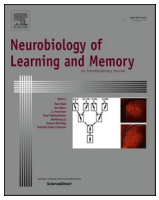
Neurobiology of Learning and Memory 199 (2023) 107732

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

journal homepage: www.elsevier.com/locate/ynlme



Dorsomedial striatum, but not dorsolateral striatum, is necessary for rat category learning

Matthew B. Broschard [a], Jangjin Kim [b], Bradley C. Love [c], John H. Freeman [d,*]

[a] The Picower Institute for Learning & Memory, Massachusetts Institute of Technology, Cambridge, MA, USA
[b] Department of Psychology, Kyungpool National University, Daegu, Republic of Korea
[c] Department of Experimental Psychology and The Alan Turing Institute, University College London, London, UK
[d] Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, USA

ARTICLE INFO

Keywords:
Category learning
Rats
Dorsal striatum
SUSTAIN
Computational modeling
Touchscreen

ABSTRACT

Categorization is an adaptive cognitive function that allows us to generalize knowledge to novel situations. Converging evidence from neuropsychological, neuroimaging, and neurophysiological studies suggest that categorization is mediated by the basal ganglia; however, there is debate regarding the necessity of each subregion of the basal ganglia and their respective functions. The current experiment examined the roles of the dorsomedial striatum (DMS; homologous to the head of the caudate nucleus) and dorsolateral striatum (DLS; homologous to the body and tail of the caudate nucleus) in category learning by combining selective lesions with computational modeling. Using a touchscreen apparatus, rats were trained to categorize distributions of visual stimuli that varied along two continuous dimensions (i.e., spatial frequency and orientation). The tasks either required attention to one stimulus dimension (spatial frequency or orientation; 1D tasks) or both stimulus dimensions (spatial frequency and orientation; 2D tasks). Rats with NMDA lesions of the DMS were impaired on both the 1D tasks and 2D tasks, whereas rats with DLS lesions showed no impairments. The lesions did not affect performance on a discrimination task that had the same trial structure as the categorization tasks, suggesting that the category impairments effected processes relevant to categorization. Model simulations were conducted using a neural network to assess the effect of the DMS lesions on category learning. Together, the results suggest that the DMS is critical to map category representations to appropriate behavioral responses, whereas the DLS is not necessary for categorization.

1. Introduction

The neural mechanisms underlying categorization are complex and likely involve multiple neural substrates, among these include the basal ganglia (Seger & Miller, 2010, Zeithamova et al., 2019; Ashby et al., 1998; Ashby & O'Brien, 2005). In a seminal paper, Knowlton, Mangels, & Squire (1996) had patients with Parkinson's Disease learn the Weather Prediction task, a probabilistic categorization task in which participants learn to combine information from sets of cards to predict the weather (i.e., 'sunny' or 'rainy'). Compared to healthy comparisons, the patients with Parkinson's Disease were impaired to learn this task, suggesting that the basal ganglia are necessary for categorization. Since this seminal work, the importance of the basal ganglia has been replicated by neuroimaging (Seger & Cincotta, 2002), neuropsychological (Maddox, Aparicio, Marchant, & Ivry, 2005), and neurophysiological experiments (Antzoulatos & Miller, 2014).

Multiple functions have been proposed regarding the role of the basal ganglia in categorization, including selective attention (Swainson et al., 2006; Brown & Marsden, 1988), set shifting (Owen, Roberts, Hodges, & Robbins, 1993; Volz et al., 1998; Lombardi et al., 1999), feedback processing (Schultz, Dayan, & Montague, 1997; Shohamy, Myers, Kalanithi, & Gluck, 2008), and stimulus generalization (Seger & Cincotta, 2005; see Shohamy, Myers, Kalanithi, & Gluck, 2008; Ashby & Ennis, 2006 for reviews). The diversity of these functions has been explained by assuming that each function is mediated through a distinct corticostriatal loop (Seger, 2008; Yahya, 2020). For instance, the head of the caudate nucleus may be critical for executive functioning via connections to the prefrontal cortex (i.e., the executive loop), whereas the tail and body of the caudate nucleus may be critical for stimulus generalization via connections to the visual cortex (i.e., the visual loop; Seger &

* Corresponding author at: Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242, USA.
E-mail address: john-freeman@uiowa.edu (J.H. Freeman).

https://doi.org/10.1016/j.nlm.2023.107732
Received 4 November 2022; Received in revised form 19 January 2023; Accepted 5 February 2023
Available online 9 February 2023

Cincotta, 2005; Lopez-Paniagua & Seger, 2011). Under this framework, the executive loop uses selective attention and feedback processing to test simple category rules (see Antzoulatos & Miller, 2011, 2014; Villagrasa et al., 2018; Shohamy, Myers, Onlaor, & Gluck, 2004; Schultz, Dayan, & Montague, 1997; Shohamy, Myers, Kalanithi, & Gluck, 2008). The visual loop, on the other hand, relies on the large convergence of sensory information from the visual cortex, allowing for generalization across stimuli that are perceptually similar (Seger, 2013).

These hypotheses have been formalized by a model of human category learning, COVIS (COmpetition between Verbal and Implicit Systems; Ashby et al., 1998). COVIS posits that humans learn new categories through two independent systems: the declarative system and the procedural system. The declarative system utilizes the executive loop and tests simple category rules, whereas the procedural system

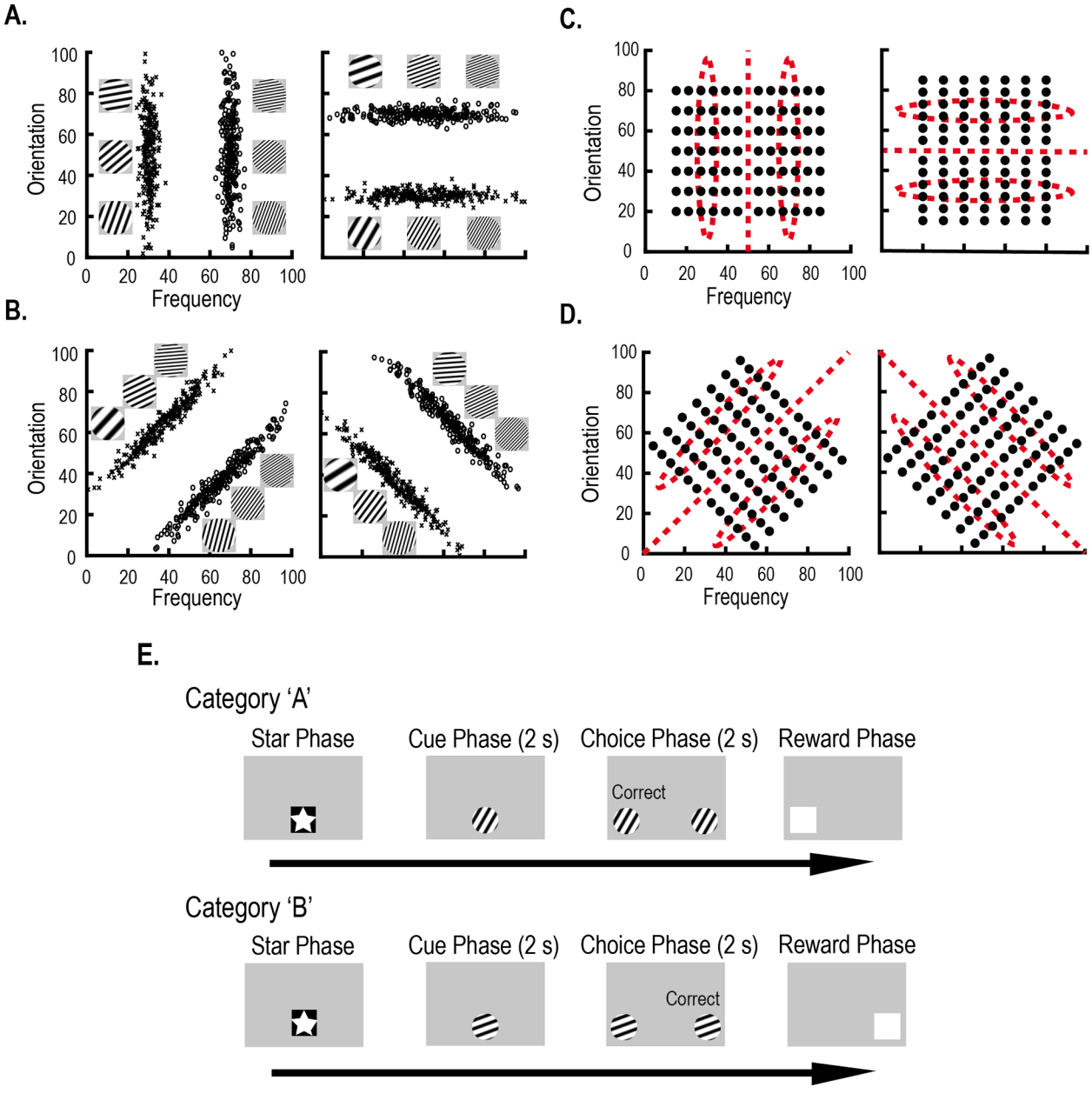


Fig. 1. A-B, Rats were randomly assigned to learn one of four category tasks. Category exemplars contained black and white gratings that varied in spatial frequency and orientation. Categories were created by placing normal distributions on this two-dimensional stimulus space. **A,** 1D tasks had category distributions that were perpendicular to a stimulus axis; therefore, only one stimulus dimension was category-relevant (i.e., the dimension perpendicular to the distributions). **B,** 2D tasks had category distributions that were not perpendicular to a stimulus axis; therefore, both stimulus dimensions were category-relevant. **C-D,** Rats were then given testing sessions to examine category generalization. Stimuli were configured into a grid that sampled across the entire stimulus space. One third of the stimuli overlapped with the training distributions (Trained; red ellipses), one third of the stimuli were closer to the category boundary compared to the training distributions (Proximal), and the remaining stimuli were farther from the category boundary compared to the training distributions (Distal). **E,** Each trial was initiated by touching the start stimulus (Star Phase). Then, a category stimulus was randomly sampled from one of the category distributions and was presented at the center of the screen (Cue phase). After three touches of the category stimulus, two copies of the stimulus appeared on the left and right sides of the screen, acting as report keys (Choice phase). Members of category 'A' required a touch to the left report key, and members of category 'B' required a touch to the right report key. After correct responses, a white box appeared on the screen (Reward phase); one touch of the white box delivered a food pellet. Correction trials were initiated after incorrect responses; here, the trial repeated from the Cue phase without food reinforcement. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

utilizes the visual loop and learns to incrementally associate category stimuli to appropriate behavioral responses. For decades, the predictions of the COVIS model have been tested by training participants to categorize distributions of visual stimuli that change along two continuous dimensions (e.g., spatial frequency and orientation; Fig. 1A-B; Ashby & Maddox, 2005). For some tasks, the categories can be learned using a unidimensional rule (i.e., rule-based; RB tasks; Fig. 1A). For other tasks, the categories must be learned by combining information from both stimulus dimensions (i.e., information integration; II tasks). Importantly, COVIS makes specific predictions such that the head of the caudate nucleus (and the declarative system) is critical to learn the RB tasks, whereas the tail of the caudate nucleus (and the procedural system) is critical to the learn the II tasks.

Support for these predictions is mixed. Patients with Parkinson's Disease are typically impaired to learn RB tasks (Maddox, Aparicio, Marchant, & Ivry, 2005; see Price, 2006; Ashby, Noble, Filoteo, Waldron, & Ell, 2003 for reviews) and not II tasks (Ashby et al., 2003; Filoteo, Maddox, Salmon, & Song, 2005). Impairments in the RB tasks are generally attributed to deficits in selective attention (Filoteo, Maddox, Ing, Zizak, & Song, 2005, 2007; Filoteo, Maddox, & Davis, 2001). A few neuroimaging studies have found dissociable basal ganglia activity in the RB and II tasks that aligns with the predictions of COVIS (Soto, Waldschmidt, Helie, & Ashby, 2013; Nomura et al., 2006); however, other studies do not replicate these results (Carpenter, Wills, Benattayallah, & Milton, 2016; Milton & Pothos, 2011). In sum, there is some support for the COVIS model; however, a more thorough test of these predictions may be necessary.

To clarify the roles of the basal ganglia in RB and II learning, we lesioned subregions of the rodent striatum. Multiple anatomical studies have compared homologies between the primate and rodent striatum (e.g., Heilbronner et al., 2016; Balsters et al., 2020). The head of the caudate nucleus in primates is likely homologous to medial portions of the rodent striatum (i.e., dorsomedial striatum; DMS), characterized by direct inputs from the prefrontal cortex. The tail of the caudate nucleus in primates is likely homologous to lateral portions of the rodent striatum (i.e., dorsolateral striatum; DLS), characterized by sensory and sensorimotor inputs (Foster et al., 2021; West et al., 1990). These homologies are supported by work comparing the functional similarity between the human and rodent striatum in the context of action selection (Balleine & O'Doherty, 2009; Yin & Knowlton, 2006). Specifically, the DMS and head of the caudate nucleus are both critical for selecting goal-directed behaviors, whereas the DLS and tail of the caudate nucleus are both critical for mediating habitual behaviors.

In the current experiment, we trained rats to learn adapted versions of the RB and II tasks using a touchscreen apparatus (i.e., 1D tasks and 2D tasks, respectively; Broschard, Kim, Love, Wasserman, & Freeman, 2019; Broschard, Kim, Love, & Freeman, 2020, 2021) to examine the roles of the DMS and DLS in category learning. Computational modeling simulated the functions of the DMS and DLS during category learning (Love, Medin, & Gureckis, 2004; Broschard et al., 2021; Broschard et al., 2019). We found that rats with lesions to the DMS were impaired on both task types, whereas lesioning the DLS had no effect on learning. Model simulations suggest that the DMS serves a general role in category learning by mapping category representations to appropriate behavioral responses. These results do not support the predictions of COVIS.

2. Methods & materials

2.1. Subjects

Sixty-two Long Evans rats (28 females; ~8 rats per group) were used for the following experiment. A power analysis was conducted based on previous experiments using the same experimental design (Broschard, 2021) to ensure the current experiment had adequate power (7 rats were needed per group to achieve 0.8 power; alpha = 0.05; effect size = 0.4). Rats were put on a 12-hour light/dark cycle and given *ad libitum* access to food and water. After acclimating to the new environment for a week, food was restricted. Weights were recorded daily to ensure each rat's weight did not go below 85 % of its free feeding weight. All procedures were approved by the Institutional Animal Care and Use Committee at the University of Iowa.

2.2. Touchscreen apparatus

Experimental sessions were conducted in custom-built touchscreen chambers (6 × 41 × 36 cm). A computer monitor (Model 1550 V, NEC, Melville, NY) was mounted on one wall of each chamber to present visual stimuli to the rats. A touchscreen (15-in, Elo Touch Systems, Fremont, CA) overlaid the computer monitor so that the rats could interact with the screen. A food tray (6.5 × 13 × 4.5 cm) was positioned at the wall opposite the computer monitor and delivered food pellets to the rats via a rotary pellet dispenser (Med Associates Inc., Georgia, VT, model ENV-203IR) that was controlled by an electrical board (Model RS-232, National Control Devices, Osceola, MO). A house light above the food tray was always on during experimental sessions. White noise was played in the room to minimize distractions. Custom MATLAB scripts controlled all sessions and procedures (MathWorks, Natick, MA; see Broschard et al., 2020 for more detail). A camera (model ELP-USB100W05MT-RL36) was mounted to the ceiling of the chamber to observe the rat's behavior.

2.3. Pre-training procedures

Each rat was handled daily for 1 week to reduce the stress of interacting with experimenters. Next, each rat was placed on the surface of a laboratory cart and was encouraged to forage for 45-mg pellets scattered on the cart's surface. This procedure allowed each rat to further habituate to the new environment and was repeated daily until the rat consumed twenty pellets within fifteen minutes. Finally, each rat underwent a daily shaping procedure to learn to interact with the touchscreen (see Broschard, Kim, Love, & Freeman, 2020 for details). This procedure included four separate phases; each phase was incrementally similar to the trial sequence used during training and testing sessions. All shaping procedures required about 14 days.

2.4. Surgery

After all pre-training procedures, each rat underwent stereotaxic surgery. Under isoflourane (1 % − 4 %) anesthesia, a Hamiltonian syringe (1 μL; 26 gauge) was lowered bilaterally into the DMS (AP: −0.4; ML: ±2.25; DV: −4.5) or the DLS (AP: 0.7; ML: ±3.6; DV: −4.8). NMDA (20 mg/ml; 0.4 μL) or PBS (0.4 μL) was infused (0.1 μL/min) into the target site. Stereotaxic coordinates were based on previous publications (Yin, Knowlton, & Balleine, 2004; Yin, Ostlund, Knowlton, & Balleine, 2005). After surgery, rats were placed on a heating pad until awake and mobile to prevent hypothermia. Meloxicam (1 mg/ml) was administered as an analgesic during surgery and 24 h after surgery to increase recovery. Rats were permitted at least one week to recover. Then, shaping sessions were given as a refresher to ensure each rat had maintained its performance.

2.5. Behavioral testing: An overview

After recovering from surgery, rats were trained to categorize distributions of visual stimuli. Briefly, rats were presented hundreds of training trials across multiple sessions. On each trial, a unique stimulus was displayed on the screen, and the rat decided the category membership of that stimulus (i.e., category 'A' or category 'B') by selecting one of two report keys. Food pellets were delivered after correct responses to guide learning.

2.6. Category stimuli

The category stimuli (239 × 239 pixels) used in this experiment contained black and white gratings (Fig. 1A-B) that varied along two continuous dimensions, spatial frequency and orientation. The spatial frequency of the gratings ranged from 0.2532 cycles per visual degree (cpd) to 1.2232 cpd, and the orientation of the gratings ranged from 0 rad to 1.75 rad. These ranges are within the perceptual limits of rats (Crijns & Op de Beeck, 2019). Linear transformations of these dimensions were made so that they had a common range (i.e., 0 to 100). Specifically,

$$\text{Normalized frequency} = \frac{cpd}{0.0097} - 26.10$$

$$\text{Normalized orientation} = radians * \frac{180}{pi}$$

A two-dimensional stimulus space was created using these transformed stimulus dimensions (Fig. 1A-B).

2.7. Category tasks

Category tasks were created by placing bivariate normal distributions on this transformed stimulus space (Fig. 1A; Category A: $\mu_X = 30$, $\sigma_X = 2.5$, $\mu_Y = 50$, $\sigma_Y = 20$; Category B: $\mu_X = 70$, $\sigma_X = 2.5$, $\mu_Y = 50$, $\sigma_Y = 20$; Broschard et al., 2019; Broschard et al., 2020; O'Donoghue, Broschard, & Wasserman, 2020). Each point within a distribution represented a category stimulus and each distribution constituted a category. Additional category tasks were created by rotating these distributions in 45-degree increments (Fig. 1A-B). This transformation does not affect any physical property of the distributions (Ashby, Smith, & Rosedahl, 2020; e.g., standard deviation, mean between-category distance, etc.), but it changes how the distributions are oriented in relation to the axes of the stimulus space. Specifically, 1D tasks had distributions that were perpendicular to one of the stimulus dimensions (Fig. 1A). For these tasks, only one stimulus dimension (i.e., the perpendicular dimension) was category-relevant and had to be considered when deciding category membership. The dimension parallel to the distributions was category-irrelevant and could be ignored. Conversely, 2D tasks had distributions that were not aligned with either stimulus axis (Fig. 1B). For these tasks, both dimensions were category-relevant, and category decisions involved combining information from both stimulus dimensions. COVIS predicts that lesioning the DMS would impair learning the 1D tasks, whereas lesioning the DLS would impair learning the 2D tasks.

2.8. Category training

Each rat was randomly assigned to learn one of the four category tasks (Broschard et al., 2019, 2020). Rats were given fifteen training sessions; each session contained 80 training trials. On each trial, a star stimulus was presented at the center of the screen (Fig. 1E; Star Phase). After one touch of the star, a category exemplar was randomly selected from the training distributions (Fig. 1A-B) and was presented on the screen (Cue Phase). After three observing touches of this exemplar, copies of the exemplar acted as report keys and were presented on the left and right sides of the screen (Choice Phase). The categories were mapped spatially, such that members of category 'A' required a touch to the left report key, and members of category 'B' required a touch to the right report key. If the correct side was chosen, a white box replaced the report key (Reward Phase); one touch of the white box delivered a food reward. If the incorrect side was chosen, a correction trial was initiated, such that the trial repeated from the Cue Phase after a 5 to 10 s time-out. Correction trials were repeated without reinforcement until the correct side was chosen or after 3 consecutive incorrect responses. Inter-trial intervals ranged from 5 to 10 s.

2.9. Category generalization

Rats were then presented with five testing sessions to examine generalization to novel stimuli. Each session contained 84 trials and included stimuli that sampled from a grid spanning the entire stimulus space (Broschard et al., 2019, 2020; Fig. 1C-D). The grid was designed such that a third of the stimuli overlapped with the training distributions (i.e., Trained; within two standard deviations), a third of the stimuli were closer to the category boundary relative to the training distributions (i.e., Proximal), and a third of the stimuli were farther from the category boundary relative to the training distributions (i.e., Distal). The trial procedure was identical to the training sessions except that correction trials were not initiated after incorrect responses. Therefore, all responses were reinforced.

2.10. Simple discrimination

As a control, rats were given training sessions to learn a simple discrimination task. All trial procedures were identical to category training, except that instead of distributions of stimuli, only two images were trained (i.e., a light box and a dark box; a common pattern of dots were added to the stimuli to add perceptual complexity; Fig. 6A; Kim, Castro, Wasserman, & Freeman, 2018). The white stimulus was mapped to the left report key, and the black stimulus was mapped to the right report key. Each session contained 72 training trials; sessions continued until reaching a learning criterion (i.e., at least 75 % accuracy for both images on two consecutive sessions). Critically, this discrimination acted as a control task to test whether the differences across groups were related to deficits in movement, motivation, or perception.

2.11. Histology

After all behavioral testing, each rat was perfused to verify the location and spread of the lesions. Each rat was given a lethal dose of euthanasia solution (sodium pentobarbital) and then perfused with ~ 150 mL PBS and ~ 150 mL formalin. Brains were stored at 4 °C. Then, a sliding microtome made 50 μm coronal sections of the target brain region. Slides were stained with thionin and cover slipped. Once dry, each slide was observed under a light microscope to examine the location and spread of each lesion. The boundaries of the DMS and DLS were defined according to Paxinos & Watson, 1998. The size of each lesion was quantified using the software ImageJ. Lesion volume was calculated using the equation of a sphere. The radius of this sphere was determined according to the brain section containing the track of the infusion needle (or the section containing the largest lesion if the track was not visible). Rats with lesions largely outside these regions were excluded from all analyzes.

2.12. Statistical analysis

Performance was quantified through two dependent measures. First, session accuracy was defined as the proportion of correct responses during the Choice phase. Second, reaction time was calculated during the Cue phase and Choice phase to quantify the amount of time to 1) observe the stimulus and 2) make a category decision. Reaction times from incorrect trials were excluded from all analyses. Additionally, reaction times that exceeded two standard deviations of the mean were excluded from all analyses, a criterion that is commonly used to eliminate outliers (O'Donoghue et al., 2020). These outliers rarely occurred.

Linear mixed effects modeling was used to analyze the dependent measures across groups (R, version 3.4.2). Models used for training sessions included fixed effects for experimental group, training session, and a quadratic function across training sessions, as well as random effects for slope, intercept, and the quadratic function. Models for testing sessions included fixed effects for experimental group, trial type (Distal, Trained, and Proximal), and a quadratic function across trial types, as

well as random effects for slope, intercept, and the quadratic function. Quadratic functions were used because they best fit the data, and higher order terms did not significantly improve these fits. Sex was added as a covariate for all models to check whether there were any significant differences between male and female rats. To find the simplest model that fit the data, we used a model simplification strategy (Crawley 2007). Briefly, random effects were systematically removed from the full model until the estimates were significantly different from the larger model before it.

2.13. SUSTAIN modeling

SUSTAIN is a computational model of human category learning that can reproduce benchmark categorization behavior (Love, Medin, & Gureckis, 2004; Fig. 7A). Here, we used SUSTAIN as a computational tool to simulate the effect of each lesion on specific components of the learning process. The SUSTAIN model was selected because multiple experiments have mapped specific components of the network to neural activity (e.g., Mack, Love, & Preston, 2016; Mack, Preston, & Love, 2020; Broschard et al., 2021). This suggests that specific brain regions may be functionally similar to computations performed by each layer within the network, providing a data-driven framework to interpret the results. Additionally, SUSTAIN has been useful in capturing both human and rat category learning behavior using the same stimuli and trial procedures (Broschard et al., 2020). This adds translational value to the interpretation of our results and suggests that SUSTAIN contains learning mechanisms that are relevant for both rats and humans.

SUSTAIN assumes that categories are represented by single or multiple clusters; each cluster represents a learned group of similar training experiences. Categorizing a new stimulus involves comparing that stimulus to the existing cluster representations; each cluster is activated according to its similarity to the stimulus. Cluster activations are projected to a decision layer, which makes a probabilistic decision regarding the category membership of that stimulus. Cluster activations are weighted by an attention mechanism that learns to emphasize stimulus dimensions that are category-relevant.

To understand the function of the dorsal striatum in rat category learning, we first fit SUSTAIN to the averaged learning data of the control groups using the MATLAB function *fmincon* (the fitted parameters were constant between the controls learning the 1D tasks and 2D tasks). This served as a baseline and provided a model that learned the category tasks at the same rate as a typical rat. Then, multiple experimental models were designed to investigate the effects of the striatal lesions. These experimental models were created by disrupting single components of the baseline model. Each experimental model assumed that the lesion produced a unique deficit during category learning; therefore, these models were useful in testing potential functions of the dorsal striatum, including selective attention (Swainson et al., 2006; Brown & Marsden, 1988), response mapping (Balleine & O'Doherty, 2009), and feedback processing (Schultz, Dayan, & Montague, 1997; Shohamy, Myers, Kalanithi, & Gluck, 2008). A full description of the experimental models is provided below.

The first model assumed that the target region was critical for selective attention (Selective Attention). For this model, the lesion groups were simulated by shuffling the attention weights of the attention mechanism before each trial, increasing the probability that attention would be directed toward category-irrelevant information. The Cluster Recruitment model assumed that the target region was critical for recruiting new cluster representations. In SUSTAIN, new clusters are formed after 'surprising' stimuli, where the model was confident in an ultimately incorrect category decision. Cluster recruitment is controlled by a threshold parameter; increasing this threshold parameter for the lesion groups impaired the ability of the model to recruit new clusters. The Feedback Processing model assumed that the target region was critical for processing feedback. This model simulated the lesion groups by adding a normal distribution of noise to the model's feedback after each trial. The mean of this distribution was 0, and the standard deviation of this distribution was a positive free parameter. The Response Mapping model assumed the target region was critical for mapping cluster representations to appropriate category labels. This model simulated the lesion groups by reducing the learning rate parameter that updates the connection weights between the cluster layer and the decision layer. Finally, the Control model assumed that the target region was not necessary for category learning.

Each experimental model was fit to the averaged learning data of the lesion groups using the MATLAB function *fmincon*. The goodness-of-fit was assessed for each model by calculating a BIC value (Neath & Cavanaugh, 2011). The underlying function of the target region was inferred from the experimental model that best fit the learning data (i.e., the model with the smallest BIC value).

2.14. Perceptual recency effects

Because each rat completed a large number of training trials, we could track category learning on a trial-by-trial basis. We leveraged this sensitivity to observe how category decisions were influenced by the identity of the most recent training exemplar (i.e., perceptual recency effects; Jones, Love, & Maddox, 2006). Trial-order effects assume that category representations are regularly updated, since decisions are biased according to local training experiences. Therefore, the current analysis examined perceptual recency effects as a proxy for understanding representational updating.

Recency effects often interact with the perceptual similarity between exemplars. For example, performance is facilitated if the exemplar is perceptually similar to the most recent exemplar (Jones et al., 2006). Thus, for the current analysis, we binned the accuracy of training trials according to the perceived similarity between the current exemplar (*n*) and the most recent exemplar (*n-1*; Nosofsky, 1986). Perceptual similarity between exemplars *i* and *j* was calculated as,

$$s_{ij} = e^{-d_{ij}}$$

where *d* is the psychological distance between exemplars *i* and *j*. Psychological distance was defined as,

$$d_{ij} = \sum_{m=1}^{M} w_m * \left| x_i - x_j \right|$$

where w_m was SUSTAIN's estimated attention weight for dimension *m* on trial *n,* and *x* was the physical value of the exemplar along dimension *m.* Trial effects were isolated by subtracting the binned accuracies by the average of 1,000 permutations where trial order was shuffled. Positive recency scores indicate increased accuracy due to trial order, negative scores indicate decreased accuracy due to trial order, and 0 indicates no effect of trial order.

3. Results

3.1. Histological assessment of DMS and DLS lesions

Representative lesions are shown in Fig. 2. Each lesion was examined under a light microscope to ensure that it was contained within the target region. Boundaries of the DMS and DLS were determined according to Paxinos & Watson (1998). All lesions were centered within their target site, and the data from all rats were included in all analyses. Along the rostral/caudal axis, all DMS lesions were contained between bregma + 0.2 and −0.75, and all DLS lesions were contained between bregma + 1.00 and + 0.45. These sizes are equivalent to previous studies that have lesioned the DMS and DLS (Yin, Knowlton, & Balleine, 2004; Yin, Ostlund, Knowlton, & Balleine, 2005). One of the DMS lesions (learning a 1D task) extended into the DLS, and two of the DLS lesions (learning a 2D task) extended into the DMS. None of the lesions

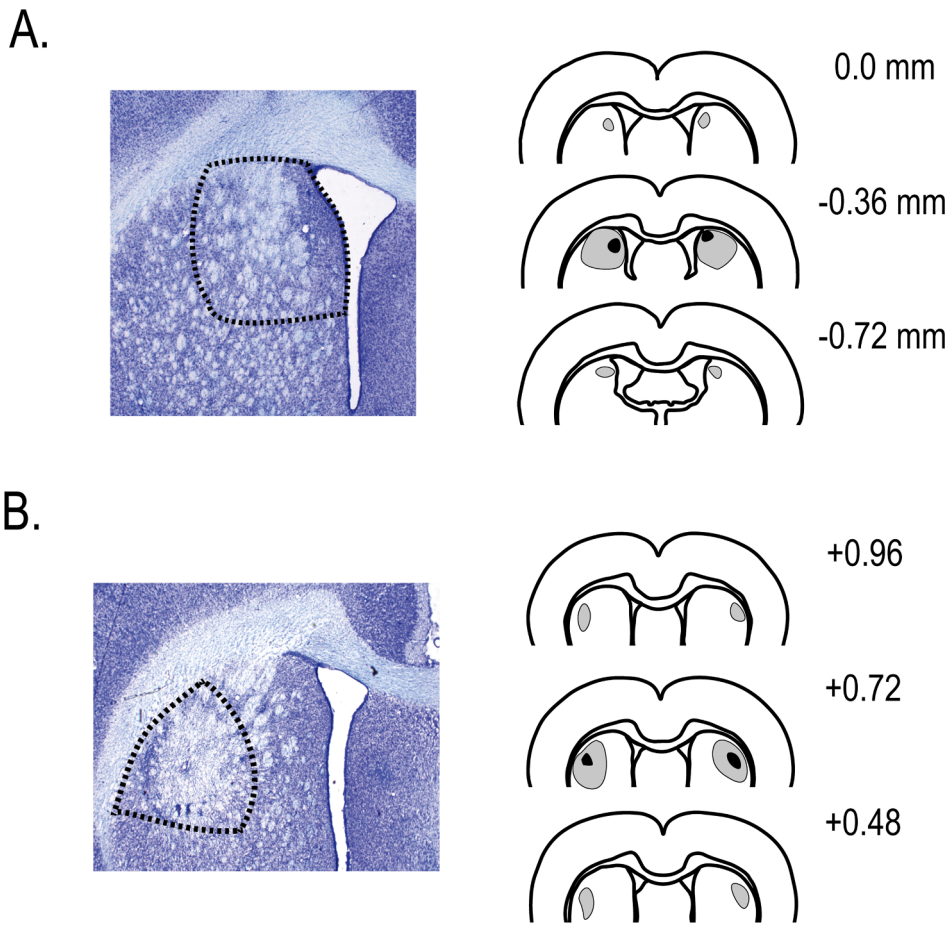


Fig. 2. Representative lesions of the DMS and DLS. **A,** Left: representative spread of the DMS lesions. Right: a comparison of lesion size and location for the smallest lesion (dark gray) and the largest lesion (light gray). All lesions were centered in the DMS and were contained within bregma + 0.2 and −0.75. One of the lesions extended into the DLS, and none of the lesions extended into the ventral striatum. **B,** Left: representative spread of the DLS lesions. Right: a comparison of lesion size and location for the smallest lesion (light gray) and the largest lesion (dark gray). All lesions were centered in the DLS and were contained within bregma + 1.00 and + 0.45. Two of the lesions extended into the DMS, and none of the lesions extended into the ventral striatum.

extended into the ventral striatum.

3.2. Lesions of the DMS, but not DLS, impair category learning

Session accuracy and reaction time were analyzed through linear mixed effects modeling. Each full model contained fixed effects for experimental group, training session, and a quadratic function across sessions, as well as random effects for slope, intercept, and the quadratic function. A covariate for sex was added to observe any differences between male and female rats.

The first model examined session accuracy. Accuracy increased significantly across sessions, replicating multiple experiments that demonstrate robust category learning in rats (Fig. 3; $t(57.85) = 8.51$, $p < .001$; Broschard et al., 2019; 2020; 2021). There were no significant differences in accuracy between males and females ($t(64.95) = 0.76$, $p = .450$), suggesting that both males and females learned the tasks at similar rates. Additionally, there were no significant differences in accuracy between controls learning the 1D tasks and 2D tasks ($t(64.36) = 0.64$, $p = .523$). This replicates past experiments and suggests that rats typically learn 1D tasks and 2D tasks at the same rate and to equal levels (Broschard et al., 2019). Rats with DMS lesions had significantly impaired accuracy across training compared to controls. This was true for rats learning the 1D tasks (Fig. 3A; $t(64.56) = 3.30$, $p = .002$) and the 2D tasks (Fig. 3B; $t(55.64) = 3.41$, $p = .001$). Conversely, lesioning the DLS did not affect session accuracy (Fig. 2C-D; 1D tasks: $t(52.92) = -0.161$, $p = .873$; 2D tasks: $t(50.00) = -0.27$, $p = .791$). Finally, accuracy impairments were significantly larger for rats with DMS lesions compared to rats with DLS lesions (1D tasks: $t(63.43) = 3.25$, $p = .003$; 2D tasks: $t(60.21) = 3.32$, $p = .002$). These results together indicate that the DMS, but not the DLS, is critical for category learning. Given that these areas have close anatomical proximity, these results indicate dramatic functional differences between striatal subregions.

We then tested whether accuracy was significantly correlated with the size of each lesion (Fig. 2E-F). Lesion size was quantified using image processing software (ImageJ), and mean accuracy of each rat was converted to a z-score relative to the control groups. For the rats with DMS lesions, accuracy was negatively correlated with lesion size, such that larger lesions produced a larger accuracy deficit ($r(16) = -0.56$, $p = .029$). This supports our finding that lesioning the DMS impairs category learning. For the DLS lesions, accuracy was not correlated with lesion size ($r(15) = -0.33$, $p = .250$).

3.3. Rats with DMS lesions, but not DLS lesions, require more time to categorize each stimulus

Next, we examined how the lesions affected reaction time during both the Cue phase and the Choice phase of each trial (i.e., Cue RT and Choice RT, respectively). First, Choice RT, but not Cue RT, decreased significantly across training sessions, suggesting that the amount of time to observe each stimulus remained constant across sessions, but the speed to make category decisions increased as rats learned the category tasks (Fig. 4 & Supplementary Fig. 1; Cue RT: $t(59.84) = -0.74$, $p = .461$; Choice RT: $t(57.15) = -5.64$, $p < .001$). Similar to session accuracy, there were no significant differences in reaction time between male and female rats, suggesting that the category tasks are learned equivalently for both males and females (Cue RT: $t(59.36) = 1.58$, $p = .121$; Choice RT: $t(57.13) = -0.28$, $p = .777$). Additionally, there were no significant differences in reaction time between control rats learning the 1D tasks and the 2D tasks (Cue RT: $t(51.95) = -0.05$, $p = .962$; Choice RT; $t(50.62) = -0.12$, $p = .909$), supporting our finding that rats typically learn the task

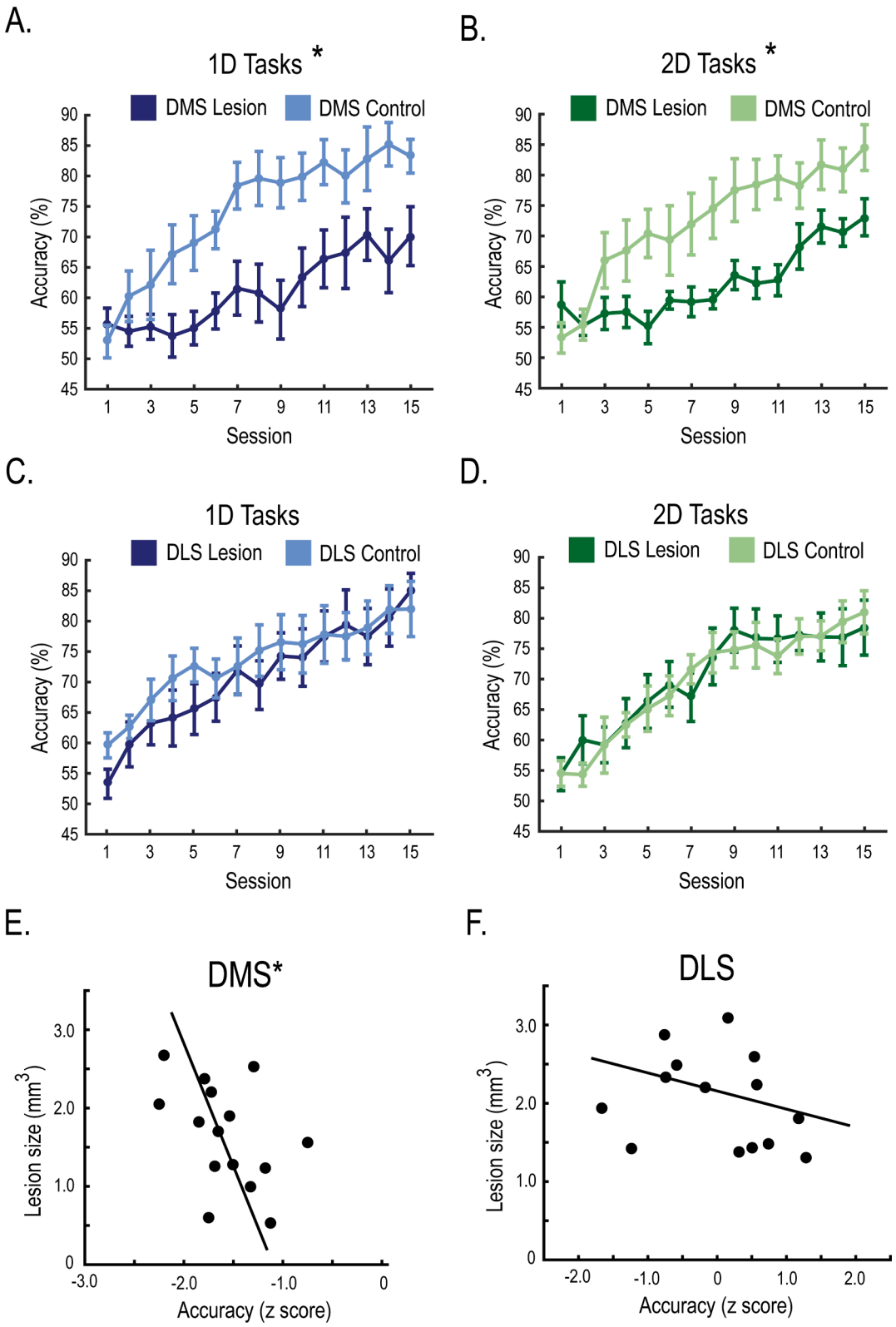


Fig. 3. Lesions of the DMS, but not the DLS, impaired category learning. **A-B,** Mean session accuracy for rats with DMS lesions learning the 1D tasks (A) and the 2D tasks (B) (n = 8 per group). Compared to controls, rats with DMS lesions had impaired accuracy for both task types. **C-D,** Mean session accuracy for rats with DLS lesions learning the 1D tasks (C) and the 2D tasks (D) (n = 8 per group). There were no significant differences in accuracy between rats with DLS lesions and controls. Together, these results suggest that the DMS, but not the DLS, is critical for category learning. All error bars indicate the *SEM*. **E-F,** Pearson's correlations between volume size (mm^3) and mean accuracy (z-score relative to controls). DMS lesion size was negatively correlated with mean accuracy (E), whereas DLS lesion size was not significantly correlated with mean accuracy (F). '*' indicates statistical significance.

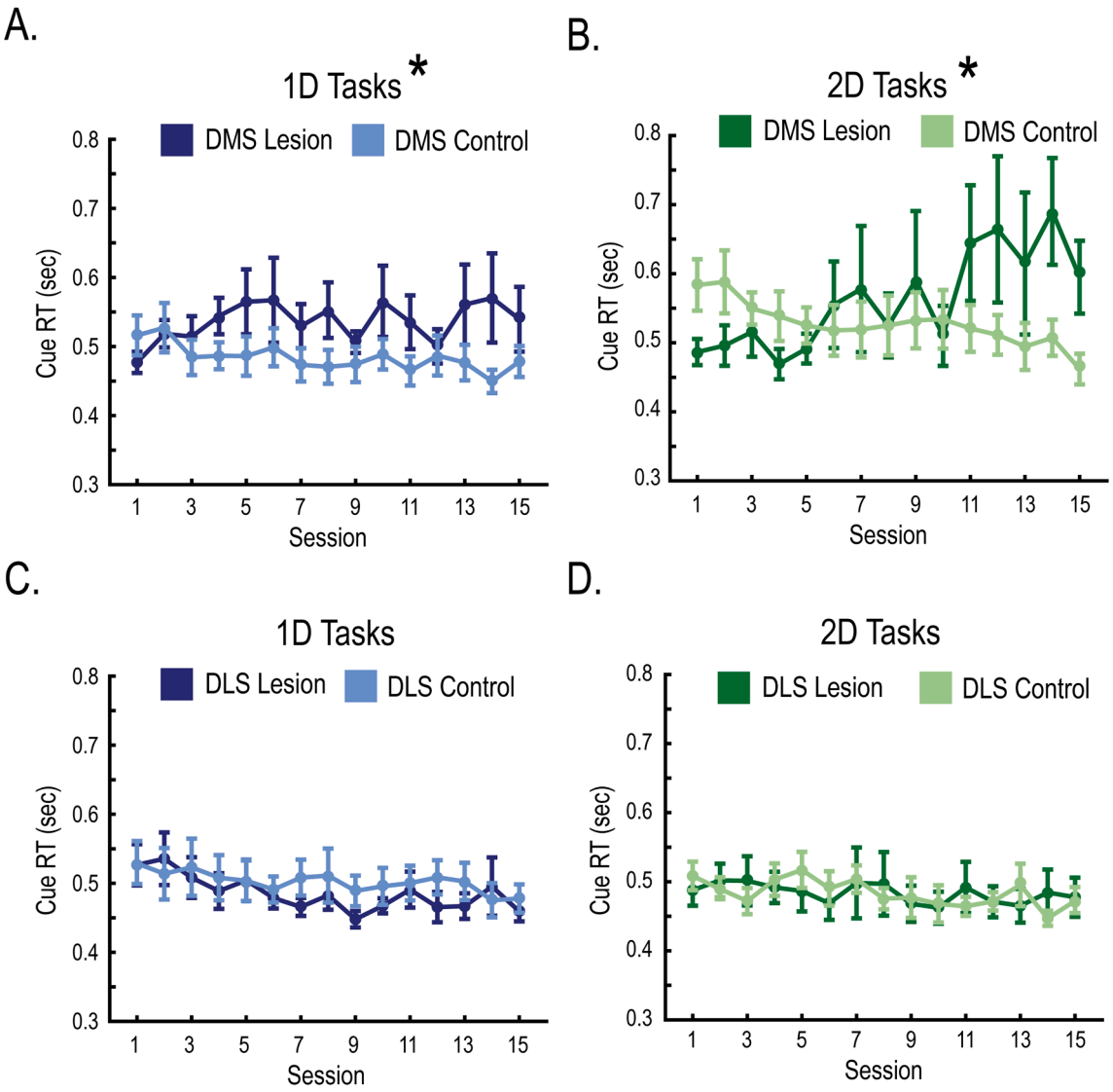


Fig. 4. Lesions of the DMS, but not DLS, impaired Cue RT during category learning. **A-B,** Rats with DMS lesions had larger Cue RT (i.e., mean time to observe and categorize each stimulus) compared to controls. This was true for rats learning the 1D tasks (A) and rats learning the 2D tasks (B). **C-D,** Cue RT for DLS rats learning the 1D tasks (C) and the 2D tasks (D). There were no significant differences in Cue RT between rats with DLS lesions and controls. '*' indicates statistical significance. All error bars indicate *SEM*.

types at equal rates (Broschard et al., 2019). For rats with DMS lesions, Cue RT, and not Choice RT, was significantly longer than controls (Fig. 4A-B & Supplementary Fig. 1A-B). This was true for both task types (1D tasks: Cue RT- $t(52.10) = -2.28$, $p = .045$, Choice RT- $t(54.30) = -0.76$, $p = .448$; 2D tasks: Cue RT- $t(59.81) = -2.50$, $p = .015$, Choice RT- $t(49.02) = -0.27$, $p = .789$). These results suggest that rats with DMS lesions required more time to observe and categorize each stimulus but did not require more time to make category decisions. Therefore, the lesions did not impair overall motor behavior but targeted processes critical for categorization. Conversely, lesioning the DLS had no effect on reaction time compared to controls (Fig. 4C-D & Supplementary Fig. 1C-D; 1D tasks: Cue RT- $t(51.50) = 1.11$, $p = .273$, Choice RT- $t(54.30) = -0.76$, $p = .448$; 2D tasks: Cue RT- $t(50.48) = 0.23$, $p = .816$ Choice RT- $t(47.60) = 0.29$, $p = .775$). Finally, rats with DMS lesions had significantly larger Cue RT than rats with DLS lesions (1D tasks: $t(54.14) = 2.34$, $p = .040$; 2D tasks: $t(60.61) = -2.65$, $p = .012$). Together, these results support our finding that the DMS, but not the DLS, is critical for rat category learning.

3.4. Lesions of the DMS, but not the DLS, impair category generalization

Each rat was given five testing sessions to examine generalization to novel stimuli. Testing stimuli were configured into a grid that sampled from the entire stimulus space. We first examined performance across the stimulus space by generating heatmaps that averaged the accuracy of each stimulus within the grid (Fig. 5A&C). Each grid was rotated so that all task types had the same orientation (i.e., the relevant axis was perpendicular to the x-axis, and the irrelevant axis was perpendicular to the y-axis). Accuracy was largely affected by the distances along the relevant axis. Specifically, accuracy was higher for stimuli farther from the category boundary, and accuracy was lower for stimuli closer to the category boundary. Accuracy was unaffected by distances along the irrelevant axis. Compared to controls, accuracy was impaired across the entire stimulus space for rats with lesions of the DMS (Fig. 5A), but not rats with lesions of the DLS (Fig. 5C).

Linear mixed effects modeling was used to analyze generalization performance quantitatively. Stimuli from the testing sessions were segregated into three trial types. Trained stimuli were sampled from regions of the stimulus space that overlapped with the training

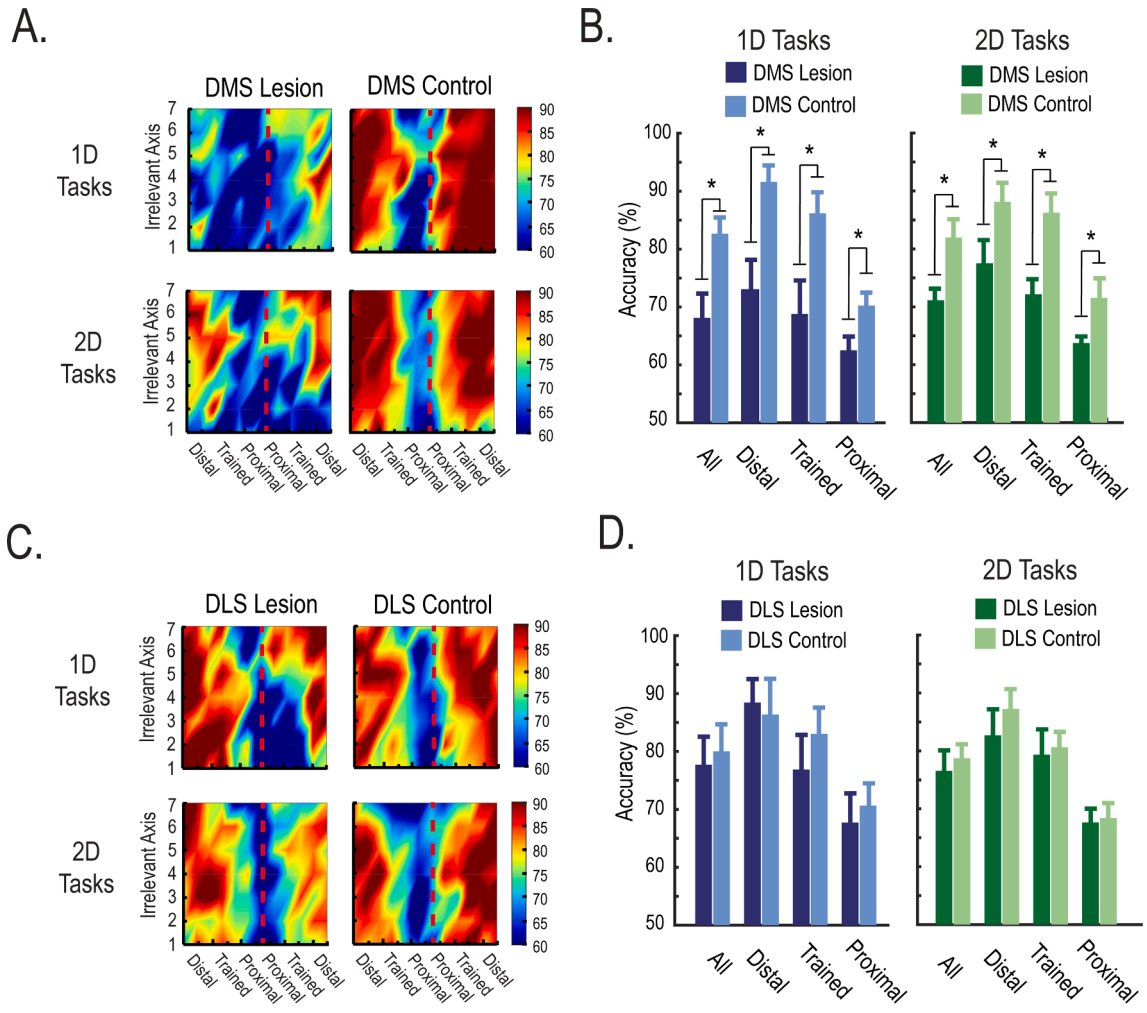


Fig. 5. Each rat was presented five testing sessions to examine category generalization. Testing stimuli were configured into a grid containing 84 stimuli. A third of these stimuli overlapped with the training distributions (Trained), whereas the remaining stimuli sampled from novel portions of the stimulus space. Of these novel stimuli, half were close to the category boundary (Proximal) and half were far from the category boundary (Distal). Generally, performance increased for stimuli farther from the category boundary. **A-B,** Mean accuracy across trial types for DMS rats that learned the 1D tasks and the 2D tasks. Compared to controls, rats with DMS lesions had lower accuracy for both the 1D tasks and the 2D tasks. There were no significant interactions across trial types. **C-D,** Mean accuracy across trial types for DLS rats that learned the 1D tasks and the 2D tasks. There were no significant differences in accuracy between rats with DLS lesions and controls. Together, these results suggest that the DMS, but not the DLS, was critical for category generalization. '*' indicates statistical significance. All error bars indicate the *SEM*.

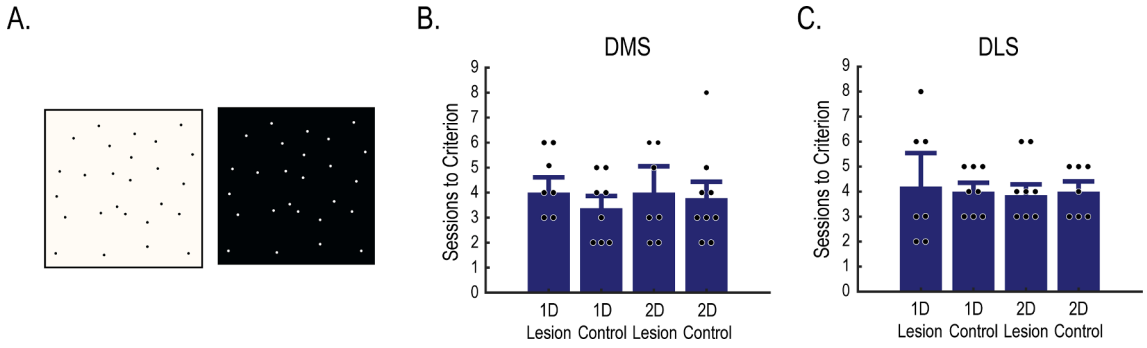


Fig. 6. Control discrimination task. **A,** Rats were presented training sessions to learn to discriminate two stimuli (i.e., a black box and a white box). This task acted as a control and was used to ensure that the impairments observed during category learning were specific to processes related to categorization and were not caused by confounding factors (e.g., deficits in motivation, motor activity, perception, etc.). **B-C,** All rats were given training sessions until they reached at least 75% accuracy for both stimuli on two consecutive sessions. There were no significant differences in the number of training sessions to reach this learning criterion across groups. This was true for rats with DMS lesions (B) and rats with DLS lesions (C). These results suggest that the observed impairments during category training were not caused by deficits unrelated to categorization. All error bars indicate the *SEM*.

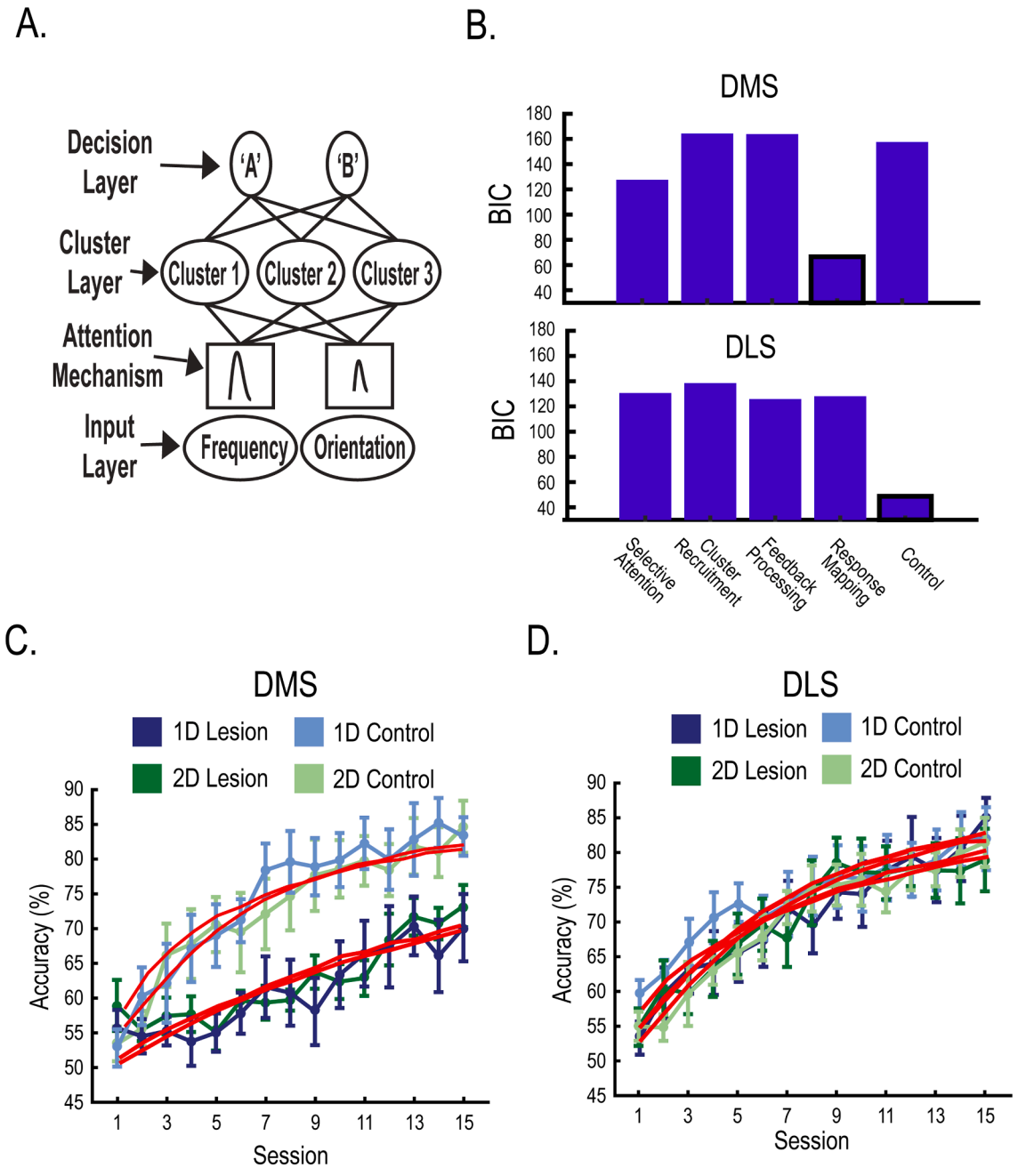


Fig. 7. SUSTAIN model fitting. **A,** Diagram of the neural network model SUSTAIN, which contains three distinct layers: an input layer, a cluster layer, and a decision layer. The input layer loads the current stimulus, the cluster layer compares that stimulus to existing category representations, and the decision layer determines the category membership of the current stimulus. An attention mechanism amplifies category-relevant stimulus before it is compared to the cluster representations. **B,** Multiple models were designed to test the underlying deficit produced by the lesions. Each model disrupted a single component of the baseline model and were compared by calculating BIC values. The training data of rats with DMS lesions (top) were best simulated when it was assumed that the DMS is critical for mapping cluster representations to appropriate category responses. This was simulated by reducing the learning rate between the cluster layer and the decision layer. Conversely, the training data of rats with DLS lesions (bottom) were best simulated when it was assumed that the DLS is not necessary for category learning (i.e., the control model). **C-D,** SUSTAIN's fits for the best fitting models for rats with DMS lesions (C; Response Mapping) and rats with DLS lesions (D; Control). Red lines indicate SUSTAIN's fit. All error bars indicate the *SEM*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distributions (i.e., within two standard deviations), Proximal stimuli were sampled from novel portions of the stimulus space close to the category boundary, and Distal stimuli were sampled from novel portions of the stimulus space far from the category boundary (Fig. 1C-D). The full linear models contained fixed effects for experimental group, trial type, and a quadratic function across trial types, and random effects for slope, intercept, and the quadratic function. As expected, accuracy was related to the distance from the category boundary. Specifically, accuracy was significantly impaired for Proximal stimuli compared to Trained stimuli (Fig. 5B&D; $t(96.00) = 5.36$, $p < .001$), indicating that generalization became more difficult for stimuli close to the category boundary. Conversely, there was no significant difference between Trained stimuli and Distal stimuli ($t(96.00) = 1.44$, $p = .153$), suggesting that rats could easily generalize to novel stimuli far from the category boundary.

Accuracy was equivalent for controls that learned the 1D tasks and

the 2D tasks ($t(48.00) = -0.15$, $p = .880$), supporting the finding that rats learned the task types at the same rate and to equal levels. Contrary to category training, accuracy was significantly lower for female rats compared to male rats ($t(54.00) = -2.10$, $p = .040$). However, there were no significant interactions across trial types (all $p > .05$), implying that the overall accuracy degraded between category training sessions and testing sessions for females more than males.

Rats with DMS lesions had impaired accuracy compared to controls for both task types (Fig. 5B; 1D tasks: $t(48.00) = -3.07$, $p = .004$; 2D tasks; $t(48.00) = -1.99$, $p = .045$). There were no significant interactions across trial types (all $p > .05$), suggesting that overall accuracy was impaired. Lesioning the DLS had no effect on category generalization (Fig. 5D; 1D tasks: $t(64.75) = -1.15$, $p = .253$; 2D tasks: $t(64.75) = -0.26$, $p = .800$). Finally, accuracy impairments were significantly larger for rats with DMS lesions compared to rats with DLS lesions (1D tasks: $t(49.50) = -3.10$, $p = .003$; 2D tasks: $t(49.50) = -2.30$, $p = .012$). Together, these results support our finding that categorization behavior is mediated through the DMS.

3.5. Lesions of the dorsal striatum do not affect learning a control discrimination task

Finally, rats were given training sessions to learn a control discrimination task. All training procedures were equivalent to the categorization sessions except that only two stimuli were presented (rather than distributions of stimuli). Sessions continued until rats reached a learning criterion (i.e., at least 75 % accuracy for both stimuli on two consecutive sessions). Using 2x2 ANOVAs, there were no significant differences in the number of sessions for each group to reach the learning criterion (Fig. 6B-C; DMS: $F(3,34) = 1.09$, $p = .367$; DLS: $F(3,32) = 0.89$, $p = .457$). This is especially informative for rats with DMS lesions, which exhibited large accuracy impairments during category training. These results suggest that the impairments during categorization sessions were likely not attributable to unrelated processes, such as deficits in perception, motor behavior, and motivation.

3.6. The DMS learns to map category representations to appropriate responses

The neural network SUSTAIN was used to investigate the neural functions of the dorsal striatum in rat category learning (Fig. 7A). This was accomplished by using a model comparison approach. Briefly, a baseline model was created by fitting SUSTAIN to the averaged learning data of the control groups. Then, multiple experimental models were designed by disrupting single components of the baseline model. Each experimental model assumed that the lesions produced a unique deficit during category learning (see Material & Methods for a complete description of each model). Experimental models were fit to the averaged learning data of the lesion groups. The quality of each fit was compared by calculating the BIC value for each model. The underlying function of each brain region was inferred from the experimental model with the smallest BIC (the best fitting model).

The BIC values for each experimental model are shown in Fig. 7B. For the rats with DMS lesions, the best fitting model was the Response Mapping model which assumed that the lesion groups had difficulty associating cluster representations to appropriate category responses (Fig. 7C). This was simulated by reducing the learning rate parameter that updates the connection weights between the cluster layer and the decision layer. For the rats with DLS lesions, the best fitting model was the Control Model, which assumed that the lesions had no effect on category learning (Fig. 7D). Together, these results suggest that the DMS is critical for mapping responses to category labels, whereas the DLS is not necessary for categorization.

3.7. Lesions of the DMS, but not the DLS, impair perceptual recency effects

The current experimental design lends itself well to trial-based analyses, since each rat completes hundreds of training trials. We have found that category decisions are influenced by the identity of the most recent training stimulus (Broschard et al., 2021). Specifically, accuracy on a given trial is facilitated if the current stimulus is perceptually similar to the last stimulus (i.e., a positive recency score; see Material & Methods), and accuracy is impaired if the current stimulus is dissimilar to the last stimulus (i.e., a negative recency score). These trial order effects imply that rats are regularly updating category representations to include recent training exemplars. Importantly, inactivating the rodent prelimbic prefrontal cortex or the dorsal hippocampus impair these trial order effects (i.e., recency scores near 0; Broschard et al., 2021), suggesting that these regions are critical for updating category representations (Love & Gureckis, 2007). Here, we repeated this analysis with the current dataset. If we assume the DMS is critical for mapping representations to behavioral responses, then lesioning the DMS would impair the expression of these category updates on a trial-by-trial basis.

We first replicated the trial order effects in controls (Fig. 8). Specifically, recency scores were significantly above 0 for high similarity pairs and significantly below 0 for low similarity pairs (all $p < .05$). Compared to controls, recency scores were significantly impaired for rats with DMS lesions (Fig. 8A-B; $F(3,34) = 3.38$, $p = .030$), but not rats with DLS lesions (Fig. 8C-D; $F(3,32) = 1.90$, $p = .148$). The interaction between striatal subregions was significant, suggesting that the perceptual recency effects were significantly larger for rats with DMS lesions compared to rats with DLS lesions ($ps < 0.05$). We predict that category representations are updated through an interaction between the prefrontal cortex and the hippocampus (Love & Gureckis, 2007) and then these updates are expressed through the DMS.

4. Discussion

Human models of category learning (e.g., COVIS) posit that the head and tail of the caudate nucleus contribute unique functions to category learning through independent corticostrial loops (Ashby & Maddox, 2005; Seger, 2008). The current experiment tested these predictions by selectively lesioning the rodent homologs of these regions (i.e., DMS and DLS) while rats learned to categorize visual stimuli. Rats with DMS lesions were impaired on both the 1D tasks and the 2D tasks (Figs. 3-5). These rats were not impaired on a control discrimination task that used similar trial procedures (Fig. 6), suggesting that the impairments during categorization sessions were likely not related to deficits in motor behavior, motivation, or perception. Conversely, lesioning the DLS had no effect on category learning or category generalization. Together, these results suggest that the DMS serves a general role in category learning, whereas the DLS is not necessary for categorization. These results conflict with the COVIS framework or suggest that the functional organization of the basal ganglia differs between rats and humans.

To better understand the function of the DMS during category learning, model simulations were conducted using the neural network SUSTAIN. This involved designing and fitting multiple experimental models to the rat data; each model tested a potential function of the striatum (e.g., selective attention, response mapping, feedback processing, etc.). Model fits were compared by calculating BIC values for each model; the model with the smallest BIC value was assumed to indicate the underlying deficit produced by the lesion. Results suggest that the DMS is critical for associating category representations with appropriate category responses (Fig. 7). This was simulated by reducing the learning rate parameter between the cluster layer and the decision layer. These results align with Shohamy et al., 2004, who successfully simulated data from patients with Parkinson's Disease by reducing a learning rate parameter. For rats with DLS lesions, the best fitting model was the control model, which assumed that the lesions had no effect on

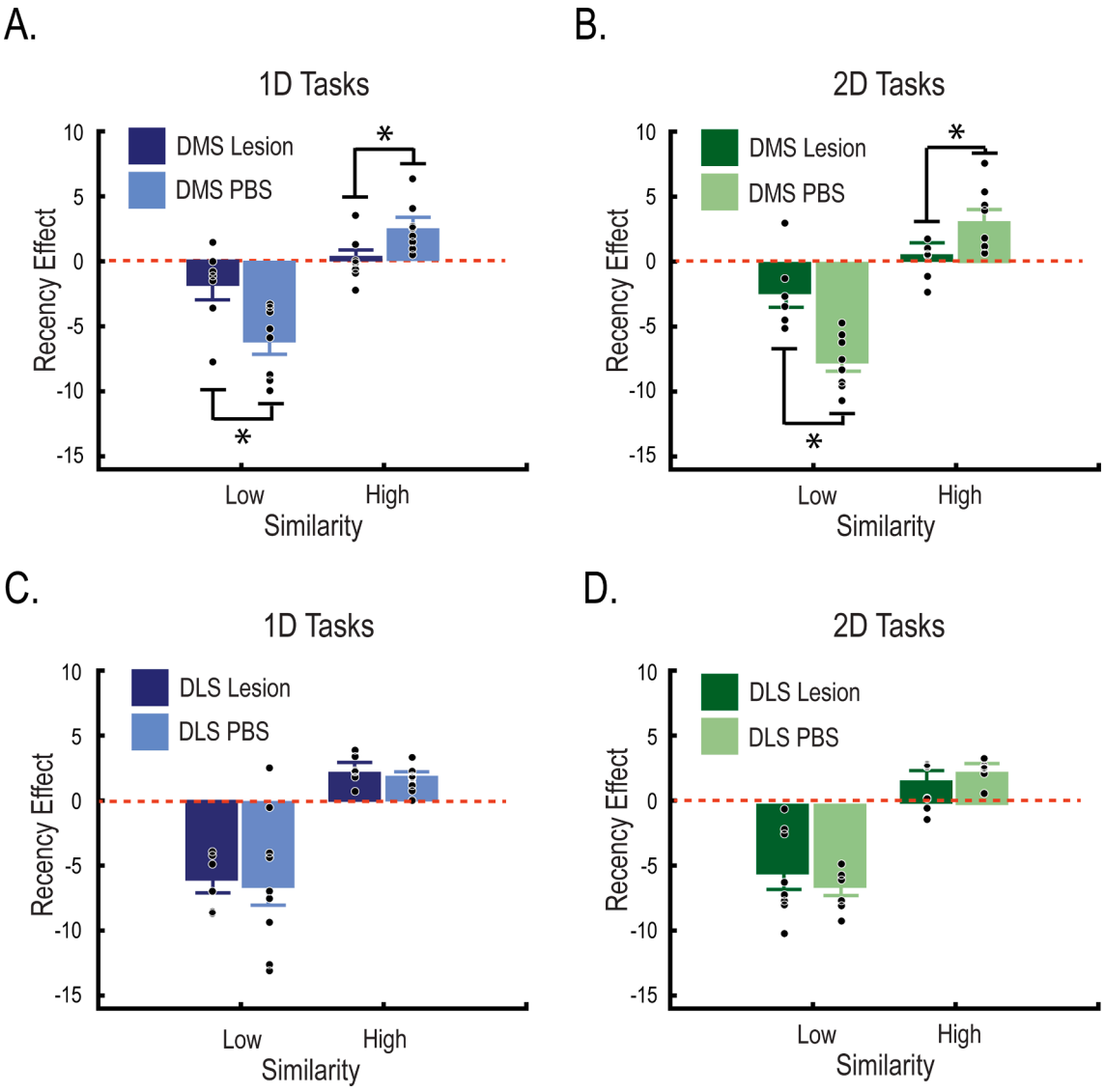


Fig. 8. Perceptual recency effects. Trial accuracy was binned according to the perceptual similarity between the current stimulus and the most recent stimulus. These binned accuracies were subtracted from iterations where trial order was randomized. For all control groups, accuracy was facilitated if the current stimulus was perceptually similar to the previous trial (i.e., a positive recency score). Accuracy was impaired if the current stimulus was perceptually dissimilar to the previous trial (i.e., a negative recency score). These trial order effects suggest that rats are regularly updating category representations, since their decisions are biased towards the most recent training experience. **A-B,** Lesioning the DMS impaired these trial order effects for rats learning the 1D tasks (A) and the 2D tasks (B). **C-D,** Lesioning the DLS did not affect the trial order effects compared to controls. This was true for rats learning the 1D tasks (C) and rats learning the 2D tasks (D). All error bars indicate the *SEM*.

category learning. Together, these results suggest that the DMS serves a general role in category learning by mapping category representations to behavioral responses, and the DLS does not contribute to category learning.

Multiple models of category learning posit that category representations are maintained by the hippocampus (Mack, Love & Preston, 2018; Schapiro, Turk-Browne, Botvinick, & Norman, 2017; Broschard, 2021). This hypothesis is supported by studies demonstrating that the hippocampus creates 'cognitive maps' of non-spatial multidimensional stimulus spaces (Solomon, Lega, Sperling, & Kahana, 2019; Aronov, Nevers, & Tank, 2017). Accordingly, we predict that category learning in rats involves an interaction between the hippocampus and the DMS, where the hippocampus builds category representations, and the DMS connects these representations to appropriate responses. This interpretation conforms to the general conceptualization of the learning mechanisms achieved by these regions (Packard, 1999; McDonald & White, 2013); namely, the hippocampus learns relationships among stimuli (i.e., S—S learning) and the striatum learns relationships between stimuli and responses (i.e., S-R learning). There are many examples of functional communication between the striatum and the hippocampus; commonly, the DMS produces goal-directed behavior by utilizing task-relevant information organized by the hippocampus (Delcasso et al., 2014; Devan & White, 1999; Fouquet, Babayan, Watilliaux, Bontempi, Tobin, & Rondi-Reig, 2013; Johnson, Van der Meer, & Redish, 2007; Goodroe, Starnes, & Brown. 2018; DeCoteau et al., 2007). We hypothesize that this interaction underlies a fundamental mechanism of category learning in rats.

The functional interaction between the hippocampus and the DMS may be observed indirectly in Fig. 8. Here, we analyzed category performance on a trial-by-trial basis as a proxy for examining single-trial updates to category representations. For controls, accuracy was affected by the perceptual similarity between the current stimulus and the last stimulus, suggesting that rats regularly update category representations. Importantly, these trial order effects were impaired in rats

with DMS lesions, but not rats with DLS lesions (Fig. 8), suggesting that the DMS may be important for updating the category representations after each trial. Broschard et al., 2021 and Broschard, 2021 found similar impairments in rats with inactivation of the prefrontal cortex or hippocampus, respectively. Assuming the hippocampus is important for maintaining category representations, we predict that representational updates originate in the hippocampus (potentially through an interaction with the prelimbic prefrontal cortex). Then, these updates are expressed through the DMS. Future experiments can test these predictions directly through projection-specific inactivation.

The current results conflict with the predictions of the COVIS model, which posits that the 1D tasks and 2D tasks are mediated by separate striatal subregions (Ashby et al., 1998). Instead, we found that the DMS mediates learning for both task types. An interesting alternative explanation is that the task differences described by COVIS do not exist between striatal subregions but instead lie within smaller subdivisions of the DMS in rodents. A recent anatomical study comprehensively mapped the corticostriatal system in mice (Foster et al., 2021). The rodent striatum is much more heterogeneous than previously thought; within the DMS exists multiple smaller domains that have distinct connectivity profiles. Critically, the DMS domains that receive prefrontal input are separable from domains that receive visual input (Foster et al., 2021, Fig. 4). Following the COVIS framework, it is possible that the domains receiving prefrontal input are responsible for learning the 1D tasks, whereas the domains receiving visual input are responsible for learning the 2D tasks. Future experiments can test this hypothesis directly by targeting each domain within the DMS separately.

Lesioning the DLS did not affect categorization in the current experiment; however, this does not rule out the possibility that the DLS is critical for learning other categorization tasks. For example, the current experiment used a supervised learning paradigm, where learning was driven through consistent food reinforcement. This design may have favored the recruitment of the DMS, which is important for selecting behaviors that produce a desired outcome (in this case, a food reward; Yin, Knowlton, & Balleine, 2004; Turner, Svegborn, Langguth, McKenzie, & Robbins, 2021; Vandaele, Ottenheimer, & Janak, 2021). The DLS, on the other hand, mediates behaviors that are more habitual and less sensitive to consistent feedback (Yin, Ostlund, Knowlton, & Balleine, 2005; Crego et al., 2020; Lipton, Gonzales, & Citri, 2019). Perhaps the DLS is more critical to learn task designs that do not rely of receiving explicit feedback, such as unsupervised paradigms (Broker, Love, & Dayan, 2021; Love, 2002; Lake, Vallabha, & McClelland, 2008).

To conclude, the current experiment adds to a growing literature that examines the neural mechanisms underlying rat category learning (Broschard et al., 2019; Broschard et al., 2020; Broschard et al., 2021; Broschard, 2021). We combined selective lesions and computational modeling approaches to understand the roles of the rodent dorsal striatum in category learning. Lesioning the DMS produced impairments for both 1D tasks and 2D tasks, while lesioning the DLS had no effect on categorization. Model simulations suggest that the DMS serves a general role in category learning by learning to associate category representations with appropriate behavioral responses. We hypothesize that the DMS participants in a larger circuit with the prelimbic prefrontal cortex and dorsal hippocampus that supports category learning in rats. Future experiments will examine these interactions in more detail.

Funding information

National Institutes of Health Grant P01-HD080679 to J.H.F. and B.C. L. and Wellcome Trust Investigator Award WT106931MA to B.C.L.

CRediT authorship contribution statement

Matthew B. Broschard: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Writing - original draft, Writing - review & editing. **Jangjin Kim:** Software, Methodology, Writing - review & editing. **Bradley C. Love:** Funding acquisition, Methodology, Writing - review & editing. **John H. Freeman:** Supervision, Funding acquisition, Resources, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nlm.2023.107732.

References

Antzoulatos, E., & Miller, E. (2011). Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. *Neuron, 71*(2), 243–249. https://doi.org/10.1016/j.neuron.2011.05.040

Antzoulatos, E., & Miller, E. (2014). Increases in functional connectivity between prefrontal cortex and striatum during category learning. *Neuron, 83*(1), 216–225. https://doi.org/10.1016/j.neuron.2014.05.005

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature, 543*(7647), 719–722. https://doi.org/10.1038/nature21692

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*(3), 442–481. https://doi.org/10.1037//0033-295X.105.3.442

Ashby, F. G., & Ennis, J. M. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation, 1–36*. https://doi.org/10.1016/s0079-7421(06)46001-1

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., Noble, S., Filoteo, J. V., Waldron, E. M., & Ell, S. W. (2003). Category learning deficits in parkinson's disease. *Neuropsychology, 17*(1), 115–124. https://doi.org/10.1037/0894-4105.17.1.115

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences, 9*(2), 83–89. https://doi.org/10.1016/j.tics.2004.12.003

Ashby, F. G., Smith, J. D., & Rosedahl, L. A. (2020). Dissociations between rule-based and information-integration categorization are not caused by differences in task difficulty. *Memory & Cognition, 48*(4), 541–552. https://doi.org/10.3758/s13421-019-00988-4

Balleine, B. W., & O'Doherty, J. P. (2009). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology, 35*(1), 48–69. https://doi.org/10.1038/npp.2009.131

Balsters, J. H., Zerbi, V., Sallet, J., Wenderoth, N., & Mars, R. B. (2020). Primate homologs of mouse cortico-striatal circuits. *eLife, 9*, e53680.

Bröker, F., Love, B. C., & Dayan, P. (2021). When unsupervised training benefits category learning. 10.31234/osf.io/k5pzu.

Broschard, M. B., Kim, J., Love, B. C., & Freeman, J. H. (2020). Category learning in rodents using touchscreen-based tasks. *Genes, Brain and Behavior.*. https://doi.org/10.1111/gbb.12665

Broschard, M. B., Kim, J., Love, B. C., Wasserman, E. A., & Freeman, J. H. (2019). Selective attention in rat visual category learning. *Learning & Memory, 26*(3), 84–92. https://doi.org/10.1101/lm.048942.118

Broschard, M. B., Kim, J., Love, B. C., Wasserman, E. A., & Freeman, J. H. (2021). Prelimbic cortex maintains attention to category-relevant information and flexibly updates category representations. *Neurobiology of Learning and Memory, 185*, Article 107524. https://doi.org/10.1016/j.nlm.2021.107524

Broschard, M.B. (2021, November 8-11). *Hippocampus stores category representations in rats*. Society for Neuroscience, virtual meeting. https://www.sfn.org/meetings/neuroscience-2021.

Brown, R., & Marsden, C. (1988). Internal versus external cues and the control of attention in parkinson's disease. *Brain, 111*(2), 323–345. https://doi.org/10.1093/brain/111.2.323

Carpenter, K. L., Wills, A. J., Benattayallah, A., & Milton, F. (2016). A comparison of the neural correlates that underlie rule-based and information-integration category learning. *Human Brain Mapping, 37*(10), 3557–3574. https://doi.org/10.1002/hbm.23259

Crawley, M. J. (2007). *The R book*. Chichester: John Wiley & Sons Ltd.

Crego, A. C., Štoček, F., Marchuk, A. G., Carmichael, J. E., Van der Meer, M. A., & Smith, K. S. (2020). Complementary control over habits and behavioral vigor by phasic activity in the dorsolateral striatum. *The Journal of Neuroscience, 40*(10), 2139–2153. https://doi.org/10.1523/jneurosci.1313-19.2019

Crijns, E., & Op de Beeck, H. (2019). The Visual Acuity of Rats in Touchscreen Setups. *Vision, 4*(1), 4. https://doi.org/10.3390/vision4010004

DeCoteau, W. E., Thorn, C., Gibson, D. J., Courtemanche, R., Mitra, P., Kubota, Y., et al. (2007). Learning-related coordination of striatal and hippocampal theta rhythms during acquisition of a procedural maze task. *Proceedings of the National Academy of Sciences, 104*(13), 5644–5649. https://doi.org/10.1073/pnas.0700818104

Delcasso, S., Huh, N., Byeon, J. S., Lee, J., Jung, M. W., & Lee, I. (2014). Functional relationships between the hippocampus and dorsomedial striatum in learning a visual scene-based memory task in rats. *Journal of Neuroscience, 34*(47), 15534–15547. https://doi.org/10.1523/jneurosci.0622-14.2014

Devan, B. D., & White, N. M. (1999). Parallel information processing in the dorsal striatum: Relation to hippocampal function. *The Journal of Neuroscience, 19*(7), 2789–2798. https://doi.org/10.1523/jneurosci.19-07-02789.1999

Filoteo, J. V., Maddox, W. T., Ing, A. D., & Song, D. D. (2007). Characterizing rule-based category learning deficits in patients with parkinson's disease. *Neuropsychologia, 45*(2), 305–320. https://doi.org/10.1016/j.neuropsychologia.2006.06.034

Filoteo, J. V., Maddox, W. T., Ing, A. D., Zizak, V., & Song, D. D. (2005). The impact of irrelevant dimensional variation on rule-based category learning in patients with parkinson's disease. *Journal of the International Neuropsychological Society, 11*(05). https://doi.org/10.1017/s1355617705050617

Filoteo, J. V., Maddox, W. T., Salmon, D. P., & Song, D. D. (2005). Information-integration category learning in patients with striatal dysfunction. *Neuropsychology, 19*(2), 212–222. https://doi.org/10.1037/0894-4105.19.2.212

Filoteo, J. V., Maddox, W. T., & Davis, J. D. (2001). A possible role of the striatum in linear and nonlinear category learning: Evidence from patients with Hungtington's disease. *Behavioral Neuroscience, 115*(4), 786–798. https://doi.org/10.1037/0735-7044.115.4.786

Foster, N. N., Barry, J., Korobkova, L., et al. (2021). The mouse cortico–basal ganglia–thalamic network. *Nature, 598*, 188–194. https://doi.org/10.1038/s41586-021-03993-3

Fouquet, C., Babayan, B. M., Watilliaux, A., Bontempi, B., Tobin, C., & Rondi-Reig, L. (2013). Complementary roles of the hippocampus and the dorsomedial striatum during spatial and sequence-based navigation behavior. *PLoS ONE, 8*(6). https://doi.org/10.1371/journal.pone.0067232

Goodroe, S. C., Starnes, J., & Brown, T. I. (2018). The complex nature of hippocampal-striatal interactions in spatial navigation. *Frontiers in Human Neuroscience, 12*. https://doi.org/10.3389/fnhum.2018.00250

Heilbronner, S. R., Rodriguez-Romaguera, J., Quirk, G. J., Groenewegen, H. J., & Haber, S. N. (2016). Circuit-Based Corticostriatal Homologies Between Rat and Primate. *Biological psychiatry, 80*(7), 509–521. https://doi.org/10.1016/j.biopsych.2016.05.012

Johnson, A., Van der Meer, M. A., & Redish, A. D. (2007). Integrating Hippocampus and striatum in decision-making. *Current Opinion in Neurobiology, 17*(6), 692–697. https://doi.org/10.1016/j.conb.2008.01.003

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of experimental psychology. Learning, memory, and cognition, 32*(2), 316–332.

Kim, J., Castro, L., Wasserman, E. A., & Freeman, J. H. (2018). Dorsal hippocampus is necessary for visual categorization in rats. *Hippocampus, 28*(6), 392–405. https://doi.org/10.1002/hipo.22839

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A Neostriatal Habit Learning System in Humans. *Science, 273*(5280), 1399–1402. https://doi.org/10.1126/science.273.5280.1399

Lake, B. M., Vallabha, G. K., & McClelland, J. L. (2008). Modeling unsupervised perceptual category learning. *2008 7th IEEE International Conference on Development and Learning*. 10.1109/devlrn.2008.4640800.

Lipton, D. M., Gonzales, B. J., & Citri, A. (2019). Dorsal striatal circuits for habits, compulsions and addictions. *Frontiers in Systems Neuroscience, 13*. https://doi.org/10.3389/fnsys.2019.00028

Lombardi, W. J., Andreason, P. J., Sirocco, K. Y., Rio, D. E., Gross, R. E., Umhau, J. C., et al. (1999). Wisconsin Card Sorting Test performance following head injury: Dorsolateral Fronto-striatal circuit activity predicts perseveration. *Journal of Clinical and Experimental Neuropsychology, 21*(1), 2–16. https://doi.org/10.1076/jcen.21.1.2.940

Lopez-Paniagua, D., & Seger, C. A. (2011). Interactions within and between corticostriatal loops during component processes of category learning. *Journal of Cognitive Neuroscience, 23*(10), 3068–3083. https://doi.org/10.1162/jocn_a_00008

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review, 9*(4), 829–835. https://doi.org/10.3758/bf03196342

Love, B. C., & Gureckis, T. M. (2007). Models in search of a brain. *Cognitive, Affective, & Behavioral Neuroscience, 7*(2), 90–108. https://doi.org/10.3758/cabn.7.2.90

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). Sustain: A network model of category learning. *Psychological Review, 111*(2), 309–332. https://doi.org/10.1037/0033-295x.111.2.309

Mack, M. L., Love, B. C., & Preston, A. R. (2016). *Dynamic updating of hippocampal object representations reflects new conceptual knowledge.*. https://doi.org/10.1101/071118

Mack, M. L., Love, B. C., & Preston, A. R. (2018). Building concepts one episode at a time: The hippocampus and concept formation. *Neuroscience Letters, 680*, 31–38. https://doi.org/10.1016/j.neulet.2017.07.061

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nat Commun, 11*(46). https://doi.org/10.1038/s41467-019-13930-8

Maddox, W. T., Aparicio, P., Marchant, N. L., & Ivry, R. B. (2005). Rule-based category learning is impaired in patients with parkinson's disease but not in patients with cerebellar disorders. *Journal of Cognitive Neuroscience, 17*(5), 707–723. https://doi.org/10.1162/0898929053747630

McDonald, R. J., & White, N. M. (2013). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience, 127*(6), 835–853. https://doi.org/10.1037/a0034883

Milton, F., & Pothos, E. M. (2011). Category structure and the two learning systems of COVIS. *European Journal of Neuroscience, 34*(8), 1326–1336. https://doi.org/10.1111/j.1460-9568.2011.07847.x

Neath, A. A., & Cavanaugh, J. E. (2011). The Bayesian information Criterion: Background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics, 4*(2), 199–203. https://doi.org/10.1002/wics.199

Nomura, E., Maddox, W., Filoteo, J., Ing, A., Gitelman, D., Parrish, T., … Reber, P. (2006). Neural Correlates of Rule-Based and Information-Integration Visual Category Learning. *Cerebral Cortex, 17*(1), 37–43. https://doi.org/10.1093/cercor/bhj122

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*(1), 39–57. https://doi.org/10.1037/0096-3445.115.1.39

Owen, A. M., Roberts, A. C., Hodges, J. R., & Robbins, T. W. (1993). Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or parkinson's disease. *Brain, 116*(5), 1159–1175. https://doi.org/10.1093/brain/116.5.1159

O'Donoghue, E. M., Broschard, M. B., & Wasserman, E. A. (2020). Pigeons exhibit flexibility but not rule formation in dimensional learning, stimulus generalization, and task switching. *Journal of Experimental Psychology: Animal Learning and Cognition, 46*(2), 107–123. https://doi.org/10.1037/xan0000234

Packard, M. G. (1999). Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. *Proceedings of the National Academy of Sciences, 96*(22), 12881–12886. https://doi.org/10.1073/pnas.96.22.12881

Paxinos, G., & Watson, C. (1998). *The Rat Brain in Stereotaxic Coordinates*. San Diego: Academic Press.

Price, A. L. (2006). Explicit category learning in parkinson's disease: Deficits related to impaired rule generation and selection processes. *Neuropsychology, 20*(2), 249–257. https://doi.org/10.1037/0894-4105.20.2.249

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Seger, C. A. (2008). How do the basal ganglia contribute to categorization? their roles in generalization, response selection, and learning via feedback. *Neuroscience & Biobehavioral Reviews, 32*(2), 265–278. https://doi.org/10.1016/j.neubiorev.2007.07.010

Seger, C. A. (2013). The visual corticostriatal loop through the tail of the caudate: Circuitry and function. *Frontiers in Systems Neuroscience, 7*. https://doi.org/10.3389/fnsys.2013.00104

Seger, C. A., & Cincotta, C. M. (2002). Striatal activity in concept learning. *Cognitive, Affective & Behavioral Neuroscience, 2*(2), 149–161. https://doi.org/10.3758/CABN.2.2.149

Seger, C. A., & Cincotta, C. M. (2005). The roles of the caudate nucleus in human classification learning. *Journal of Neuroscience, 25*(11), 2941–2951. https://doi.org/10.1523/jneurosci.3401-04.2005

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience, 33*(1), 203–219. https://doi.org/10.1146/annurev.neuro.051508.135546

Shohamy, D., Myers, C. E., Onlaor, S., & Gluck, M. A. (2004). Role of the basal ganglia in category learning: How do patients with parkinson's disease learn? *Behavioral Neuroscience, 118*(4), 676–686. https://doi.org/10.1037/0735-7044.118.4.676

Shohamy, D., Myers, C., Kalanithi, J., & Gluck, M. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience & Biobehavioral Reviews, 32*(2), 219–236. https://doi.org/10.1016/j.neubiorev.2007.07.008

Solomon, E. A., Lega, B. C., Sperling, M. R., & Kahana, M. J. (2019). Hippocampal theta codes for distances in semantic and temporal spaces. *Proceedings of the National Academy of Sciences, 116*(48), 24343–24352. https://doi.org/10.1073/pnas.1906729116

Soto, F. A., Waldschmidt, J. G., Helie, S., & Ashby, F. G. (2013). Brain activity across the development of automatic categorization: A comparison of categorization tasks using multi-voxel pattern analysis. *NeuroImage, 71*, 284–297. https://doi.org/10.1016/j.neuroimage.2013.01.008

Swainson, R., Sengupta, D., Shetty, T., Watkins, L., Summers, B., Sahakian, B., et al. (2006). Impaired dimensional selection but intact use of reward feedback during visual discrimination learning in parkinson's disease. *Neuropsychologia, 44*(8), 1290–1304. https://doi.org/10.1016/j.neuropsychologia.2006.01.028

Turner, K. M., Svegborn, A., Langguth, M., McKenzie, C., & Robbins, T. W. (2022). Opposing roles of the dorsolateral and dorsomedial striatum in the acquisition of skilled action sequencing in rats. *The Journal of Neuroscience, 42*(10), 2039–2051. https://doi.org/10.1523/jneurosci.1907-21.2022

Vandaele, Y., Ottenheimer, D. J., & Janak, P. H. (2021). Dorsomedial striatal activity tracks completion of behavioral sequences in rats. *Eneuro, 8*(6). https://doi.org/10.1523/eneuro.0279-21.2021

Villagrasa, F., Baladron, J., Vitay, J., Schroll, H., Antzoulatos, E. G., Miller, E. K., et al. (2018). On the role of cortex-basal ganglia interactions for category learning: A

neurocomputational approach. *The Journal of Neuroscience, 38*(44), 9551–9562. https://doi.org/10.1523/jneurosci.0874-18.2018

Volz, H., Gaser, C., Hager, F., Rzanny, R., Mentzel, H., Kaiser, W., et al. (1998). Brain activation during cognitive stimulation with the Wisconsin card sorting test—a functional MRI study on healthy volunteers and Schizophrenics. *Schizophrenia Research, 29*(1–2), 110. https://doi.org/10.1016/s0920-9964(97)88579-9

West, M. O., Carelli, R. M., Pomerantz, M., Cohen, S. M., Gardner, J. P., Chapin, J. K., et al. (1990). A region in the dorsolateral striatum of the rat exhibiting single-unit correlations with specific locomotor limb movements. *Journal of Neurophysiology, 64*(4), 1233–1246.

Yahya, K. (2020). The basal ganglia corticostriatal loops and conditional learning. *Reviews in the Neurosciences, 32*(2), 181–190. https://doi.org/10.1515/revneuro-2020-0047

Yin, H., & Knowlton, B. (2006). The role of the basal ganglia in habit formation. *Nature ReviewsNeurosci, 7*, 464–476. https://doi.org/10.1038/nrn1919

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2004). Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience, 19*(1), 181–189. https://doi.org/10.1111/j.1460-9568.2004.03095.x

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience, 22*(2), 513–523. https://doi.org/10.1111/j.1460-9568.2005.04218.x

Zeithamova, D., Mack, M. L., Braunlich, K., Davis, T., Seger, C. A., Van Kesteren, M. T., et al. (2019). Brain mechanisms of concept learning. *The Journal of Neuroscience, 39*(42), 8259–8266. https://doi.org/10.1523/jneurosci.1166-19.2019