striatal activity

Received Aug. 18, 2009; revised Sept. 29, 2009; accepted Oct. 14, 2009.

This research was supported by U.S. National Institutes of Health Medical Scientist Training Program Grant 5T32GM07205 (E.Y.K.) and funds from the John B. Pierce Laboratory (M.L.). E.Y.K. and M.L. conceived and designed the experiments, analyzed the data, and wrote the paper. E.Y.K. performed the experiments. We thank the Instruments Shop at the John B. Pierce Laboratory for technical support and Hyoung Seo for helpful comments on a draft of this manuscript.

The authors declare that no competing interests exist.

Correspondence should be addressed to Dr. Mark Laubach, The John B. Pierce Laboratory, 290 Congress Avenue, New Haven, CT 06519. E-mail: mlaubach@jbpierce.org.

DOI:10.1523/JNEUROSCI.4060-09.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/2914891-12\$15.00/0

changed during the rats' first experience with a reversal of the stimulus–reward contingencies. Throughout training, animals experienced catch trials (no stimulus presented) on 10% of trials. These trials allowed us to measure response bias, the tendency to respond independent of the stimulus (Macmillan and Creelman, 2005).

Materials and Methods

General procedure. Briefly, rats were trained to make nosepoke responses to produce a stimulus signaling reward. They were then implanted with arrays of electrodes in the dorsomedial striatum. Neural activity was recorded while rats performed a simple reaction time task. This was followed by recordings during discrimination learning in a Go/NoGo paradigm. After successful discrimination learning, neural recordings were continued as rats learned a reversal of the previous stimulus–reward contingencies.

Subjects. Nine male Long–Evans rats (300–450 g, Charles–River Laboratories) were individually housed and kept on a 12 h light dark cycle with lights on at 7:00 A.M. Facilities were temperature- and humidity-controlled. Rats were given several days to acclimate to the facilities, during which they were handled and had free access to food and water. All animal procedures were approved by the Institutional Animal Care and Use Committee at the John B. Pierce Laboratory (New Haven, CT).

Fluid restriction. To motivate behavior, rats had regulated access to water for 18 h before behavioral sessions. Rats earned about half of their daily water during the behavioral sessions, and the remaining volume of the average water consumption was provided in the home cage, 1–2 h after the behavioral sessions. Rats were given free access to water at least 1 d every week. Food was always available *ad libitum*. Weights were monitored to ensure that rats maintained 90–95% of their free water body weight throughout the experiment.

Behavioral chamber. Initial training took place in a standard operant chamber housed within a sound-attenuating box (ENV-008, Med Associates). The floors of the boxes were rectangular (30 cm wide by 24 cm deep). Behavioral devices were located on the narrow walls. One wall had a central nosepoke hole with a photobeam sensor to identify nosepoke entry (ENV-114BM, Med Associates). The opposite wall had a metal spout connected to a water pump (PHM-100, Med Associates). Spout contact was detected with a custom-built circuit that measured electrical resistance between the spout and the floor (John B. Pierce Laboratory Instruments Shop). The volume of water delivered was calibrated by adjusting the duration of pump activation (1 s = 40 μ l). Above the spout was a speaker (ES1, TDT Technologies) for presentation of acoustic stimuli generated by a digital signal processor (RP2.1, TDT Technologies) and an electrostatic speaker driver (ED1, TDT Technologies). Above the speaker was a 28 V, 100 mA incandescent light (ENV-215M, Med Associates). A fan was located on the inside of each sound attenuating box to provide constant background noise and ventilation. Behavioral protocols were controlled using MedPC software (Med Associates).

Tone conditioning. After handling and fluid restriction, rats experienced tone conditioning. Rats were placed in the operant chamber with the nosepoke hole closed. The houselight was turned on at the beginning of the behavioral session and remained on until the end. On the first day, an 8 kHz tone (5 s, ~60 dB) was presented approximately every 30 s, and 120 μ l of water was delivered at the spout. Starting on the second day, water delivery occurred only if the rats licked the spout during the tone. To diminish the rate of spontaneous licking, spout contacts in the four seconds before stimulus onset delayed stimulus presentation. On one-third of trials, selected pseudo-randomly, no tone was presented and we measured the probability that rats would lick at the metal spout in the

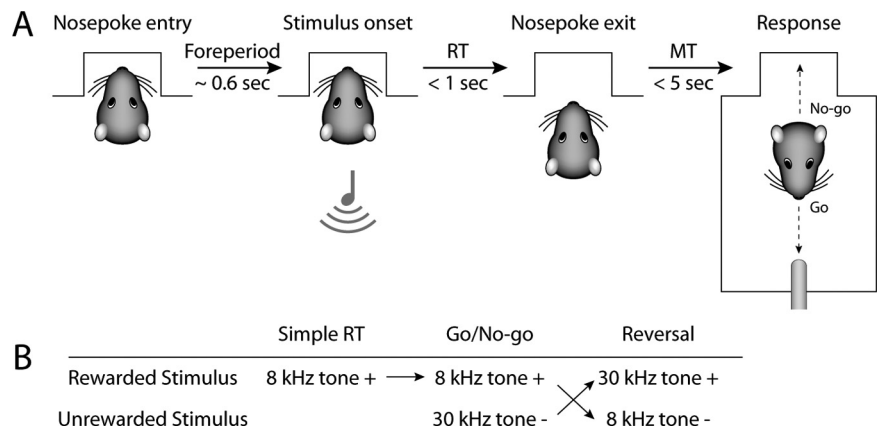


Figure 1. Behavioral paradigm. **A**, Trials began when rats inserted their snout into the nosepoke aperture. If they maintained this posture for 0.6 s (delay period), an auditory stimulus was presented. Rats then had to withdraw from the nosepoke within 1 s from stimulus onset (RT) and cross the chamber to collect a water reward within 5 s (MT). This was called a Go response. If rats did not attempt to contact the spout promptly this was called a No-go response. On No-go responses, rats typically initiated a new trial by reinserting their snout in the nosepoke aperture. **B**, The first stage of training used a simple reaction time task, in which an 8 kHz tone was used as the rewarded stimulus (8k+). Go responses to this stimulus were always rewarded. Next, animals experienced discrimination training, and were presented with an equal number of trials using 30 kHz tones as an unrewarded stimulus (30k-). Go responses to the unrewarded stimulus led to a timeout. In reversal training, the values of the two tones reversed, i.e., the 30 kHz tone was now rewarded (30k+) whereas the 8 kHz tone was now unrewarded (8k-). In all sessions, 10% of trials had no stimulus and served as catch trials to assess the rats' bias to respond for reward independent of the stimulus.

absence of the stimulus. Once rats licked consistently more on tone trials than on catch trials ($\geq 15\%$ difference on at least two consecutive days, or $\geq 10\%$ difference on at least three consecutive days, range 4–8 sessions), they were advanced to nosepoke training. Sessions lasted ~60 min.

Nosepoke training. The nosepoke aperture was unveiled and rats now initiated presentations of the 8 kHz tone by inserting their snout into the nosepoke aperture (Fig. 1). Rats had to maintain this position for a set time (delay period) to receive a stimulus. The delay period was increased from 50 to 600 ms both within sessions and across the first few days of nosepoke training. If rats withdrew from the nosepoke early, no stimuli were presented and responses were not reinforced. No timeout or other form of punishment was used at this stage of training. Following presentation of the 8 kHz tone, rats had to withdraw from the nosepoke within a set time (shortened over sessions to 1 s) and cross to the opposite wall of the chamber to make contact with the spout within a set time (shortened over sessions to 5 s) to receive water (80 μ l). We use the term "Go response" to describe a prompt nosepoke withdrawal followed by contact with the water spout within the requisite time. The term "No-go response" is used to describe entry into the nosepoke before any spout contact. On all stimulus trials, stimuli remained on until the time window for a response had elapsed or a new trial was initiated. Rats were trained until they responded to 80% of stimuli on at least two consecutive days (10–19 sessions). Trials were self-paced and sessions lasted ~90 min.

Transfer to recording chamber. Rats were then transferred to the chamber in which physiological recordings would take place after electrode implantation. The same nosepoke protocol described above was used for training. The recording chamber was similar to the standard operant box but modified for electrical recordings (custom built by Med Associates). All walls, floor bars, and behavioral devices were made of acrylic plastic and the long walls sloped diagonally outward. The fluid spout was made of plastic and responses were detected by an infrared photobeam crossing in front of it (John B. Pierce Laboratory Instruments Shop). Pump onset was delayed by a brief amount of time following the break of the photobeam (250 ms for 6 rats, 150 ms for 3 rats). Auditory stimuli were presented using the same equipment as previously, but stimuli lasted only up to 500 ms. (Stimuli were shorter than 500 ms if the rats initiated a new trial by reinserting their snout into the nosepoke before the stimulus time had elapsed. This was a rare type of event, but did occur on a small fraction of trials. In these cases, the stimulus was extinguished so that the delay period remained silent until presentation of the next stimulus.) The speaker was wrapped in additional copper foil shielding. Be-

havioral protocols in the recording chamber were controlled using a digital input-output card (PCI-DIO-96, National Instruments) and the Matlab Data Acquisition Toolbox (MathWorks), as well as the freely available Psychophysics Toolbox (Brainard, 1997). The chamber was placed on a steel plate within a copper wire Faraday cage. Rat behavior was monitored using an infrared camera and videotaped for offline analysis. Rats were trained to ensure that behavior was stable in the recording chamber (3–9 sessions), at which point they were given a week of unregulated access to water and food in preparation for surgery.

Surgery. Anesthesia was initiated with 4% vaporized halothane. Rats were then injected intraperitoneally with ketamine (100 mg/kg) and xylazine (10 mg/kg). Smaller supplementary doses of ketamine were given as needed to suppress responses to toe pinches. The scalp was shaved and the rat was placed in a stereotaxic apparatus using blunt 45° ear bars to prevent eardrum rupture. Eyes were covered with ophthalmic antibiotic ointment to prevent desiccation. The scalp was disinfected with iodine, injected subcutaneously with lidocaine (0.3 ml), and incised and retracted to expose the skull surface. Lambda and Bregma were leveled and bilateral craniotomies were made for implantation of multielectrode arrays. One skull screw was inserted anterior and posterior to each craniotomy (four total).

Electrode implantation. Multielectrode arrays were composed of 50 μm stainless steel wires coated with Teflon, arranged in 2×8 configurations with 250 μm spacing between wires and mounted on Microtech (NB Labs) or Omnetics connectors (Neuroline). Neural activity was monitored as electrodes were lowered into the dorsomedial striatum. One array was placed in the dorsomedial striatum of each hemisphere (centered on 0.2 mm AP, ± 2.5 mm ML, -4.6 mm DV from brain surface).

Recovery. After electrode implantation, the craniotomy was covered with cyanoacrylate (Slo-Zap) and cyanoacrylate accelerator (Zip-Kicker). The skull was covered with methyl methacrylate (AM Systems). Wound margins were dabbed with vectromycin antibiotic ointment. Postsurgical analgesia was provided by subcutaneous injections of buprenorphine over the first 2 d (0.03 mg/kg) and oral enrofloxacin in drinking water for a week (600 mg of enrofloxacin diluted in 500 ml of water). After recovery from surgery, rats had regulated access to fluid to reinstate behavioral training.

Handling before recording sessions. Rats were briefly anesthetized with 2% halothane to connect cables to the implanted connectors. Rats recovered spontaneous motor activity within a few minutes and were placed in a separate acrylic chamber for at least half an hour before behavioral recordings. During this time, spike sorting was performed and spontaneous recordings of neural activity were made, including recordings of wideband signals from each electrode (at 20 kHz). Once this was complete, rats were moved into the behavioral chamber and connected to a 32 channel slip contact commutator (Plexon) mounted in the center of the ceiling of the chamber. This device connected cables from the implanted probes to the recording system and allowed relatively free movement.

Simple reaction time task. Neural activity was recorded from the medial striatum of each rat first while performing the simple reaction time task (Fig. 1). Each session started with a delay period of 100 ms. If rats made sustained responses for two trials in a row (nosepoke duration > delay period), the delay period increased by 100 ms to a maximum of 600 ms. If rats made premature responses for four trials in a row (nosepoke duration < delay period), the delay period decreased by 100 ms to a minimum of 100 ms. Additionally, 10% of the trials were “catch” trials. On these trials, no stimulus was presented at the end of the delay period and responses had no consequence and the “reaction time” was measured relative to the time at which the stimulus would have occurred. The vast majority of trials occurred with a fixed delay of 600 ms. Therefore, responses were initially similar since the rats were anticipating the timing of the stimulus.

The time from stimulus onset to nosepoke exit was defined as the reaction time (RT). The time from nosepoke exit until lick entry was defined as the movement time (MT). RTs and MTs are reported as median \pm interquartile range (IQR). Rats drank for several seconds, at which point they returned across the chamber to initiate a new trial. We defined the end of the lick period as the last lick before an inter-lick interval of >300 ms (Wilson and Bowman, 2005). Rats were run post-

operatively for several days on this task (3–4 d). Neural activity from the last day was selected for neural analysis.

Discrimination training. Rats were then advanced to a Go/NoGo reaction time task (Fig. 1). On any trial, after the rat inserted its snout into the nosepoke aperture, it could receive one of several stimuli, with only one stimulus presented on each trial. If the rat received an 8 kHz tone, it could make a Go response to collect a water reward (rewarded 8k+, or S+). However, if the rat received a 30 kHz tone (S–), Go responses led to a timeout for 4–8 s (unrewarded 30k–). Timeouts were reinitiated if the rat entered the nosepoke during the timeout period or were prolonged by 1 s for continued spout approaches. All behavioral devices, including the houselights, were turned off until the end of the timeout period. In pilot behavioral studies, rats learned equally well with the 8 kHz or 30 kHz tone serving as the rewarded stimulus. For simplicity, all rats were trained with the 8 kHz tone serving as the rewarded stimulus. Stimuli were selected pseudorandomly using the following probabilities: 8 kHz tone: 45%, 30 kHz tone: 45%, catch trial: 10%. There were no consequences of responding on catch trials, i.e., Go responses did not lead to water reinforcement or a timeout. No stimulus was presented more than three times in a row.

Discrimination learning was assessed by comparing responses to the 8 kHz tones with responses to the 30 kHz tones. Two criteria were used to assess successful discrimination. Go response rates to the two stimuli had to be different (>30% difference in Go responses over one entire session or $\geq 85\%$ accuracy on a contiguous block of 20 trials). Additionally, RTs to the two stimuli had to be significantly different (ranksum test, $p < 0.05$). These criteria were selected to ensure reliable differences in responding to the two stimuli without the risk of overtraining. Rats were trained for 2–6 sessions until they achieved both the accuracy and RT criteria in the same session. The first and last session of discrimination training were selected for data analysis.

It is important to point out that the rats did not always respond to S+ during discrimination training. Instead of attempting to collect rewards on these trials, rats would initiate a new trial by reentering the nosepoke. This kind of responding may have developed for several reasons. First, the rats were not intensively fluid restricted, i.e., body weight remained between 90 and 95%. Second, rats experienced timeouts during discrimination training for the first time in the overall training procedure. The resulting slowing of the rate of reward may have contributed to the increase in trial initiation rather than reward collection. Third, we did not want to overtrain the rats at the discrimination stage of training. If the rats had been trained on the original discrimination or on the simple RT task (using just the S+ stimulus), we expect that they would have eventually achieved a level of performance were they made many more Go responses to the S+ than were found in our study.

Reversal training. Once rats demonstrated selective responding to the two stimuli, the stimulus–reward contingencies were reversed in the next session (Fig. 1B). Responses to 30 kHz tones were now rewarded (30k+), and responses to 8 kHz tones were now punished with a timeout (8k). Rats were trained until they demonstrated successful reversal of responding, using the same criteria as in discrimination learning. Three sessions were selected from the reversal training for data analysis: the first session of reversal training, a session in the middle of reversal training in which rats responded at similar rates to the two stimuli without an RT difference, and a session in which rats achieved criteria performance on the reversed discrimination. Rats were trained for 6–17 sessions until achieving the new criteria.

Histology. At the conclusion of the recording sessions, rats were killed with an intraperitoneal injection of pentobarbital (≥ 100 mg/kg). Rats were then perfused with chilled saline followed by 4% formaldehyde. Brains were cut horizontally on a frozen sliding microtome, stained with thionin, dehydrated in an increasing concentration of ethanol followed by xylenes, and mounted and coverslipped on gelatin subbed slides. Electrode holes were identified using light microscopy and plotted onto a rat brain atlas (Paxinos and Watson, 1998). Three-dimensional models of the rat striatum were constructed using custom written software in Matlab (<http://spikelab.jbpierce.org/Resources/Code/3DAnatomy.zip>).

Physiological recordings. Signals from the implanted electrodes were transmitted to a recording system (Plexon Multichannel Acquisition

Processor) and amplified 1000–20,000 times. Spike activity was collected by setting a voltage threshold. Waveforms that crossed the threshold were time-stamped, sampled, and stored at 40 kHz. Unique waveforms were identified online and recorded. Online root-mean-square values, while rats were quietly resting, were typically $\sim 20 \mu\text{V}$ (calculated within Plexon OnLine Sorter). Waveforms were then processed off-line (Plexon OffLine Sorter) to remove artifacts and sorted into different units using principal component analysis and template-based methods. After processing, units had to meet several criteria to be considered single units: (1) mean peak-to-peak voltage had to be at least $100 \mu\text{V}$; (2) signal-to-noise ratio had to be at least 3:1; (3) fewer than 2% of interspike intervals could be $< 2 \text{ ms}$; (4) the mode of the interspike interval histogram had to be $> 5 \text{ ms}$; (5) mean firing rates had to be $< 30 \text{ Hz}$; (6) the distribution of maximal waveform points had to be relatively normal (skewness < 0.75). This latter measure ensured that waveforms were appropriately isolated from the threshold. Additionally, we evaluated neural activity during the drinking period to ensure that neurons were stationary. Neurons had to fire at least once on 10% of reinforcements and have a Z-score of < 3.5 on a runs test (Siegel, 1956). These characteristics were additionally verified by online and offline experimenter assessment.

The results of this study do not depend on recording activity from the same neuron across multiple days. It is difficult to determine conclusively whether electrophysiological signals recorded across multiple days reflect the activity of the same neuron. Practically, it is also unnecessary to do so, given that we are interested in studying the relationship of a particular striatal region to behavior. As such, we elected to use the statistical assumption that neurons represent repeated samples from the same location. Since our electrodes were chronically fixed and immovable, we can confidently state that neural activity was recorded from the same locations in striatum across training.

Our results are based on analysis of overall task-related activity for the neural populations. There are severe technical limitations to tracking neurons across sessions. We make no claim of recording from the same neurons across sessions, as we know of no completely satisfying way to identify a given neuron across a large number of sessions. Therefore, we are limiting our interpretations to those based on the population analysis of recordings of similar representative ensembles of neurons. The major population effects that are described in this study (i.e., overrepresentation of Go responses as training progresses, no change in fraction of neurons responsive to the discriminative stimuli) are relevant despite our not being able to comment on neural plasticity at the level of single identified neurons. We have more directly addressed this in a recent study that allowed for tracking some individual neurons (Kimchi et al., 2009) and did not find neurons that showed altered task-related activity when learning did not occur.

Data analysis. Data analysis was done in Matlab (Mathworks) and R (<http://www.R-project.org>). Custom-written software and Matlab functions from Plexon were used to analyze behavioral and neural data in Matlab. Exploratory and statistical analyses were performed using the statistics toolbox for Matlab and a variety of functions in R. Data were exchanged between these programs using the R.matlab library for R and the R (D)COM server. For all analyses of neural activity, the timestamps from identified single units were aligned to the behavioral events to create peri-event rasters and peri-event time histograms. Spike density functions were calculated by convolving the spike trains with a Gaussian ($\sigma = 20 \text{ ms}$) and used only for display purposes. Behavioral data were analyzed using nonparametric tests as implemented in Matlab (signrank

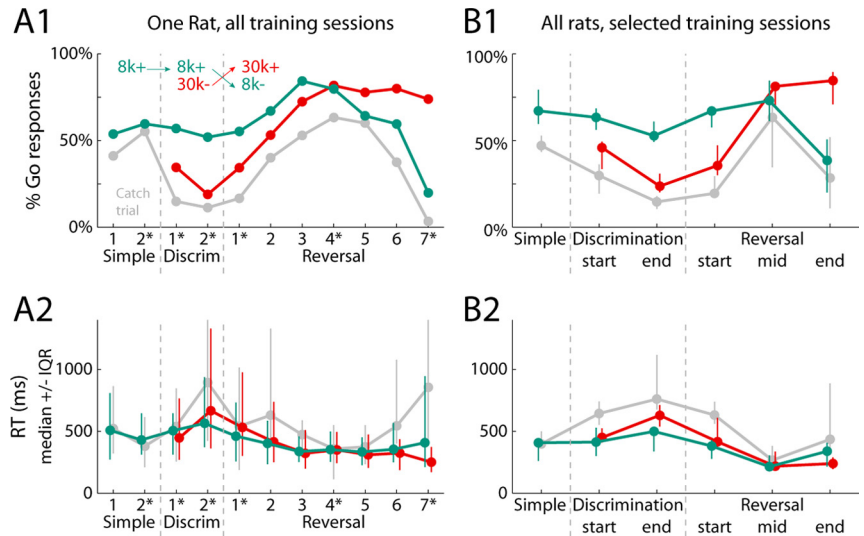


Figure 2. Behavior measures of discrimination and reversal learning. **A**, Behavioral data for a single rat across all training sessions. **A1**, The likelihood of a Go response after each stimulus is shown across training (green = 8 kHz tone; red = 30 kHz tone; gray = no stimulus/catch trial). The inset depicts the reward contingencies in effect at each stage of training. The dashed gray lines mark transitions between training stages (simple RT task, discrimination learning, and reversal). **A2**, Reaction time data for each trial type are shown across training (median \pm IQR). **B**, Behavioral data are shown from all rats at each stage of training (the first session of reversal training, a session in the middle of reversal training in which rats responded at similar rates to the two stimuli without an RT difference, and a session in which rats achieved criteria performance on the reversed discrimination). It was necessary to select sessions for analysis in this study due to rats requiring different numbers of training sessions to progress through the entire training process. Starred numbers in A indicate the sessions selected for analysis for that particular subject. Conventions for Go responses (**B1**) and RT data (**B2**) are as in A.

and ranksum) and ANOVA as implemented in R (function *aov*). Separate ANOVAs were used to compare neural activity over sessions during discrimination and reversal learning.

Identifying neural modulations. Neural modulations were identified by comparing changes in firing rates in two windows surrounding a behavioral event. This analysis specifically tests whether there is a reliable change in firing rate within trials rather than an increase or decrease relative to a predetermined epoch. Such a test was necessary given the rapid pace of the task and heterogeneity of firing patterns. One window encompassed spikes preceding the behavioral event and the second window encompassed spikes following the event. Firing rates in these two windows were compared using the signrank test and judged to be significantly modulated if $p < 0.05$. Only trials with a delay period at least 400 ms long and an RT $> 100 \text{ ms}$ were used for analysis.

Several time windows were explored, but most analyses presented below used a window of 200 ms before and after the event as stated. The duration was chosen to correspond roughly with that of earlier work on rat dorsomedial striatum (Dolbakyan et al., 1977; White and Rebec, 1993; Teagarden and Rebec, 2007). Only one test was performed for each event for each neuron, i.e., windows were not optimized to capture exact onset and duration of responses. While we may have underestimated the percentage of neurons modulated by each event, our approach allowed us to maintain a consistent p -value/ α level across comparisons.

To assess effects of different stimuli and responses on neural activity, we compared firing rates in a 200 ms window around the time of the relevant behavioral events. The brief duration of the window for firing rate analysis was chosen based on video assessment of behavior in the task. Behavior during the epochs was consistent over trials and animals. For comparisons between stimuli (trials with 8 kHz tones vs trials with 30 kHz tones), the analysis window was the 200 ms following stimulus onset. For comparisons between response choices (trials with Go responses vs trials with No-go responses), the analysis window preceded the response and was the 200 ms epoch before nosepoke exit. For analyses during the delay period (while the rat was waiting for a new stimulus), the analysis window was the 200 ms epoch before stimulus onset. Firing rates were compared for the relevant sets of trials using the ranksum test ($p <$

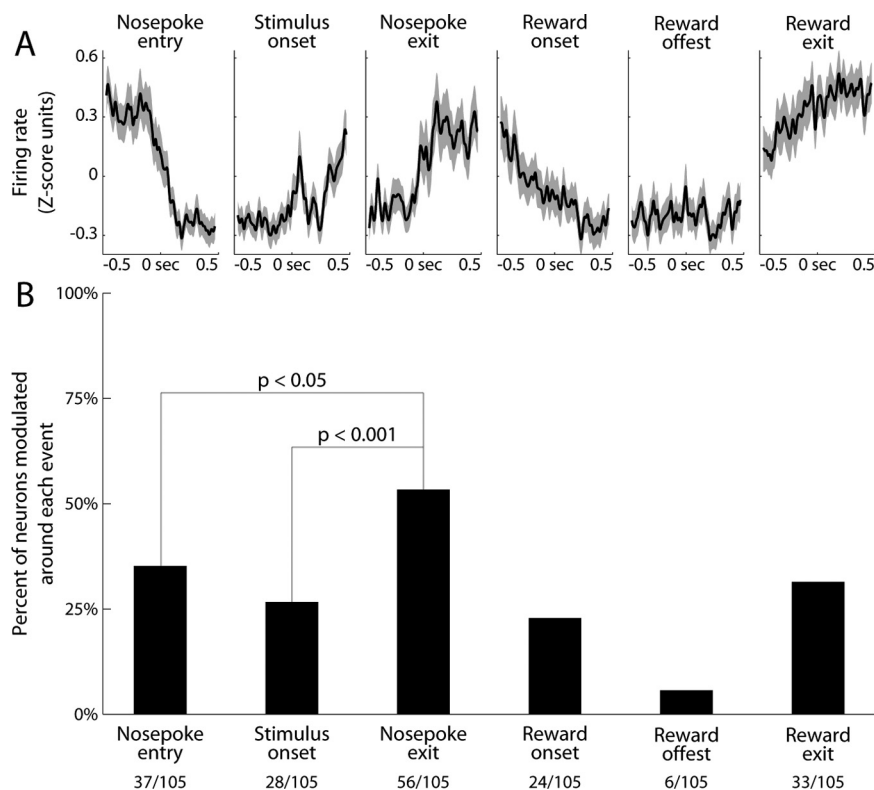


Figure 3. Task-related modulation of firing rate during the simple reaction time task. **A**, The mean firing rate was visibly modulated around most task events. Neural activity is depicted as the mean normalized firing rate of all neurons. Gray bands depict the SEM. Activity was highest during times of locomotion, but visibly modulated even around stimulus onset. **B**, The percentage of neurons modulated around each task event was determined by comparing firing rates before and after the event (± 200 ms windows, signrank test, $p < 0.05$). Just over half of all neurons had a modulation of firing rate at nosepoke exit, more than at any other single event in the task.

0.05). We also used a sliding window, in which bins of 200 ms were used every 50 ms, to capture dynamics of neural sensitivity.

Numbers of trials. For neural activity collected during the simple reaction time task, we analyzed one session from each rat and evaluated all rewarded trials from that session. This equilibrated the number of trials analyzed across different events within each rat. However, across days of training, the number of trials that were performed was more variable. To ensure that the number of available trials did not influence our statistical analysis of firing rates, we required a minimum of 50 and a maximum of 100 trials for any given event or set of trials in a session. This was done to equate statistical sensitivity across the training stages and is reflected in differences in the numbers of neurons used to evaluate different events as noted in the Results.

Results

Nine rats were trained to perform a simple reaction time task (Fig. 1A), and then implanted with multielectrode arrays to record neural activity from the dorsomedial striatum. Neural activity was recorded as animals performed a previously trained a simple reaction time task, through the initial acquisition of a Go/NoGo reaction time task, and through an initial reversal of the stimulus–reward contingencies in the Go/NoGo task (Fig. 1B).

Behavior during the simple reaction time task

Rats typically responded to the rewarded 8 kHz tones by withdrawing from the nosepoke promptly and crossing the chamber to collect water (Go response). They did this on a median of 67.1% of trials (IQR 59.4–70.2%). The median RT was 407 ms (IQR 260–433 ms). The median MT was 1.54 s (IQR 1.20–2.12 s). Rats withdrew from the nosepoke prematurely on a median of

23.9% of trials (IQR 19.7–36.8%). These premature responses were not specifically punished (no timeouts).

On 10% of trials, rats received no stimulus. These catch trials served to gauge the rats' response bias at each stage of training. In the simple reaction time task, rats made Go responses on almost half of the trials without a stimulus (median 46.9%, IQR 43.3–52.8%), which was less than on trials with a stimulus (signrank $p < 0.005$). RTs on trials without stimuli were 398 ms (IQR 365–499 ms), similar to RTs on trials with stimuli (signrank $p = 0.10$).

These results suggest that rats learned to time the interval between nosepoke entry and the stimulus (which was fixed at 600 ms) and that rats did not use the stimulus to initiate withdrawal from the nosepoke. However, the rats did come to use the stimulus to select between Go (approach spout) and No-go (avoid spout) responding, as described below.

Behavior during discrimination and reversal learning

Rats learned to successfully discriminate between the two stimuli and reverse the discrimination. During initial discrimination sessions, rats had to respond to rewarded 8 kHz tones (8k+) but not unrewarded 30 kHz tones (30k-) (Fig. 1). Go responses to 8k+ led to a water reward, whereas Go responses to 30k- were punished with a timeout. Rats reached criterion performance on the discrimination within several sessions (as defined in Materials and Methods). At this point the stimulus reward contingencies were switched, with Go responses to the 30 kHz tones now being rewarded (30k+) and Go responses to the 8 kHz tones now being punished with a timeout (8k-). The behavioral results for a single subject are shown in Figure 2A. Because different rats required different numbers of sessions to reach criteria, we defined common stages of training for all rats throughout discrimination and reversal learning using behavioral criteria (detailed in Materials and Methods). Results from the selected sessions are shown in Figure 2B for all rats.

By the end of the first training session, Go response rates to 8k+ tones (63.3%) were higher than response rates to the unrewarded 30k- tones (45.8%, signrank $p < 0.005$), which in turn were higher than Go response rates on trials without stimuli (29.7%, signrank $p < 0.02$). This pattern held for the later discrimination session as well, where the difference in response rates for the two stimuli was now larger (median difference 21.4% first session vs 33.0% later session, signrank $p < 0.03$). Median RTs to the two stimuli were not significantly different in the first session of discrimination training (8k+: 412 ms, 30k-: 449 ms, signrank $p = 0.43$). By the end of discrimination training, median RTs for unrewarded tones were significantly longer (8k+: 482 ms, 30k-: 666 ms, signrank $p < 0.001$). The RT differences depended more on response choice than the stimulus prompting the choice. Nosepokes followed by Go responses had a faster RT, regardless of the preceding tone (ANOVA, effect of response: $F_{(1,182)} = 75.1$, $p \ll 0.001$, effect of stimulus: $F_{(1,182)} = 0.2$, $p = 0.09$).

During reversal training, the 30 kHz tone became the rewarded stimulus (30k+) and the 8 kHz tone was now the unrewarded stimulus (8k-) (Fig. 1). Initially, Go responses were similar to those during discrimination training: lower rates and slower RTs for 30k+ than 8k- (Fig. 2; response rate difference in the “wrong” direction: signrank $p < 0.005$; RT difference: signrank $p < 0.005$). Within several days, rats responded at equal rates and speeds to the two stimuli (response rate difference: signrank $p = 0.36$; RT difference: signrank $p = 0.25$). This state was achieved by all rats, but with different amounts of training (median of 3 sessions, IQR 2.75–4.5 sessions). After several more training sessions, 8/9 rats successfully reversed their response preferences (median of 3.5 sessions, IQR 2.5–8 sessions). These rats now made Go responses more frequently to 30k+ than 8k- stimuli (84.5% vs 38.4%, signrank $p < 0.01$). Additionally, responses to 30k+ were significantly faster than to 8k- stimuli (30k+: 223 ms, 8k-: 338 ms, $p < 0.02$).

Neural data

We recorded spiking activity from a total of 618 dorsomedial striatal neurons over the selected stages of training. The median number of neurons recorded per rat in each session was 11 (IQR 8–15). The median firing rate for all neurons was 2.22 Hz (IQR 0.83–7.22 Hz).

Neural activity during the simple reaction time task

We recorded from 105 neurons during performance of the simple reaction time task. Striatal neural activity was visibly modulated around several task events (Fig. 3A). Activity tended to be highest during times of locomotion, such as before nosepoke entry or upon nosepoke exit. This pattern held only loosely for individual neurons, however, and there was striking heterogeneity and diversity across the population. We therefore calculated the percentage of neurons modulated around each specific task event (Fig. 3B). The majority of neurons were modulated by at least one task event (71%, 75/105; using Bonferroni correction for $p < 0.05$). Just over half of all neurons were modulated around the time of Nosepoke exit (53%, 56/105). This was a higher proportion than for any other single event, such as Nosepoke entry (35%, 37/105, $\chi^2 = 6.3$, $p < 0.05$) or Stimulus onset (27%, 28/105, $\chi^2 = 14.5$, $p < 0.001$).

Stimulus-related neural activity during discrimination and reversal

We next recorded neural activity as rats were trained on a novel Go/NoGo discrimination and reversal. The average activity pat-

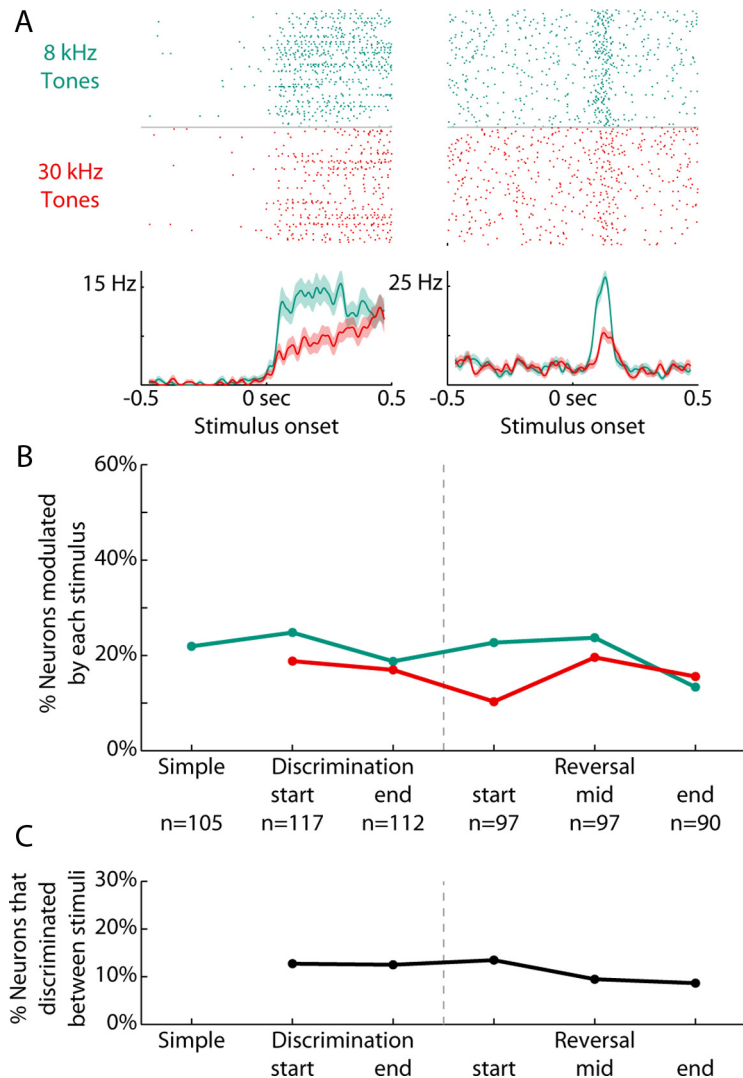


Figure 4. Stimulus-related activity did not change with training. *A*, Rasters and peri-event time histograms are presented from two neurons that discriminated between the rewarded and unrewarded stimuli. In green are the trials using the rewarded 8 kHz tone (8k+), while in red are the trials using the unrewarded 30 kHz tone (30k-). *B*, The percentage of neurons modulated around each stimulus (± 200 ms, signrank $p < 0.05$) are plotted across the various stages of training. Neural modulations to 8 kHz tones are in green, modulations to 30 kHz tones are in red. Gray dashed line indicates when reversal occurred (from 8k+/30k- to 30k+/8k-). The n refers to the number of neurons recorded at that stage across all rats. The percentage of neurons modulated around each stimulus did not change significantly over training (ANOVA, effect of training: $F_{(5,78)} = 2.0$, $p = 0.09$, interaction between training and stimuli: $F_{(4,78)} = 1.2$, $p = 0.30$). *C*, The percentage of neurons that discriminated between the two stimuli (200 ms following stimulus onset, ranksum $p < 0.05$) are plotted throughout discrimination and reversal training. The percentage of neurons whose activity discriminated between the two stimuli did not change significantly over training (ANOVA, effect of training: $F_{(4,31)} = 0.6$, $p = 0.70$).

tern and neural variability during discrimination training was qualitatively similar to that during the simple reaction time task (data not shown). Because striatal activity is known to represent the behavioral significance of stimuli during familiar learning, we measured how many neurons were responsive to each stimulus throughout training. The activities of some neurons were modulated by the two stimuli, often differentially so. Figure 4A depicts two neurons that fired more after the 8 kHz tone than the 30 kHz tone. Throughout novel discrimination and reversal training, 21% of neurons were modulated around the time of the 8 kHz tone (107/513), whereas 16% of neurons were modulated around the time of the 30 kHz tone (84/513, $\chi^2 = 3.1$, $p = 0.08$). The percentage of neurons modulated around each stimulus did not change significantly over training (Fig. 4B; ANOVA, effect of

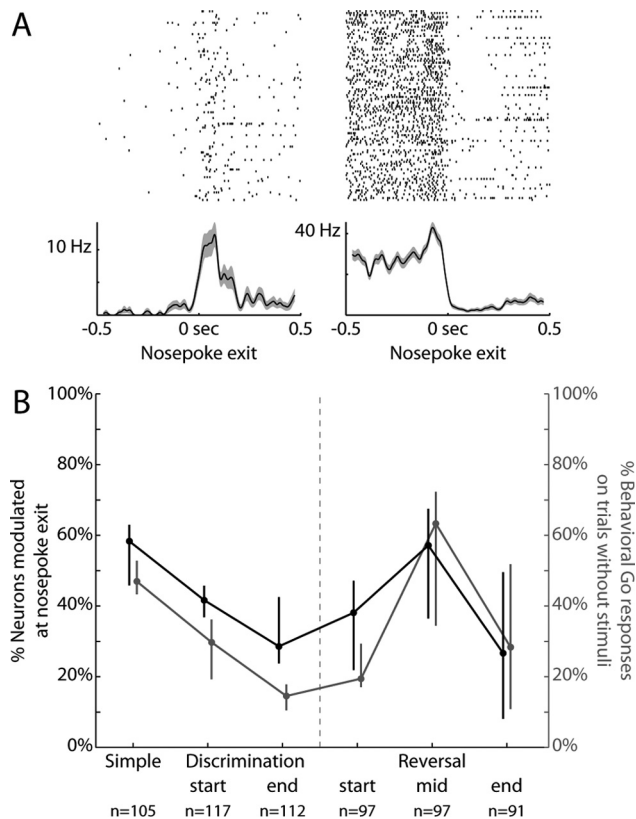


Figure 5. Response-related activity tracked the behavioral bias to respond. **A**, Rasters and peri-event time histograms are presented from two neurons whose activity changed around the time of nosepoke exit (± 200 ms, signrank $p < 0.05$). Of the neurons whose activity was modulated around response initiation at the Simple RT stage of training, slightly more than half increased their firing rates following nosepoke exit (as on the left, 61%, 34/56), while the rest decreased their firing rates (as on the right, 39%, 22/56). **B**, The percentage of neurons with changes in neural activity at nosepoke exit are plotted in black across the various stages of training. There was a significant change in the proportion of neurons modulated throughout training (ANOVA, effect of training during discrimination learning: $F_{(2,16)} = 6.8$, $p < 0.01$; effect of training during reversal learning: $F_{(2,15)} = 4.6$, $p < 0.03$). The percentage of behavioral Go responses on catch trials is replotted in gray as a measure of bias to respond. The training stages at which there was most likely to be nosepoke-related changes in neural activity were also those during which the response strategy was most likely to be driven by bias. Response-related neural modulations and the behavioral bias to respond were positively correlated for most rats (78%, 7/9; mean $r = 0.37$, $SD = 0.44$) and the correlations were significantly > 0 across subjects (t test comparison to 0, $p < 0.05$).

training: $F_{(5,78)} = 2.0$, $p = 0.09$, interaction between training and stimuli: $F_{(4,78)} = 1.2$, $p = 0.30$). Additionally, the percentage of neurons whose activity discriminated between the two stimuli (11%, 55/513; as in Fig. 4A) did not change significantly over training (Fig. 4C; ANOVA, effect of training: $F_{(4,31)} = 0.6$, $p = 0.70$). This was despite successful learning of the behavioral significance of the stimuli throughout discrimination learning and reversal (as revealed in Fig. 2). Exploratory analysis (raster plots) and more quantitative methods (receiver operating characteristic analysis) showed that there was no strong interaction between stimuli and response on neural firing.

Response-related neural activity during discrimination and reversal

During the simple RT task, more neurons were modulated around the time of Nosepoke exit than any other single event. We therefore examined this activity further to see how it might change over training. Figure 5A depicts two neurons with a mod-

ulation in firing rate closely aligned to the time of response initiation. Of the neurons whose activity was modulated around the behavioral response, slightly more than half increased their firing rates following nosepoke exit (61%, 34/56), while the rest decreased their firing rates (39%, 22/56).

In contrast to the relative stability of stimulus-related neural activity throughout discrimination and reversal learning, modulation of neural activity around the response was more dynamic (Fig. 5B, black line). There was a significant change in the proportion of neurons modulated throughout training (ANOVA, effect of training during discrimination learning: $F_{(2,16)} = 6.8$, $p < 0.01$; effect of training during reversal learning: $F_{(2,15)} = 4.6$, $p < 0.03$). The percentage of neurons with response-related changes in neural activity decreased by the end of discrimination training (35%, 39/112; vs simple response stage: $\chi^2 = 6.1$, $p < 0.05$; signrank $p < 0.05$). During the middle of reversal training, however, the percentage of neurons with response-related changes in neural activity increased again (53%, 51/97). This was significantly different than the end of both discrimination and reversal learning (vs end of discrimination learning: $\chi^2 = 6.0$, $p < 0.05$; signrank $p < 0.05$; vs end of reversal learning: $\chi^2 = 7.4$, $p < 0.01$; signrank $p < 0.05$).

The training stages at which there was most likely to be nosepoke-related modulation in neural activity were also those during which the response strategy was most likely to be driven by bias (Fig. 2). Behavioral bias could be measured by the tendency for rats to respond on silent, catch trials without stimuli. Catch trials comprised 10% of all trials at each training stage. Behavioral bias to respond without a stimulus was highest during the simple response stage and the middle of reversal training. For ease of comparison, the neural and behavioral curves are overlaid in Figure 5B (neural modulations in black, behavioral bias to respond on trials without stimuli in gray). Response-related neural modulations and the behavioral bias to respond were positively correlated for most rats (78%, 7/9; mean $r = 0.37$, $SD = 0.44$) and the correlations were significantly > 0 across subjects (t test comparison to 0, $p < 0.05$).

Neural activity related to response type

Because rats could either make Go or No-go responses at each stage of training, we examined whether neural activity was necessarily similar for each response. Figure 6A depicts two neurons whose activity in the 200 ms before nosepoke exit differed depending on whether the rat would make a Go or No-go response (ranksum $p < 0.05$).

Throughout training, 44% of neurons were modulated around nosepoke exit specifically on Go response trials (268/618, Fig. 6B). In contrast, 33% of neurons were modulated around nosepoke exit specifically on No-go response trials (166/505, $\chi^2 = 12.9$, $p < 0.001$). Across subjects, ANOVA revealed significant main effects of training (discrimination learning: $F_{(2,33)} = 5.8$, $p < 0.01$; reversal learning: $F_{(2,33)} = 5.0$, $p < 0.02$). There was a significant effect of response type (Go or No-go) during reversal learning ($F_{(1,33)} = 14.0$, $p < 0.001$), but not during discrimination learning ($F_{(1,33)} = 1.0$, $p > 0.3$). The interaction between these variables for both stages of training was not significant ($F < 1.1$ and $p > 0.3$ for both stages of training).

Across all training stages, 16% of all neurons had significantly different activity for the two response types in the 200 ms before nosepoke exit (83/505, ranksum $p < 0.05$). The percentage of neurons with such differential activity did not change with training (discrimination learning: $F_{(2,9)} = 0.1$, $p > 0.8$; reversal learning: $F_{(2,10)} = 1.7$, $p > 0.2$; Fig. 6C).

Neural activity related to the previous response

As so many neurons showed alterations in response-related activity during learning, we wondered whether there might be influences of preceding actions on neural activity. To examine this issue, we tested whether firing rate during the delay period, when rats were standing still and waiting for the next stimulus, depended on the outcome of the previous trial, the previous response that was made (Go or No-go), and the stimulus that was presented. In Figure 7A, we show activity from two neurons that fired at different rates on trials preceded by Go and No-go responses. For one neuron, activity was higher following a Go response. For the other, activity was higher following a No-go response. In the final 200 ms of the delay period, just before a stimulus was delivered, 21% of neurons were sensitive to the previous response (101/474 neurons across training). In contrast, at this time point, 11% of neurons were sensitive to the upcoming response (54/505, $\chi^2 = 19.9$, $p < 0.001$). Sensitivity to the previous rather than current response was increased throughout training and across rats (ANOVA, $F_{(1,60)} = 16.5$, $p < 0.001$), without a main effect of training (ANOVA, $F_{(5,60)} = 0.4$, $p = 0.82$) or interaction between response and training (ANOVA, $F_{(5,60)} = 0.7$, $p = 0.63$).

The transition from sensitivity to the previous response to sensitivity to the current response occurred during the RT epoch. Figure 7B depicts the percentage of neurons sensitive to the previous response (gray) or current response (black) in a 200 ms sliding window relative to stimulus onset. The percentage of neurons whose activity was sensitive to the previous response remained higher until ~300 ms after the onset of the stimulus. Figure 7C shows neural firing aligned to the onset of the response itself. Neurons were sensitive to both the preceding and current action of the animal and reflected the current response only after the animal was already moving.

Neural activity related to the trial outcome

If a rat made a Go response, it received clear feedback via the outcome of the trial. A correct Go response to a rewarded stimulus was positively reinforced with water. An incorrect Go response to an unrewarded stimulus was punished with a timeout, signaled by extinction of all behavioral devices.

Some neurons were modulated by each reinforcer (Fig. 8A). Throughout training, 28% of neurons were modulated around the onset of reward (153/552, Fig. 8B). A similar proportion of neurons were modulated around the onset of the timeout (33%, 113/347, $\chi^2 = 2.2$, $p = 0.14$). There was no effect of stage of training (ANOVA, $F_{(5,60)} = 1.6$, $p = 0.17$), reinforcer type ($F_{(1,60)} = 1.0$, $p =$

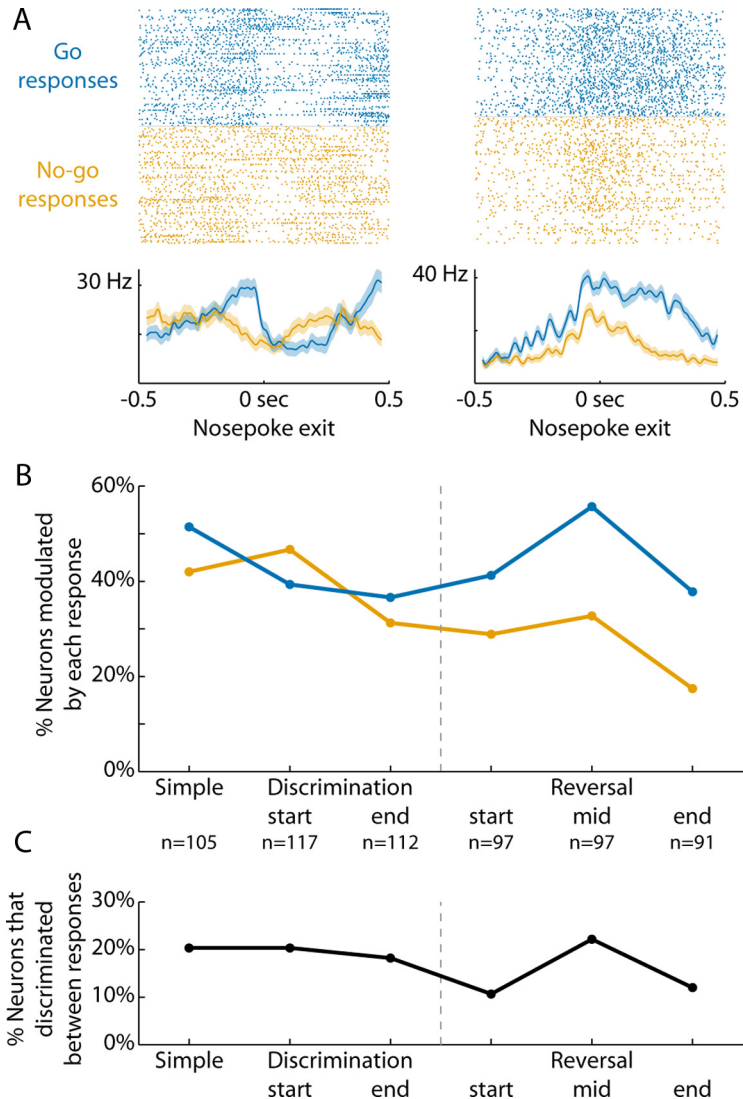


Figure 6. Delay period activity reflected action selection. **A**, Rasters and peri-event time histograms are presented from two neurons that discriminated between Go and No-go responses (Go = blue, No-go = orange; -200 ms window before nosepoke exit, ranksum $p < 0.05$). **B**, The percentage of neurons modulated on trials of each response type (± 200 ms, signrank $p < 0.05$) are plotted across the various stages of training. Across subjects, ANOVA revealed significant main effects of training (discrimination learning: $F_{(2,33)} = 5.8$, $p < 0.01$; reversal learning: $F_{(2,33)} = 5.0$, $p < 0.02$). There was a significant effect of response type (Go or No-go) during reversal learning ($F_{(1,33)} = 14.0$, $p < 0.001$), but not during discrimination learning ($F_{(1,33)} = 1.0$, $p > 0.3$). **C**, The percentage of neurons that discriminated between the two responses (-200 ms window before nosepoke exit, ranksum $p < 0.05$) are plotted throughout discrimination and reversal training. The percentage of neurons whose activity was different showed a statistical trend to change with training (ANOVA, discrimination learning: $F_{(2,9)} = 0.1$, $p > 0.8$; reversal learning: $F_{(2,10)} = 1.7$, $p > 0.2$).

0.32), or interaction between reinforcer and training ($F_{(4,60)} = 0.12$, $p = 0.97$). Of all neurons, 16% had different neural activity in the 200 ms following the onset of feedback (47/293). This proportion did not change significantly with training (ANOVA, $F_{(4,14)} = 5.5$, $p = 0.06$). Neural activity related to the previous outcome was not sustained as readily into the next trial as neural activity related to the previous response (Fig. 8C).

Discussion

We recorded neural activity in the dorsomedial striatum as rats were trained to discriminate between two stimuli and to reverse this discrimination for the first time. Changes in neural firing rates were observed across training sessions (Figs. 4A–8A), but not within individual sessions (i.e., on a trial-by-trial basis). As a

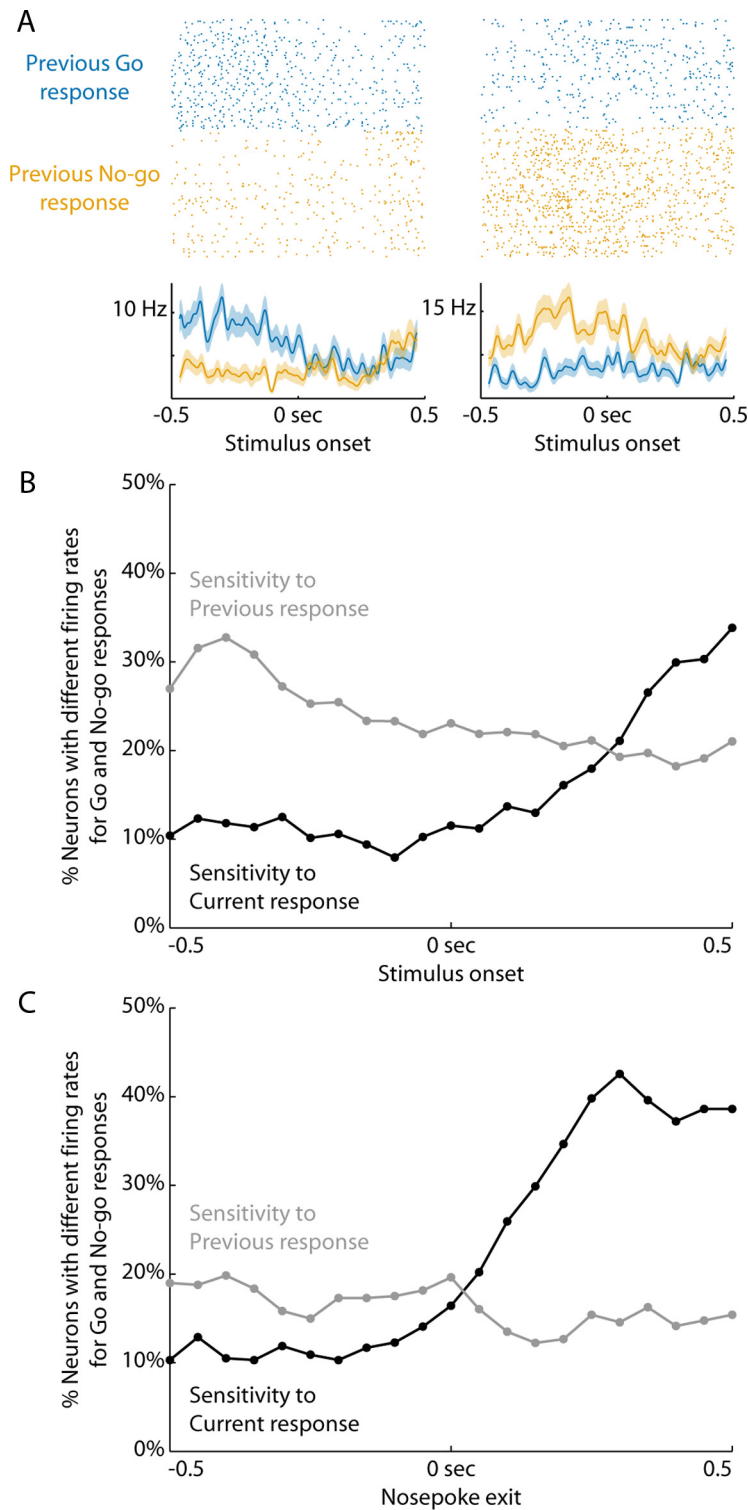


Figure 7. Neural firing rates were sensitive to the previous response. **A**, Rasters and peri-event time histograms are presented from two neurons sensitive to the previous response (200 ms window at the end of the delay period of the next trial before stimulus onset, ranksum $p < 0.05$). Trials after a Go response are blue and trials after a No-go response are orange. The neuron on the left fired more after a Go response and the neuron on the right fired more after a No-go response. In the final 200 ms of the delay period, just before a stimulus was delivered, 21% of neurons were sensitive to the previous response (101/474 neurons across training). **B**, The percentage of neurons sensitive to the previous response (gray) or current response (black) are plotted relative to the onset of the stimulus. A 200 ms sliding window was moved from -0.5 s before the stimulus until 0.5 s after, in 50 ms steps. Data are plotted at the time point signifying the end of the window, i.e., the point at 0 s includes neural activity from the 200 ms before the stimulus. The firing rates in these windows were compared for Go and No-go responses (ranksum, $p < 0.05$). The percentage of neurons that were sensitive to the previous response decreased as time progressed into the next trial. **C**, Sensitivity to the previous and current responses are plotted with neural activity aligned to nosepoke exit. Conventions are as in **B**. Neurons reflected the current response primarily after the animal was already moving.

result, we failed to observe rapid changes in neural activity during initial learning, as has been reported previously in studies that examined neural correlates of associative learning using procedurally familiar tasks (Brasted and Wise, 2004; Pasupathy and Miller, 2005; Williams and Eskandar, 2006; Kimchi and Laubach, 2009). Learning-related changes in neural activity were observed at the time of response initiation (Fig. 5) and during the delay period (Fig. 6), but not at the time of the stimulus (Fig. 4) or the consumption of reward (Fig. 8). We found no evidence for selective changes in striatal activity during reversal learning. Instead, changes in neural activity were associated with the level of response bias in the session (Fig. 4) and many neurons were sensitive to the behavioral response made on the preceding trial (Fig. 7). These findings suggest learning-related changes in the dorsomedial striatum are affected more by the animal's actions in the task than by the discriminative stimuli that actually determine how the rat should act.

Neural activity related to operant responding

Previous behavioral studies in rodents have shown that the dorsomedial striatum is necessary for response initiation in simple and choice reaction time tasks (Brown and Robbins, 1989a,b; Carli et al., 1989). As might be expected from these studies, dorsomedial striatal neurons were modulated at the time of response initiation (similar to Kimchi et al., 2009, who used a simple repetitive nosepoking task) and that the fraction of these response-related neurons increased during reversal learning. Similar to recent studies by Lau and Glimcher (2007) and Taha et al. (2007), most response-related neurons fired after response initiation (Fig. 3A, Nosepoke exit), and so it is unlikely that these neurons have a driving influence over response initiation. During the response, it is possible that there was overlap in the activation of movement- and outcome-related neurons. However, it is difficult to distinguish whether this overlap is due to true neural multiplexing of multiple types of information or to the interdependence of locomotor and consummatory behaviors in our task. Prior studies in the primate striatum have reported conflicting results related to this issue. In some cases, information is represented within the same population of neurons (Ding and Hikosaka, 2007). Other studies have reported that actions and outcomes influence distinct groups of cells (Lau and Glimcher, 2007).

We found no evidence for changes in the fraction of dorsomedial striatal neurons that fired to the discriminative stimuli (Fig. 4). However, we did find that the fraction of response-modulated neurons depended on the level of response bias in the testing session (Fig. 4): when response bias was high, more neurons showed task-related activity. During the reversal, the fraction of response-modulated neurons became as large as during the initial stages of training. As response bias has been shown to influence the striatum during the performance of procedurally familiar tasks (Lauwereyns et al., 2002; Itoh et al., 2003; Ding and Hikosaka, 2007), we suggest that the correlation between behavioral response bias and response-related activity observed in the present study is evidence for response bias as a major determinant of learning-related activity in the striatum.

Regarding the lack of stimulus-driven activity in the striatum (i.e., short latency increases in firing rate on trials with S+ or S−), we must admit it is possible that we would see such activity if animals were more extensively trained. In a recent study using well trained subjects, we failed to find such neural activity in the dorsal striatum (Kimchi and Laubach, 2009). In this task, rats learned to flexibly assign actions to acoustic stimuli and were trained over a period of several months to a relatively stringent behavioral criterion (>90%). In that study, we failed to find striatal neurons that fired in a manner that was suggestive of stimulus-driven activity (specific and robust responses at short latencies from stimulus onset).

The neural basis of reversal learning

Previous studies of the dorsomedial striatum have found that neurons are sensitive to changes in reward values of stimuli (Apicella et al., 1991; Tremblay et al., 1998), reward values of actions (Lauwereyns et al., 2002; Itoh et al., 2003; Watanabe et al., 2003; Samejima et al., 2005; Ding and Hikosaka, 2006; Lau and Glimcher, 2007), and the reward context (Kawagoe et al., 1998). Dorsomedial neurons represent arbitrary associations between stimuli and responses and adjust firing rates rapidly following changes in task contingencies (Brasted and Wise, 2004; Pasupathy and Miller, 2005; Watanabe and Hikosaka, 2005; Williams and Eskandar, 2006; Kimchi and Laubach, 2009). All of the above studies were done in animals with extensive experience with the task procedure. In contrast, in the present study we observed relatively slow changes in striatal activity during task acquisition, similar to other studies of discrimination learning outside of the basal ganglia (e.g., Shuler and Bear, 2006), and found that the level of activation in the striatum depended on the subject's response bias within the session.

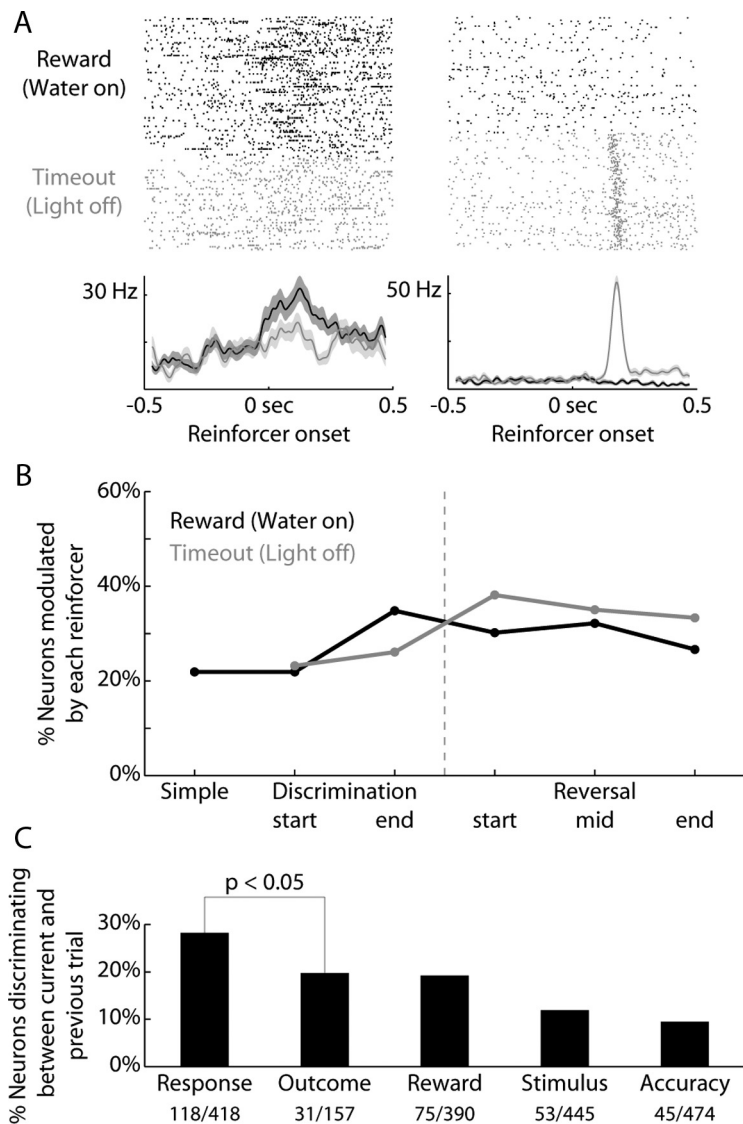


Figure 8. Reward-related activity did not change over training. **A**, Rasters and peri-event time histograms are presented from two neurons that discriminated between positive and negative outcomes (black = reward, gray = timeout; 200 ms window following reinforcer onset, ranksum $p < 0.05$). **B**, The percentage of neurons modulated on trials of each outcome (± 200 ms, signrank $p < 0.05$) are plotted across the various stages of training. There was no effect of stage of training (ANOVA, $F_{(5,60)} = 1.6$, $p = 0.17$), reinforcer type ($F_{(1,60)} = 1.0$, $p = 0.32$), or interaction between reinforcer and training ($F_{(4,60)} = 0.12$, $p = 0.97$). **C**, The percentage of neurons whose prestimulus activity reflects various previous trial types (Response = previous Go vs No-go response; Outcome = previous reward vs timeout; Reward = previous reward vs no reward, which depends on both previous response and outcome). Neural activity related to the previous outcome was not sustained as readily into the next trial as neural activity related to the previous response ($\chi^2 = 3.8$, $p < 0.05$).

Our results predict that periods of increased response bias, observed when task conditions are uncertain, may depend more on striatal processing than periods of stable task performance. From this perspective, deficits in reversal learning and behavioral flexibility that result from damage in the striatum may be due to the lack of “fallback strategy” (try to collect reward no matter what the stimulus was). New behavioral studies are needed to test this proposition. Paradoxically, one would predict that behavior might be most “automatic” before achievement of successful behavioral adjustment. This interpretation could actually account for findings that the dorsomedial striatum is active early during more familiar learning, even appearing to lead other brain areas such as other parts of striatum or prefrontal cortex (Pasupathy and Miller, 2005; Williams and Eskandar, 2006).

Our results may also provide insight into differences between early and later stages of reversal learning. Early stages of a reversal depend on orbitofrontal cortex (Schoenbaum et al., 2002, 2003; McAlonan and Brown, 2003; Tait and Brown, 2007) and later stages of reversal depend more on medial frontal cortex (Quirk et al., 2000; Rhodes and Killcross, 2007). Both of these cortical regions project to the dorsomedial striatum (McGeorge and Faull, 1989; Berendse et al., 1992; Cheatwood et al., 2003, 2005; Reep et al., 2003). Moreover, differences in reversal behavior have been noted following lesions of the cortex and striatum (Dube et al., 1996; Ragozzino, 2007; Ragozzino and Rozman, 2007; Clarke et al., 2008), with cortical lesions having pronounced effects early in a reversal. These studies support the idea that orbitofrontal neurons represent the current value of a stimulus based on a comparison between expected and obtained outcomes (Schoenbaum et al., 2007). The present study adds to this view and suggests that the striatum has a different role than orbitofrontal cortex in reversal learning: if the striatum is damaged, subjects are unable to generate responses in a spontaneous manner due to the lack of response bias signals generated by dorsomedial striatal neurons and are thus unable to recover from the reversal.

Response bias and a memory system for action

Throughout training, we observed the highest levels of task-related activity in the striatum in sessions with relatively high levels of response bias. The striatum might thus control the level of exploratory behavior emitted by the animal, especially when the outcome of a given action has become uncertain. Such a role for the basal ganglia in generating spontaneous variations in behavior has been proposed during birdsong learning (Kao et al., 2005) and may also explain classic effects of striatal lesions of the striatum on spontaneous alternation behavior (Chorover and Gross, 1963; Divac et al., 1978). From a theoretical perspective, bias-related changes in striatal activity could enable plasticity in cortical systems that receive projections from thalamic areas innervated by the basal ganglia, as proposed in a leading theory of basal ganglia function by Houk and Wise (1995) and also in a computational model of birdsong learning (Troyer and Doupe, 2000). A recent study by Andalman and Fee (2009) has provided direct evidence for bias-like signals existing within the basal ganglia components of the songbird system and being under control of cortex-like areas (the anterior forebrain pathway). Studies like this one, combining methods for neural recording, reversible inactivation, and microstimulation are needed to advance our understanding of neural circuits for operant conditioning in the mammalian brain.

In our Go/NoGo task, response bias could reflect the level of behavioral activation (Salamone and Correa, 2002), the incentive properties of the spout and other contextual elements of the experimental chamber (Berridge and Robinson, 1998), or a short-term memory of the animal's preceding action (White, 1997; Kesner and Rogers, 2004). We view this latter possibility as especially intriguing. Manipulations of processing in the dorsomedial striatum do not impair the initial stages of discrimination learning, e.g., behavior in the few trials immediately following a first reversal or shift in response strategy. However, they do impair reversal learning and lead to increases in "regressive errors" in which performance is based on previously learned task strategies (Ragozzino et al., 2002; Palencia and Ragozzino, 2004). Neural evidence for such traces of previous actions have been observed in the ventral striatum (Kim et al., 2007) and traces of prior outcomes have been found in the dorsomedial striatum (Histed et al., 2009). Based on these results, it is possible that memory of a

previous response, independent of the stimulus that triggered the response or the resulting outcome, may have lasting influences on striatal neurons. Dopaminergic signaling within the corticostriatal pathways could provide a mechanism for linking these representations from one trial to the next, as proposed by Horvitz (2009). By combining memories of prior actions, stimuli, and outcomes, each perhaps represented in different regions of the striatum or in other brain areas such as the frontal cortex, it may be possible to optimize the learning process based on recent task performance.

References

- Adams S, Kesner RP, Ragozzino ME (2001) Role of the medial and lateral caudate-putamen in mediating an auditory conditional response association. *Neurobiol Learn Mem* 76:106–116.
- Andalman AS, Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proc Natl Acad Sci U S A* 106:12518–12523.
- Apicella P, Ljungberg T, Scarnati E, Schultz W (1991) Responses to reward in monkey dorsal and ventral striatum. *Exp Brain Res* 85:491–500.
- Balleine BW, Delgado MR, Hikosaka O (2007) The role of the dorsal striatum in reward and decision-making. *J Neurosci* 27:8161–8165.
- Berendse HW, Galis-de Graaf Y, Groenewegen HJ (1992) Topographical organization and relationship with ventral striatal compartments of prefrontal corticostriatal projections in the rat. *J Comp Neurol* 316:314–347.
- Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Brain Res Rev* 28:309–369.
- Brainard DH (1997) The Psychophysics Toolbox. *Spat Vis* 10:433–436.
- Brasted PJ, Wise SP (2004) Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. *Eur J Neurosci* 19:721–740.
- Broadbent NJ, Squire LR, Clark RE (2007) Rats depend on habit memory for discrimination learning and retention. *Learn Mem* 14:145–151.
- Brown VJ, Robbins TW (1989a) Deficits in response space following unilateral striatal dopamine depletion in the rat. *J Neurosci* 9:983–989.
- Brown VJ, Robbins TW (1989b) Elementary processes of response selection mediated by distinct regions of the striatum. *J Neurosci* 9:3760–3765.
- Carelli RM, Wolske M, West MO (1997) Loss of lever press-related firing of rat striatal forelimb neurons after repeated sessions in a lever pressing task. *J Neurosci* 17:1804–1814.
- Carli M, Jones GH, Robbins TW (1989) Effects of unilateral dorsal and ventral striatal dopamine depletion on visual neglect in the rat: a neural and behavioural analysis. *Neuroscience* 29:309–327.
- Cheatwood JL, Reep RL, Corwin JV (2003) The associative striatum: cortical and thalamic projections to the dorsocentral striatum in rats. *Brain Res* 968:1–14.
- Cheatwood JL, Corwin JV, Reep RL (2005) Overlap and interdigitation of cortical and thalamic afferents to dorsocentral striatum in the rat. *Brain Res* 1036:90–100.
- Chorover SL, Gross CG (1963) Caudate nucleus lesions: behavioral effects in the rat. *Science* 141:826–827.
- 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.
- Costa RM, Cohen D, Nicolelis MA (2004) Differential corticostriatal plasticity during fast and slow motor skill learning in mice. *Curr Biol* 14:1124–1134.
- Ding L, Hikosaka O (2006) Comparison of reward modulation in the frontal eye field and caudate of the macaque. *J Neurosci* 26:6695–6703.
- Ding L, Hikosaka O (2007) Temporal development of asymmetric reward-induced bias in macaques. *J Neurophysiol* 97:57–61.
- Divac I, Markovitsch HJ, Pritzel M (1978) Behavioral and anatomical consequences of small intrastriatal injections of kainic acid in the rat. *Brain Res* 151:523–532.
- Dolbakyan E, Hernandez-Mesa N, Bures J (1977) Skilled forelimb movements and unit activity in motor cortex and caudate nucleus in rats. *Neuroscience* 2:73–80.
- Dube WV, Callahan TD, McIlvane WJ, Deutsch CK, Ullman MD, Koul O,

- McCluer RH (1996) Auditory discrimination reversal learning and assessment of behavioral teratogenesis in rats. *Behav Processes* 37:197–207.
- Featherstone RE, McDonald RJ (2005) Lesions of the dorsolateral or dorsomedial striatum impair performance of a previously acquired simple discrimination task. *Neurobiol Learn Mem* 84:159–167.
- Hikosaka O (2007) Basal ganglia mechanisms of reward-oriented eye movement. *Ann N Y Acad Sci* 1104:229–249.
- Histed MH, Pasupathy A, Miller EK (2009) Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* 63:244–253.
- Horvitz JC (2009) Stimulus-response and response-outcome learning mechanisms in the striatum. *Behav Brain Res* 199:129–140.
- Houk JC, Wise SP (1995) Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. *Cereb Cortex* 5:95–110.
- Itoh H, Nakahara H, Hikosaka O, Kawagoe R, Takikawa Y, Aihara K (2003) Correlation of primate caudate neural activity and saccade parameters in reward-oriented behavior. *J Neurophysiol* 89:1774–1783.
- Jog MS, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM (1999) Building neural representations of habits. *Science* 286:1745–1749.
- Kao MH, Doupe AJ, Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433:638–643.
- Kawagoe R, Takikawa Y, Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci* 1:411–416.
- Kesner RP, Rogers J (2004) An analysis of independence and interactions of brain substrates that subservise multiple attributes, memory systems, and underlying processes. *Neurobiol Learn Mem* 82:199–215.
- Kim YB, Huh N, Lee H, Baeg EH, Lee D, Jung MW (2007) Encoding of action history in the rat ventral striatum. *J Neurophysiol* 98:3548–3556.
- Kimchi EY, Laubach M (2009) Dynamic encoding of action selection by the medial striatum. *J Neurosci* 29:3148–3159.
- Kimchi EY, Torregrossa MM, Taylor JR, Laubach M (2009) Neuronal correlates of instrumental learning in the dorsal striatum. *J Neurophysiol* 102:475–489.
- Kirkby RJ (1969) Caudate nucleus lesions and perseverative behavior. *Physiol Behav* 4:451–454.
- Lau B, Glimcher PW (2007) Action and outcome encoding in the primate caudate nucleus. *J Neurosci* 27:14502–14514.
- Lauwereyns J, Watanabe K, Coe B, Hikosaka O (2002) A neural correlate of response bias in monkey caudate nucleus. *Nature* 418:413–417.
- Macmillan NA, Creelman CD (2005) *Detection theory: a user's guide*, Ed 2. Mahwah, NJ: Lawrence Erlbaum.
- McAlonan K, Brown VJ (2003) Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. *Behav Brain Res* 146:97–103.
- McGeorge AJ, Faull RL (1989) The organization of the projection from the cerebral cortex to the striatum in the rat. *Neuroscience* 29:503–537.
- Palencia CA, Ragozzino ME (2004) The influence of NMDA receptors in the dorsomedial striatum on response reversal learning. *Neurobiol Learn Mem* 82:81–89.
- Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433:873–876.
- Paxinos G, Watson C (1998) *The rat brain in stereotaxic coordinates*, Ed 4. San Diego: Academic.
- Pisa M, Cyr J (1990) Regionally selective roles of the rat's striatum in modality-specific discrimination learning and forelimb reaching. *Behav Brain Res* 37:281–292.
- Quirk GJ, Russo GK, Barron JL, Lebron K (2000) The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *J Neurosci* 20:6225–6231.
- Ragozzino ME (2007) The contribution of the medial prefrontal cortex, orbitofrontal cortex, and dorsomedial striatum to behavioral flexibility. *Ann N Y Acad Sci* 1121:355–375.
- Ragozzino ME, Rozman S (2007) The effect of rat anterior cingulate inactivation on cognitive flexibility. *Behav Neurosci* 121:698–706.
- Ragozzino ME, Ragozzino KE, Mizumori SJ, Kesner RP (2002) Role of the dorsomedial striatum in behavioral flexibility for response and visual cue discrimination learning. *Behav Neurosci* 116:105–115.
- Reep RL, Cheatwood JL, Corwin JV (2003) The associative striatum: organization of cortical projections to the dorsocentral striatum in rats. *J Comp Neurol* 467:271–292.
- Rhodes SE, Killcross AS (2007) Lesions of rat infralimbic cortex result in disrupted retardation but normal summation test performance following training on a Pavlovian conditioned inhibition procedure. *Eur J Neurosci* 26:2654–2660.
- Salamone JD, Correa M (2002) Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. *Behav Brain Res* 137:3–25.
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337–1340.
- Schoenbaum G, Nugent SL, Saddoris MP, Setlow B (2002) Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. *Neuroreport* 13:885–890.
- Schoenbaum G, Setlow B, Nugent SL, Saddoris MP, Gallagher M (2003) Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. *Learn Mem* 10:129–140.
- Schoenbaum G, Saddoris MP, Stalnaker TA (2007) Reconciling the roles of orbitofrontal cortex in reversal learning and the encoding of outcome expectancies. *Ann N Y Acad Sci* 1121:320–335.
- Shuler MG, Bear MF (2006) Reward timing in the primary visual cortex. *Science* 311:1606–1609.
- Siegel S (1956) *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Taha SA, Nicola SM, Fields HL (2007) Cue-evoked encoding of movement planning and execution in the rat nucleus accumbens. *J Physiol* 584:801–818.
- Tait DS, Brown VJ (2007) Difficulty overcoming learned non-reward during reversal learning in rats with ibotenic acid lesions of orbital prefrontal cortex. *Ann N Y Acad Sci* 1121:407–420.
- Tang C, Pawlak AP, Prokopenko V, West MO (2007) Changes in activity of the striatum during formation of a motor habit. *Eur J Neurosci* 25:1212–1227.
- Teagarden MA, Rebec GV (2007) Subthalamic and striatal neurons concurrently process motor, limbic, and associative information in rats performing an operant task. *J Neurophysiol* 97:2042–2058.
- Tremblay L, Hollerman JR, Schultz W (1998) Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J Neurophysiol* 80:964–977.
- Troyer TW, Doupe AJ (2000) An associational model of birdsong sensorimotor learning I. Efference copy and the learning of song syllables. *J Neurophysiol* 84:1204–1223.
- Watanabe K, Hikosaka O (2005) Immediate changes in anticipatory activity of caudate neurons associated with reversal of position-reward contingency. *J Neurophysiol* 94:1879–1887.
- Watanabe K, Lauwereyns J, Hikosaka O (2003) Neural correlates of rewarded and unrewarded eye movements in the primate caudate nucleus. *J Neurosci* 23:10052–10057.
- White IM, Rebec GV (1993) Responses of rat striatal neurons during performance of a lever-release version of the conditioned avoidance response task. *Brain Res* 616:71–82.
- White NM (1997) Mnemonic functions of the basal ganglia. *Curr Opin Neurobiol* 7:164–169.
- Williams ZM, Eskandar EN (2006) Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nat Neurosci* 9:562–568.
- Wilson DI, Bowman EM (2005) Rat nucleus accumbens neurons predominantly respond to the outcome-related properties of conditioned stimuli rather than their behavioral-switching properties. *J Neurophysiol* 94:49–61.