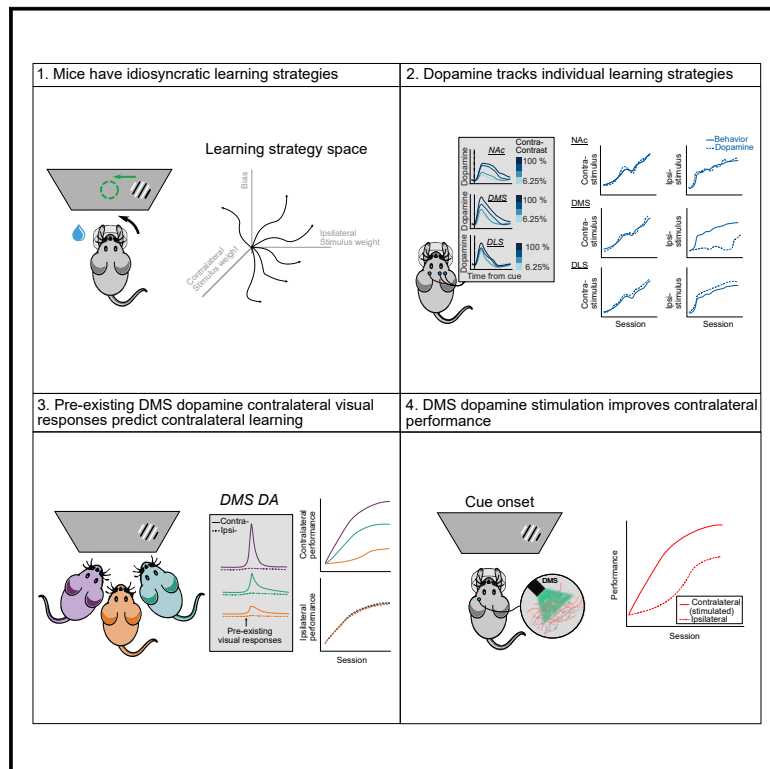


# Pre-existing visual responses in a projection-defined dopamine population explain individual learning trajectories

## Graphical abstract



## Authors

Alejandro Pan-Vazquez,  
Yoel Sanchez Araujo,  
Brenna McMannon, ...,  
Jonathan W. Pillow, Nathaniel D. Daw,  
Ilana B. Witten

## Correspondence

ndaw@princeton.edu (N.D.D.),  
iwitten@princeton.edu (I.B.W.)

## In brief

Pan-Vazquez, Sanchez Araujo, et al. show how pre-existing contralateral visual responses in dopamine terminals in the dorsomedial striatum help explain side-specific learning trajectories of a visuomotor task. These results suggest that feature- and projection-specific dopamine signals could simplify the problem of learning a new task.

## Highlights

- Dopaminergic responses track side-specific learning trajectories in a visuomotor task
- Dopamine in the DMS has pre-existing responses to contralateral visual stimuli
- Pre-existing DMS dopaminergic responses predict learning trajectories for contralateral stimuli
- Activation of dopamine terminals in the DMS improves contralateral performance



## Report

# Pre-existing visual responses in a projection-defined dopamine population explain individual learning trajectories

Alejandro Pan-Vazquez,<sup>1,5</sup> Yoel Sanchez Araujo,<sup>1,5</sup> Brenna McMannon,<sup>1</sup> Miranta Louka,<sup>1</sup> Akhil Bandi,<sup>1</sup> Laura Haetzel,<sup>1</sup> Mayo Faulkner,<sup>2</sup> International Brain Laboratory, Jonathan W. Pillow,<sup>1</sup> Nathaniel D. Daw,<sup>1,3,\*</sup> and Ilana B. Witten<sup>1,4,6,\*</sup>

<sup>1</sup>Princeton Neuroscience Institute, Princeton University, Washington Road, Princeton, NJ 08540, USA

<sup>2</sup>UCL Institute of Ophthalmology, University College London, 11-43 Bath Street, London EC1V 9EL, UK

<sup>3</sup>Department of Psychology, Princeton University, Washington Road, Princeton, NJ 08540, USA

<sup>4</sup>Howard Hughes Medical Institute, Princeton University, Washington Road, Princeton, NJ 08540, USA

<sup>5</sup>These authors contributed equally

<sup>6</sup>Lead contact

\*Correspondence: [ndaw@princeton.edu](mailto:ndaw@princeton.edu) (N.D.D.), [iwitten@princeton.edu](mailto:iwitten@princeton.edu) (I.B.W.)

<https://doi.org/10.1016/j.cub.2024.09.045>

## SUMMARY

A key challenge of learning a new task is that the environment is high dimensional—there are many different sensory features and possible actions, with typically only a small reward-relevant subset. Although animals can learn to perform complex tasks that involve arbitrary associations between stimuli, actions, and rewards,<sup>1–6</sup> a consistent and striking result across varied experimental paradigms is that in initially acquiring such tasks, large differences between individuals are apparent in the learning process.<sup>7–12</sup> What neural mechanisms contribute to initial task acquisition, and why do some individuals learn a new task much more quickly than others? To address these questions, we recorded longitudinally from dopaminergic (DA) axon terminals in mice learning a visual decision-making task.<sup>7</sup> Across striatum, DA responses tracked idiosyncratic and side-specific learning trajectories, consistent with widespread reward prediction error coding across DA terminals. However, even before any rewards were delivered, contralateral-side-specific visual responses were present in DA terminals, primarily in the dorsomedial striatum (DMS). These pre-existing responses predicted the extent of learning for contralateral stimuli. Moreover, activation of these terminals improved contralateral performance. Thus, the initial conditions of a projection-specific and feature-specific DA signal help explain individual learning trajectories. More broadly, this work suggests that functional heterogeneity across DA projections may serve to bias target regions toward learning about different subsets of task features, providing a potential mechanism to address the dimensionality of the initial task learning problem.

## RESULTS

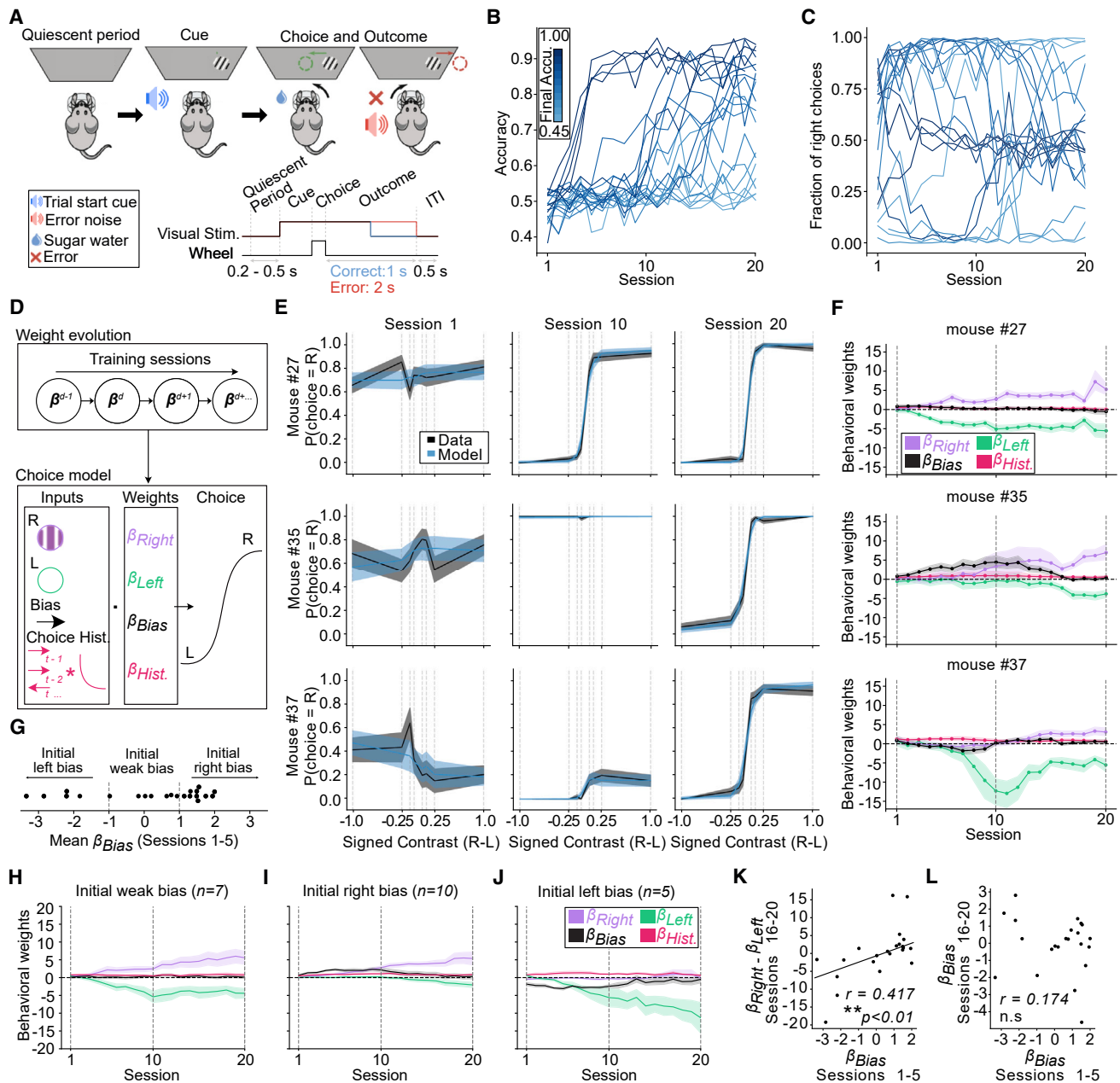
### Idiosyncratic and side-specific learning of a standardized visual decision-making task

To study individual differences in learning, we leveraged a standardized visual decision-making task.<sup>7,13–15</sup> In this task, at the beginning of each trial, a visual grating was presented on a screen on either the right or left side. Mice received a reward by turning a steering wheel with their front paws in the direction that centered the grating on the screen (Figures 1A and S1A–S1C). Correct wheel turns were rewarded with 3  $\mu$ L sucrose water, whereas incorrect ones led to a short timeout (2 s) and white noise (0.5 s). Across trials, the visual gratings varied in contrast (100%, 25%, 12%, and 6.25%), with contrasts randomly and uniformly selected. To maximize interpretability of the neural recordings over the course of learning, we did not use a shaping procedure,<sup>7</sup> a debiasing protocol,<sup>7,16</sup> or other experimenter interventions.

Mice showed a large degree of variability in their learning trajectories, both in terms of the number of sessions to reach high accuracy (Figure 1B) as well as in their probability of choosing left vs. right wheel turns (Figure 1C). Whereas some mice selected both options (right and left) to an equal extent from the beginning of training, others appeared to prefer one side or the other.

To quantify choices across learning, we constructed a behavioral model (Figure 1D) that described each session's contrast-dependent choices (i.e., psychometric curve) based on weights that evolved across sessions: a weight on the visual stimulus contrast on the left ( $\beta_{left}$ ) or right ( $\beta_{right}$ ) that captured the stimulus-dependent tendency to make left or right choices, a choice bias ( $\beta_{bias}$ ) that captured the tendency to choose one side or the other, irrespective of the stimulus, and a choice history regressor ( $\beta_{hist}$ ) that captured the tendency to repeat previous choices (see STAR Methods for model details). This model successfully reproduced the diverse psychometric curves observed across





**Figure 1. Idiosyncratic and side-specific learning trajectories**

(A) Schematic of the task. On each trial, a Gabor patch of a different contrast (6.12%, 12.5%, 25%, or 100%) is presented on the right or left side of a screen. Centering the patch with a steering wheel leads to a small water reward, whereas moving it out of the screen results in a short timeout (2 s) and white noise (0.5 s).

(B) Accuracy (fraction of trials rewarded) across training sessions.

(C) Probability of right choices across training sessions. In (B) and (C), each line represents one mouse, colored by their mean accuracy in sessions 16–20.

(D) Schematic of the behavioral model. Choice (left or right) on each trial is predicted with a logistic function based on weighting the contrast of the right and left visual stimulus ( $\beta_{right}$  and  $\beta_{left}$ ), a bias term ( $\beta_{bias}$ ) coded such that positive values indicate rightward choice, and a choice history kernel. Weights evolved across sessions (see STAR Methods for details).

(E) Psychometric curves (“data”) and model fits (“model”) from 3 example mice on the first, middle, and last session of training. Lines and shading represent mean  $\pm$  SEM.

(F) Model weights across training for the same mice from (E). Lines and shading represent mean and 95% confidence intervals.

(G) Early  $\beta_{bias}$  (average of sessions 1–5) for all the mice, showing the subdivisions used in subsequent panels between mice with weak, left, or right initial bias.

(H–J) Average trajectories of bias, right and left stimulus weights across training for mice subdivided by their initial bias as shown in (G). Lines and shading represent mean  $\pm$  SEM across mice.

(K and L) Relationship between early  $\beta_{bias}$  (average of sessions 1–5) and the late difference in stimulus sensitivity weights ( $\beta_{right} - \beta_{left}$  for sessions 16–20).  $r = 0.417$ ,  $p = 0.0007$ .

(legend continued on next page)

learning (Figure 1E; 3 example sessions from 3 example mice) and allowed us to isolate the contribution of each variable to each mouse's behavior across learning (Figure 1F).

The model revealed that whereas some mice learned to increasingly weight visual stimuli similarly on both sides as training progressed (e.g., show an increase across training in both  $\beta_{right}$  and  $\beta_{left}$  in Figure 1F, top), many instead displayed idiosyncratic learning trajectories where they preferentially weighted one stimulus vs. the other (e.g., large  $\beta_{right}$  vs. small  $\beta_{left}$  in Figure 1F, middle; large  $\beta_{left}$  vs. small  $\beta_{right}$  in Figure 1F, bottom). This side-specific learning is consistent with the choice asymmetries evident in the raw data (Figure 1C), though note that the model enables distinguishing between stimulus-contrast sensitivity (the stimulus weight) and a stimulus-independent choice bias. The side specificity of stimulus learning at the end of the training ( $\beta_{right} - \beta_{left}$ ) was predicted by the bias parameter  $\beta_{bias}$  at the beginning of training, as evidenced both by plotting the stimulus weights based on groups defined by the level of initial bias (Figures 1G–1J) and by correlating initial bias with the difference between the final stimulus sensitivity weights (Figure 1K). By contrast, early  $\beta_{bias}$  did not predict final  $\beta_{bias}$  (Figure 1L).

Based on these observations, we concluded that mice display idiosyncratic side-specific learning trajectories that could be predicted by initial bias. We next explored how individual differences in striatal dopaminergic (DA) signals might explain these individual learning patterns.

### Across the striatum, contrast-dependent DA visual responses tracked side-specific individual learning trajectories

We recorded striatal DA signals longitudinally over the course of learning using fiber photometry. To ensure consistent expression of the activity indicator across time and individuals, we used double-transgenic mice that expressed GCaMP6f in DA neurons (DAT-Cre  $\times$  Ai148). Before training, optical fibers were implanted into the following striatal subregions of each mouse: dorsomedial striatum (DMS), dorsolateral striatum (DLS), and nucleus accumbens core (NAc; Figures 2A and S2A–S2C). Each subregion was recorded unilaterally, with a mixture of left and right hemispheres across subregions within each mouse and across mice for each subregion. During each session for each mouse, we recorded simultaneously from all 3 subregions.

Given that DA neurons encode reward-predictive cues,<sup>17</sup> behavioral and DA sensitivity to the stimuli should be expected to increase during task acquisition, as animals learn that the stimuli are predictive of reward.<sup>8,9,18–21</sup> To test this, we quantified DA sensitivity to the visual stimuli by fitting the normalized fluorescence data with a linear encoding model for each session and subregion (model schematic in Figure S3A). This allowed us to estimate stimulus response kernels (Figure 2B), which reflect the contribution of the visual stimulus to the neural signals while (linearly) accounting for other task events (actions and outcomes). This model could accurately capture moment-by-

moment fluctuations in fluorescence on each trial (Figures 2C and S3B) as well as the trial-averaged PSTHs (Figures S3C and S3D).

Averaged across mice, across all regions, the magnitude of these stimulus responses increased across sessions, as mice learned that the stimuli were predictive of reward (Figure 2D; L2-norm of stimulus response kernels). Moreover, stronger stimulus responses emerged to the stronger contrasts, consistent with stronger contrasts becoming more predictive of reward as the animals acquired the task (i.e., increasing stimulus-sensitivity behavioral weight with training in Figures 1F–1J). Whereas stimulus responses in DLS and NAc dopamine were similar for contralateral and ipsilateral stimuli, the DMS had much stronger stimulus responses for contralateral than ipsilateral stimuli (Figure 2D).<sup>22,23</sup>

How do these stimulus responses relate to each mouse's idiosyncratic and side-specific learning profiles (Figure 1)? For each mouse, across sessions, we correlated the contrast dependence of the *behavioral* trajectories for stimuli on one side (stimulus contrast weight from the behavioral model; Figure 1) with the contrast dependence of the *neural* trajectories for stimuli on the same side (difference in L2-norm for highest vs. lowest contrast stimulus kernel). There was a strong correlation between the behavioral and neural measures across all regions (example trajectories, Figure 2E; neural and behavioral correlations for all mice, Figure 2F). Whereas DLS and NAc dopamine showed this correlation for both ipsilateral and contralateral stimuli, in the DMS, this correlation was only apparent for contralateral stimuli (Figure 2F), presumably reflecting the contralateral specificity of the stimulus responses themselves (Figure 2D).

Thus, across the striatum, we observed the co-evolution during task acquisition of side-specific dopamine and behavioral sensitivity to the visual stimuli, consistent with reward prediction error signaling in DA neurons.

### Pre-existing visual responses in DMS DA terminals predict learning on the contralateral side

Although the striatum-wide correlations between DA trajectories and learning trajectories were striking (Figure 2), they do not clarify whether there are DA signals that *precede* behavioral changes that could potentially explain individual differences in task acquisition. Notably, the DMS (unlike the other regions) exhibited contrast-dependent responses to the visual stimuli from the very first session (Figure 2D). To determine whether these signals existed before training or whether they emerged during the first session, we examined DA responses during an earlier pre-exposure session (“session 0,” before the first session) in which the visual gratings were presented but mice did not receive rewards nor could they turn the wheel (Figure 3A).

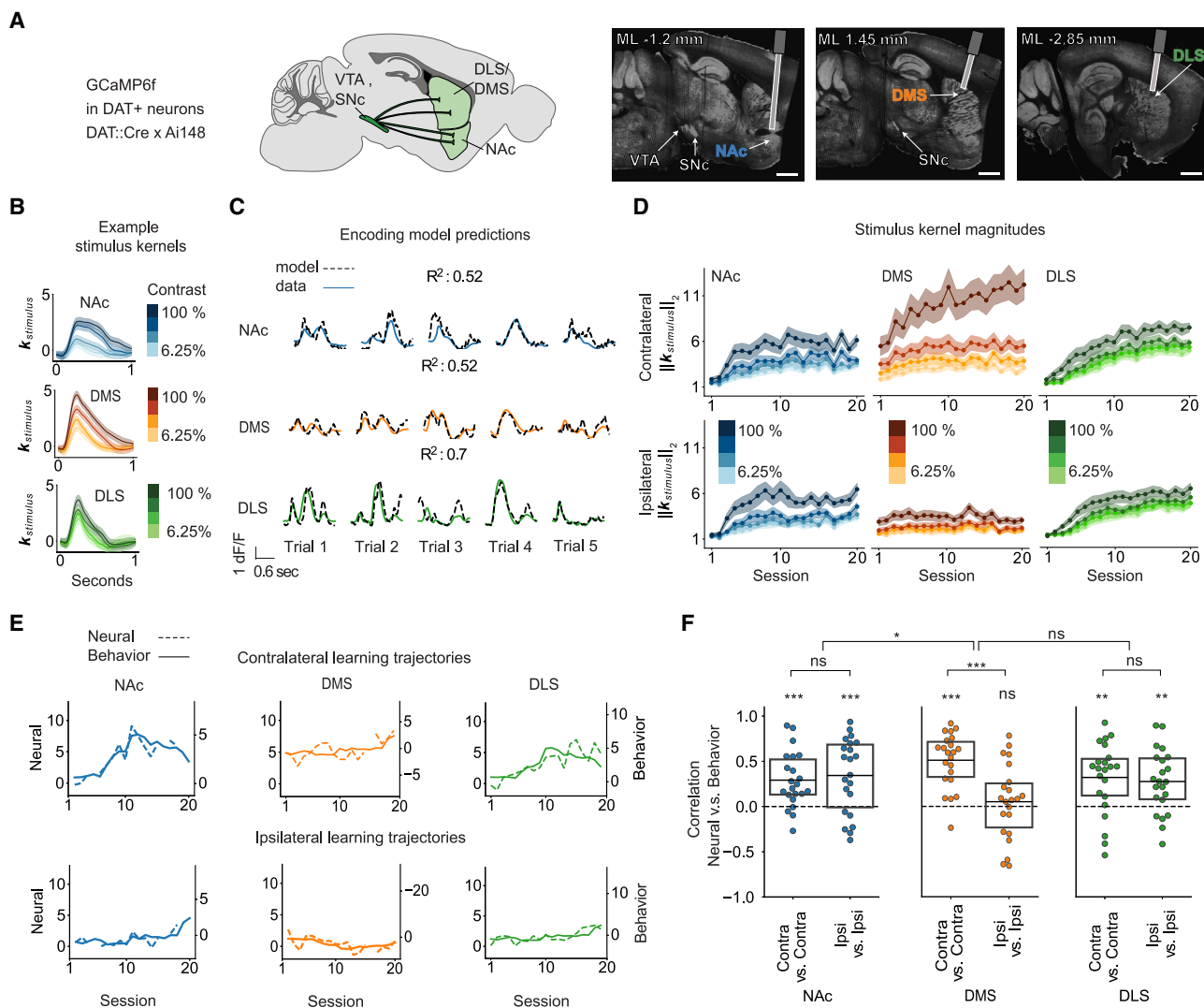
In the DMS, but not the NAc or DLS, DA terminals had contrast-dependent visual responses during stimulus pre-exposure before training (session 0; DMS, Figures 3B and 3C; NAc and DLS, Figure S4). These DMS DA responses were contralateral-side

(I) Relationship between early  $\beta_{bias}$  (sessions 1–5) and late  $\beta_{bias}$  (sessions 16–20).  $R = 0.174$ ,  $p = 0.522$ .

In (K) and (L), each dot is a mouse. Correlation and  $p$  values from robust regression. \*\* $p < 0.01$ ; ns, not significant.

Across all panels,  $n = 22$  mice.

See also Figure S1.



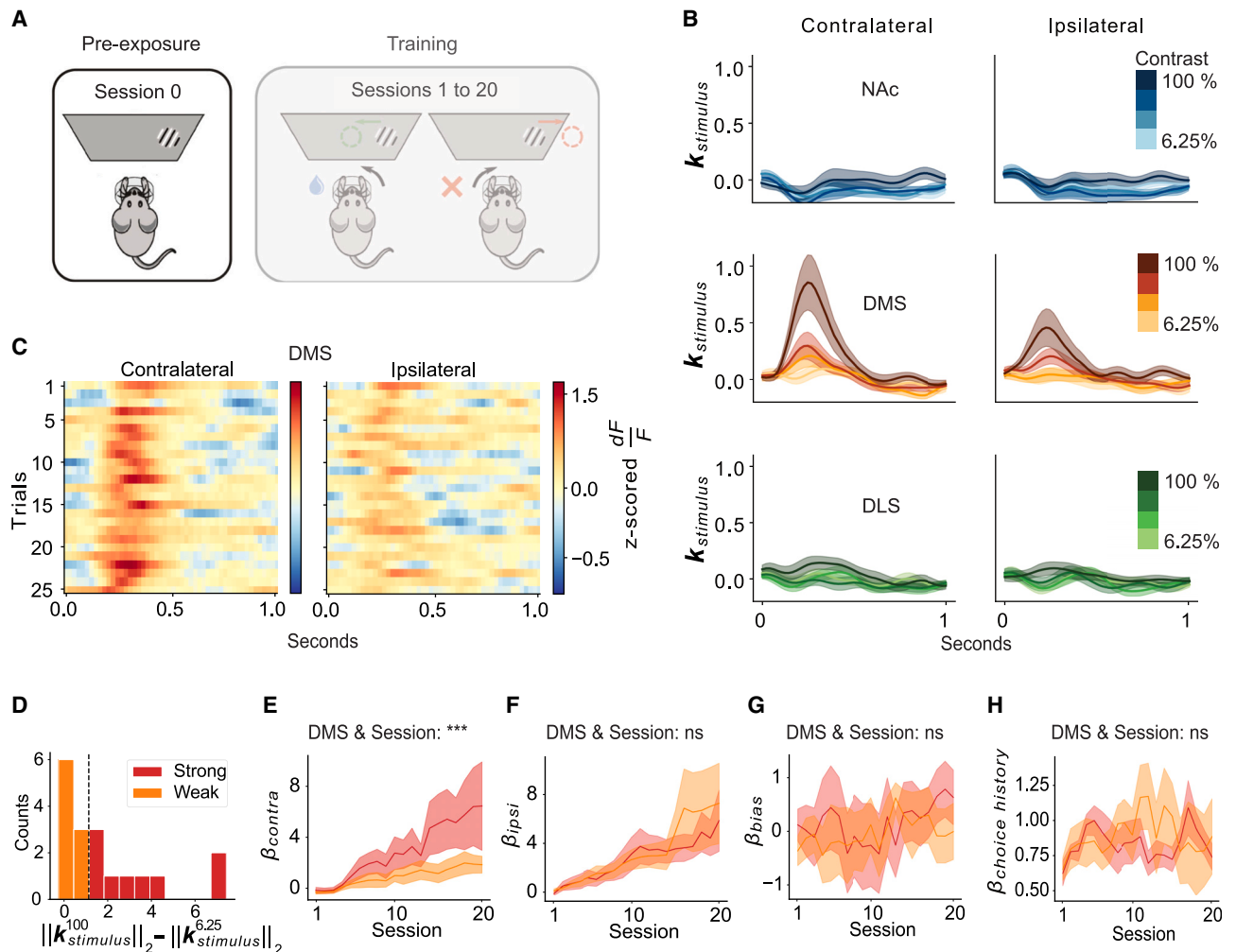
**Figure 2. In DA terminals across the striatum, contrast-dependent visual responses track individual side-specific learning trajectories**  
(A) Experimental strategy used for collecting the fiber photometry data from DA terminals. Left: schematic of the recorded projections using the GCaMP6f x DAT::Cre mouse line. Right: example histology. Scale bar, 1 mm.  
(B) Contralateral stimulus response kernels from an example mouse on an example session.  
(C) Z scored dF/F (solid line) and predictions from the encoding model (dashed line) on 5 different trials for an example mouse on an example session.  $R^2$  is the variance explained across the session within all trial epochs (from stimulus onset to 1 s after feedback).  
(D) Stimulus response magnitudes (L2-norm) in each region and session, averaged across mice, for contralateral (top) and ipsilateral (bottom) stimuli. Lines and shading represent mean  $\pm$  SEM.  
(E) Trajectories of the contrast-dependence of neural stimulus response magnitudes (“neural”; difference in L2-norm for 100% and 6.25% contrast) and the behavioral stimulus choice weights (“behavioral”) for contralateral (top) and ipsilateral (bottom) stimuli (from an example mouse in which DMS is recorded on one hemisphere and DLS/NAc on the other).  
(F) Correlations of the neural and behavioral trajectories as shown in (E).  $p$  values calculated with  $t$  tests. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ; ns, not significant. See Table S1 for statistical details for (F).  $n = 22$  mice in (D) and (F). See also Figures S2 and S3.

specific, similar to the pattern observed throughout training (Figure 2D). The stronger response to higher contrast stimuli may be interpreted as a salience-related signal.<sup>18,24–35</sup> Although dopamine has been associated with novelty coding,<sup>18,32,34–37</sup> these pre-existing visual responses did not attenuate during the 25 presentations of each stimulus (Figure 3C).

A previous theoretical account<sup>38</sup> suggests that novelty- or salience-related DA signals could provide animals with a “bonus”

(or head start) in forming stimulus-reward associations. Therefore, we wondered whether variability in these DMS-specific pre-existing DA responses to the visual stimuli might predict side-specific differences across individuals in learning trajectories.

Indeed, we observed a striking relationship between these pre-exposure stimulus responses in DMS DA and individual differences in learning about stimuli on the side contralateral to the recording. For visualization purposes, we median-split mice into



**Figure 3. Pre-existing visual responses in DMS DA terminals predict side-specific learning trajectories**

(A) Schematic of the stimulus pre-exposure session before training (“session 0”).

(B) Stimulus response kernels in the NAc, DMS, and DLS for contralateral and ipsilateral stimuli of each contrast, averaged across mice, during session 0 (pre-exposure). Lines and shading represent mean  $\pm$  SEM.

(C) Heatmap of stimulus responses on session 0 to 100% contrast stimuli in the DMS for the first 25 trials, averaged across mice.

(D) Histogram across mice of contrast-dependent contralateral stimulus responses on session 0, quantified as the difference in the L2-norm of the highest and lowest contrast contralateral stimulus, colored by a median split.

(E) Contralateral stimulus sensitivity weights from the behavioral model for mice with strong vs. weak contrast-dependent stimulus responses during session 0 (subdivision of mice shown in D). Lines and shading represent mean  $\pm$  SEM. \*\*\* $p < 0.001$  for the interaction between DMS stimulus response on session 0 and session in a two-way ANOVA (see [Data S1.1–S3.2](#) for model details and full results).

(F) Same as (E), except for the ipsilateral stimulus weight from the behavioral model. No significant interaction (ns) between DMS stimulus response on session 0 and session (see [Data S1.3–S3.4](#) for model details and full results).

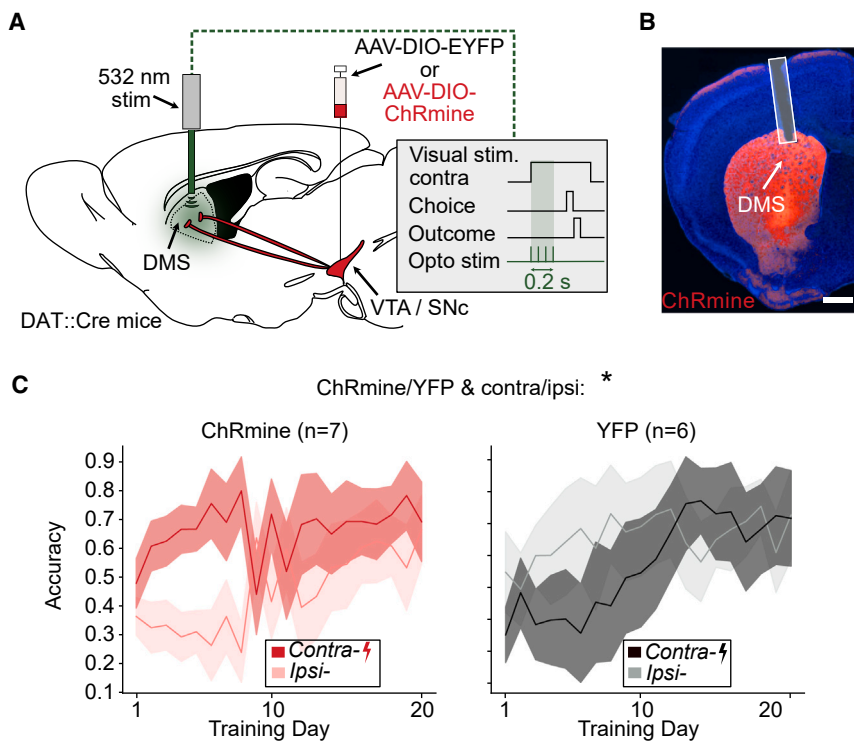
(G) Same as (E) and (F), but for the bias weights from the behavioral model (transformed such that positive means contralateral bias). No significant interaction (ns) between DMS stimulus response on session 0 and session (see [Data S1.5–S3.6](#) for model details and full results). In all panels,  $n = 18$  mice.

(H) Same as (G), but for the choice history weights from the behavioral model. No significant interaction (ns) between DMS stimulus response on session 0 and session (see [Data S1.7](#) and [3.8](#) for model details and full results).

See also [Figures S2, S4, and S5](#) and [Data S1](#).

2 groups based on their pre-exposure visual responses ([Figure 3D](#); “strong” vs. “weak” contrast-dependent contralateral DMS DA responses on session 0) and plotted the trajectory of the behavioral model weights during subsequent task training in each group ([Figures 3E–3H](#)). Over the course of training, the animals with larger pre-exposure visual stimulus responses developed larger contralateral behavioral stimulus sensitivity weights ([Figure 3E](#)).

By contrast, the pre-exposure DMS DA visual responses were not predictive of the bias, choice history, or ipsilateral stimulus sensitivity behavioral weights ([Figures 3F–3H](#)). Moreover, in the NAc and DLS, pre-exposure visual responses did not significantly correlate with behavioral weight trajectories ([Figures S4 and S5](#)), consistent with the very weak or non-existent pre-exposure visual responses in those regions ([Figure 3B](#)).



**Figure 4. Stimulating DMS DA terminals at the onset of contralateral stimulus presentation improves side-specific performance**

(A) Schematic of the optogenetic stimulation of DMS DA terminals. Mice either expressed ChRmine or a control construct in DA neurons. DA terminals in the DMS were optogenetically stimulated unilaterally (532 nm, 0.2 s burst duration, 5 ms pulse width, 20 Hz pulses,  $\sim 0.25$  mW) at the onset of the contralateral stimulus presentation throughout training.

(B) Example histology image of optical fiber location and terminal expression of ChRmine-mScarlet. Scale bar, 900  $\mu$ m.

(C) Comparison of performance for contralateral and ipsilateral stimulus trials in control ( $n = 7$ , left) and ChRmine ( $n = 6$ , right) mice. Lines and shading represent mean  $\pm$  SEM.  $*p < 0.05$  for cohort (ChRmine/YFP) and side (contra/ipsi) interaction in three-way ANOVA with cohort (ChRmine/YFP), day, and side (contralateral/ipsilateral) as factors (see Table S2 for model details and full results).

See also Figures S1 and S2 and Table S2.

Thus, pre-existing visual responses in DMS dopamine (but not NAc or DLS dopamine) predict individual differences in learning about the contralateral visual stimuli, even before any training.

### Activation of DMS DA terminals during contralateral stimulus presentation improves side-specific performance

The strength of pre-existing visual responses in DMS DA terminals predicted side-specific learning (Figures 3D, 3E, and S5). Could these visual signals causally impact side-specific performance during learning? Although it is clear that DA signals at the time of outcome reinforce previous actions,<sup>39–43</sup> whether DA signals at the time of a preceding cue impact learning is less clear.<sup>33,39,40,44–46</sup>

To test for a causal role of visual-stimulus-related DA signals in the DMS, we performed brief unilateral optogenetic stimulation of DMS dopamine terminals at the presentation of the contralateral visual stimulus throughout training (Figures 4A, 4B, and S2D). The stimulation, which terminated before the outcome period (Figures S1D and S1E), led to a significant improvement in performance between contralateral vs. ipsilateral trials for the opsin vs. control groups, which grew over early training (Figure 4C). We therefore concluded that DA signals in the DMS could improve performance for contralateral stimuli.

## DISCUSSION

As mice learn to perform a visual decision-making task, variation across individuals in learning trajectories closely tracks visual responses in DA terminals across striatum, consistent with

widespread reward prediction error coding (Figures 1 and 2). In contrast to these striatum-wide patterns, prior to any reward or training, pre-existing contralateral-side-preferring visual responses are present in DMS DA terminals, and these signals predict and help explain contralateral-side-specific learning trajectories (Figures 3 and 4). This work is significant in suggesting that (1) the initial conditions of the DA system are important in explaining individual differences in learning and (2) feature- and projection-specific DA signals could be a mechanism to simplify the problem of initial task acquisition.

### Pre-existing visual responses in DMS-projecting DA neurons could serve to simplify initial task acquisition

A major reason that initial task acquisition is challenging is the issue of credit assignment: in virtually any task, even nominally simple ones, there are multiple possible dimensions of the environment that an animal could try to learn about, with most being irrelevant to reward. In the case of the visual decision-making task used here, mice need to learn that the side-specific relationship between visual stimuli and actions is what leads to reward, whereas other stimuli, actions, or stimulus-action relationships do not (including many high-dimensional, uncontrolled incidental features of the environment<sup>47</sup>).

Our data suggest that pre-existing side-specific and projection-specific visual DA responses could serve to decrease the dimensionality of this learning problem. Different striatal subcircuits could be initialized to more easily learn reward associations for different subsets of stimuli or actions. Although our data reveal that the pre-existing response in DA terminals in the DMS to contralateral visual stimuli explains learning for stimuli that are contralateral, other striatal circuits may have pre-existing DA sensory responses to other modalities that could, in turn, predispose learning in favor of those associations. For example, DA terminals in the tail of the striatum, an area critical for auditory-motor learning,<sup>48–51</sup> have been shown to have

pre-existing auditory responses.<sup>52</sup> Under our proposed framework, these pre-existing auditory DA responses may contribute to auditory-motor learning, much like the visual responses we report here appear to contribute to visual-motor learning. However, this possibility remains to be directly tested.

Indeed, such hypothetical specialization simplifies the curse of dimensionality insofar as any given subcircuit would be initially biased to quickly discovering simple associations, appropriate for tasks in which a small set of stimuli are associated with reward. Of course, there is no free lunch. Such a bias for sparse solutions would not help and might indeed hinder detecting contingencies that depend on more complex (e.g., multimodal) combinations of features. This reasoning also leads to testable predictions about which types of contingencies are most easily learnable: those that well match the feature selectivity of a projection-defined DA population. This may explain why, in the present data, animals often learn more quickly to respond to stimuli on one side or the other, independently from those on the other side, rather than the fuller response rule combining both sides (Figure 1).

This framework is consistent with a classic idea<sup>38</sup> about the role of salience signals in DA neurons<sup>18,24–35</sup> in providing an optimistic bonus to support learning. However, it adds to this idea by suggesting that there are a number of separate and specialized circuits rather than a single, global prediction error. Specifically, the finding that a visual contralateral-side-specific signal in the DMS supports contralateral-side-specific learning is what is most novel about the current work.

### Initial conditions of the DMS DA system can explain individual differences in task acquisition

Our data also help to explain why different individuals learn the same task much more quickly than others (Figure 1). The variation across individuals in the pre-existing visual response in DMS DA signals before task training and the presentation of any reward predicts individual differences in acquisition of the visual-motor task on the contralateral side (Figure 3). This supports the idea that the pre-existing DMS DA visual responses facilitate reward learning and highlights the importance of the initial conditions of the DA system in understanding the emergence of individual differences in behavior.

### Relationship to prior experimental work on the DMS DA system

Our results complement prior work characterizing the DMS DA system, demonstrating contralateral response preferences,<sup>22,23,53</sup> visual responses,<sup>22,52</sup> and stimulus-value-related responses.<sup>54</sup> Our work adds to the literature primarily by revealing a pre-existing sensory-feature-specific and projection-specific DA signal that explains later learning (Figures 3 and 4).

On the other hand, the relationship between our work and recent work implicating DMS dopamine in individual differences in the development of a habitual<sup>55</sup> or punishment-resistant<sup>56</sup> behavioral strategy is not yet clear. Although it seems possible that a pre-existing sensory-feature-specific DA signal could contribute to the emergence of those strategies, more work is required to clarify a potential connection.

Whereas here we focus on visual responses in DMS DA, previous work from our group and others has identified putative

action correlates in this population.<sup>23,53,57,58</sup> Lateralized action-related DMS DA responses may relate to head-orienting behavior, as DMS is a target of frontal-orienting fields,<sup>59,60</sup> has been implicated in tasks with orienting behavioral outputs,<sup>61–63</sup> and the DMS DA signal itself reverses with changes in orientation.<sup>53</sup> For this reason, in the present task, action responses may be at least partially obscured by head-fixation, which prevents orienting behavior. Regardless, we are confident that the contrast-dependent responses examined here are visual, as we isolated them from wheel movement with an encoding model and, furthermore, confirmed their presence during stimulus pre-exposure in the absence of task-related movements (Figures 3A and 3B).

### Relationship to the feature-specific RPE model of DA heterogeneity

The presence of a visual-feature-specific DMS DA signal is consistent with a recent model that proposes that response variation across DA neurons can be explained, at least in part, by differences in the feature representations in the inputs that are used to calculate reward prediction errors (“feature-specific reward prediction error model”<sup>47</sup>). In this framework, different dopamine neurons calculate different reward prediction errors based on different subsets of the full feature space, based on the corticostriatal inputs they preferentially receive. This model thus predicts similar feature selectivity of the DMS-projecting DA neurons relative to the DMS neurons themselves, assuming an anatomical arrangement where the DMS projects (indirectly or directly) primarily to DMS-projecting DA neurons.<sup>64</sup> Consistent with this prediction, the DMS receives direct visual cortical inputs<sup>65</sup> and has visual responses before task training.<sup>66</sup> Alternatively, the visual responses we observed in DMS DA neurons could originate from the superior colliculus rather than the basal ganglia.<sup>67–69</sup>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Ilana Witten ([iwitten@princeton.edu](mailto:iwitten@princeton.edu)).

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- All the code for the analyses and visualizations performed in this project has been made publicly available. The URLs are listed in the [key resources table](#).
- All behavioral data have been made publicly available. The URLs are listed in the [key resources table](#).
- All photometric data have been made publicly available. The URLs are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

### ACKNOWLEDGMENTS

We thank Oliver Huang, Angela Chan, and the PNI Viral Core Facility for AAV production; Adrian Sirko and the Princeton Laboratory Animal Resources staff for help with animal husbandry; Zoe Ashwood, Sam Zorowitz, and Harrison Ritz for helpful discussion about behavioral modeling; and V. Roser, M. Siniscalchi, R. Fetcho, and other members of the Witten lab for feedback on this work. This work was supported by NIH 1DP1MH136573 (I.B.W.), the Simons

Collaboration on the Global Brain (I.B.W. and J.W.P.), the Princeton Neuroscience Institute Innovation Fund (I.B.W. and N.D.D.), and Wellcome Trust – Sir Henry Wellcome Fellowship 221657/Z/20/Z (A.P.-V.) and 1U19NS123716-01 (I.B.W. and J.W.P.).

### AUTHOR CONTRIBUTIONS

Conceptualization, A.P.-V., Y.S.A., J.W.P., N.D.D., and I.B.W.; investigation, A.P.-V., B.M., M.L., L.H., and A.B.; methodology, A.P.-V., Y.S.A., J.W.P., N.D.D., and I.B.W.; data curation, A.P.-V. and M.F.; formal analysis, Y.S.A.; software, Y.S.A.; visualization, A.P.-V. and Y.S.A.; writing, A.P.-V., Y.S.A., J.W.P., N.D.D., and I.B.W.; supervision, A.P.-V., N.D.D., J.W.P., and I.B.W.; funding acquisition, A.P.-V., J.W.P., N.D.D., and I.B.W.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Mice
- METHOD DETAILS
  - Surgery
  - Behavioral task
  - Fiber photometry
  - Histology
  - Optogenetic stimulation
  - Behavioral model
  - Neural model
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Linear Mixed Models variable coding

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2024.09.045>.

Received: March 5, 2024

Revised: June 11, 2024

Accepted: September 17, 2024

Published: October 15, 2024

### REFERENCES

1. Hu, F., Kamigaki, T., Zhang, Z., Zhang, S., Dan, U., and Dan, Y. (2019). Prefrontal corticotectal neurons enhance visual processing through the superior colliculus and pulvinar thalamus. *Neuron* **104**, 1141–1152.e4.
2. Siniscalchi, M.J., Phoumthippavong, V., Ali, F., Lozano, M., and Kwan, A.C. (2016). Fast and slow transitions in frontal ensemble activity during flexible sensorimotor behavior. *Nat. Neurosci.* **19**, 1234–1242.
3. Pinto, L., and Dan, Y. (2015). Cell-type-specific activity in prefrontal cortex during goal-directed behavior. *Neuron* **87**, 437–450.
4. Mah, A., Schierbeck, S.S., Bossio, V., and Constantinople, C.M. (2023). Distinct value computations support rapid sequential decisions. Preprint at bioRxiv. <https://doi.org/10.1101/2023.03.14.532617>.
5. Platt, M.L., and Glimcher, P.W. (1999). Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238.
6. Britten, K.H., Shadlen, M.N., Newsome, W.T., and Movshon, J.A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* **12**, 4745–4765.
7. International Brain Laboratory, Aguillon-Rodriguez, V., Angelaki, D., Bayer, H., Bonacchi, N., Carandini, M., Cazes, F., Chapuis, G., Churchland, A.K., Dan, Y., et al. (2021). Standardized and reproducible measurement of decision-making in mice. *eLife* **10**, e63711. <https://doi.org/10.7554/eLife.63711>.
8. Bernklau, T.W., Righetti, B., Mehrke, L.S., and Jacob, S.N. (2024). Striatal dopamine signals reflect perceived cue-action-outcome associations in mice. *Nat. Neurosci.* **27**, 747–757. <https://doi.org/10.1038/s41593-023-01567-2>.
9. Garcia, S.L., Laffere, A., Toschi, C., Schilling, L., Podlaski, J., Fritsche, M., Zatzka-Haas, P., Li, Y., Bogacz, R., Saxe, A., and Lak, A. (2023). Striatal dopamine reflects individual long-term learning trajectories. Preprint at bioRxiv. <https://doi.org/10.1101/2023.12.14.571653>.
10. Moore, S., and Kuchibhotla, K.V. (2022). Slow or sudden: Re-interpreting the learning curve for modern systems neuroscience. *IBRO Neurosci. Rep.* **13**, 9–14.
11. Gallistel, C.R., Fairhurst, S., and Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proc. Natl. Acad. Sci. USA* **101**, 13124–13131.
12. Papachristos, E.B., and Gallistel, C.R. (2006). Autoshaped head poking in the mouse: a quantitative analysis of the learning curve. *J. Exp. Anal. Behav.* **85**, 293–308.
13. Burgess, C.P., Lak, A., Steinmetz, N.A., Zatzka-Haas, P., Bai Reddy, C., Jacobs, E.A.K., Linden, J.F., Paton, J.J., Ranson, A., Schröder, S., et al. (2017). High-yield methods for accurate two-alternative visual psychophysics in head-fixed mice. *Cell Rep.* **20**, 2513–2524.
14. International Brain Laboratory, Benson, B., Benson, J., Birman, D., Bonacchi, N., Carandini, M., Catarino, J.A., Chapuis, G.A., Churchland, A.K., Dan, Y., et al. (2023). A brain-wide map of neural activity during complex behaviour. Preprint at bioRxiv. <https://doi.org/10.1101/2023.07.04.547681>.
15. International Brain Laboratory, Banga, K., Benson, J., Bonacchi, N., Buijins, S.A., Campbell, R., Chapuis, G.A., Churchland, A.K., Felicia Davatolhagh, M., Lee, H.D., et al. (2022). Reproducibility of in-vivo electrophysiological measurements in mice. Preprint at bioRxiv. <https://doi.org/10.1101/2022.05.09.491042>.
16. Pinto, L., Koay, S.A., Engelhard, B., Yoon, A.M., Deverett, B., Thiberge, S.Y., Witten, I.B., Tank, D.W., and Brody, C.D. (2018). An accumulation-of-evidence task using visual pulses for mice navigating in virtual reality. *Front. Behav. Neurosci.* **12**, 36.
17. Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. *Science* **275**, 1593–1599.
18. Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* **13**, 900–913.
19. Hollerman, J.R., and Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* **1**, 304–309.
20. Pan, W.-X., Schmidt, R., Wickens, J.R., and Hyland, B.I. (2005). Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J. Neurosci.* **25**, 6235–6242.
21. Coddington, L.T., and Dudman, J.T. (2018). The timing of action determines reward prediction signals in identified midbrain dopamine neurons. *Nat. Neurosci.* **21**, 1563–1573.
22. Moss, M.M., Zatzka-Haas, P., Harris, K.D., Carandini, M., and Lak, A. (2021). Dopamine axons in dorsal striatum encode contralateral visual stimuli and choices. *J. Neurosci.* **41**, 7197–7205.
23. Parker, N.F., Cameron, C.M., Taliaferro, J.P., Lee, J., Choi, J.Y., Davidson, T.J., Daw, N.D., and Witten, I.B. (2016). Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* **19**, 845–854.
24. Menegas, W., Babayan, B.M., Uchida, N., and Watabe-Uchida, M. (2017). Opposite initialization to novel cues in dopamine signaling in ventral and

- posterior striatum in mice. *eLife* 6, e21886. <https://doi.org/10.7554/eLife.21886>.
25. Lutas, A., Kucukdereli, H., Alturkistani, O., Carty, C., Sugden, A.U., Fernando, K., Diaz, V., Flores-Maldonado, V., and Andermann, M.L. (2019). State-specific gating of salient cues by midbrain dopaminergic input to basal amygdala. *Nat. Neurosci.* 22, 1820–1833.
  26. de Jong, J.W., Afjei, S.A., Pollak Dorocic, I., Peck, J.R., Liu, C., Kim, C.K., Tian, L., Deisseroth, K., and Lammel, S. (2019). A neural circuit mechanism for encoding aversive stimuli in the mesolimbic dopamine system. *Neuron* 101, 133–151.e7.
  27. Brischoux, F., Chakraborty, S., Brierley, D.I., and Ungless, M.A. (2009). Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proc. Natl. Acad. Sci. USA* 106, 4894–4899.
  28. Lerner, T.N., Shilyansky, C., Davidson, T.J., Evans, K.E., Beier, K.T., Zalocusky, K.A., Crow, A.K., Malenka, R.C., Luo, L., Tomer, R., and Deisseroth, K. (2015). Intact-brain analyses reveal distinct information carried by SNc dopamine subcircuits. *Cell* 162, 635–647.
  29. Wang, D.V., and Tsien, J.Z. (2011). Convergent processing of both positive and negative motivational signals by the VTA dopamine neuronal populations. *PLoS One* 6, e17047.
  30. Gore, B.B., and Zweifel, L.S. (2013). Genetic reconstruction of dopamine D1 receptor signaling in the nucleus accumbens facilitates natural and drug reward responses. *J. Neurosci.* 33, 8640–8649.
  31. Yuan, L., Dou, Y.-N., and Sun, Y.-G. (2019). Topography of reward and aversion encoding in the mesolimbic dopaminergic system. *J. Neurosci.* 39, 6472–6481.
  32. Horvitz, J.C., Stewart, T., and Jacobs, B.L. (1997). Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. *Brain Res.* 759, 251–258.
  33. Cai, L.X., Pizano, K., Gundersen, G.W., Hayes, C.L., Fleming, W.T., Holt, S., Cox, J.M., and Witten, I.B. (2020). Distinct signals in medial and lateral VTA dopamine neurons modulate fear extinction at different times. *eLife* 9, e54936. <https://doi.org/10.7554/eLife.54936>.
  34. Kutlu, M.G., Zachry, J.E., Melugin, P.R., Cajigas, S.A., Chevee, M.F., Kelly, S.J., Kutlu, B., Tian, L., Siciliano, C.A., and Calipari, E.S. (2021). Dopamine release in the nucleus accumbens core signals perceived saliency. *Curr. Biol.* 31, 4748–4761.e8.
  35. Kutlu, M.G., Zachry, J.E., Melugin, P.R., Tat, J., Cajigas, S., Isiktas, A.U., Patel, D.D., Siciliano, C.A., Schoenbaum, G., Sharpe, M.J., and Calipari, E.S. (2022). Dopamine signaling in the nucleus accumbens core mediates latent inhibition. *Nat. Neurosci.* 25, 1071–1081.
  36. Lak, A., Stauffer, W.R., and Schultz, W. (2016). Dopamine neurons learn relative chosen value from probabilistic rewards. *eLife* 5, e18044. <https://doi.org/10.7554/eLife.18044>.
  37. Willmore, L., Minerva, A.R., Engelhard, B., Murugan, M., McMannon, B., Oak, N., Thiberge, S.Y., Peña, C.J., and Witten, I.B. (2023). Overlapping representations of food and social stimuli in mouse VTA dopamine neurons. *Neuron* 111, 3541–3553.e8. <https://doi.org/10.1016/j.neuron.2023.08.003>.
  38. Kakade, S., and Dayan, P. (2002). Dopamine: generalization and bonuses. *Neural Netw.* 15, 549–559.
  39. Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Vander Weele, C.M., Kennedy, R.T., Aragona, B.J., and Berke, J.D. (2016). Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* 19, 117–126.
  40. Lak, A., Okun, M., Moss, M.M., Gurnani, H., Farrell, K., Wells, M.J., Reddy, C.B., Kepecs, A., Harris, K.D., and Carandini, M. (2020). Dopaminergic and prefrontal basis of learning from sensory confidence and reward value. *Neuron* 105, 700–711.e6.
  41. Witten, I.B., Steinberg, E.E., Lee, S.Y., Davidson, T.J., Zalocusky, K.A., Brodsky, M., Yizhar, O., Cho, S.L., Gong, S., Ramakrishnan, C., et al. (2011). Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* 72, 721–733.
  42. Steinberg, E.E., Keiflin, R., Boivin, J.R., Witten, I.B., Deisseroth, K., and Janak, P.H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* 16, 966–973.
  43. Kim, K.M., Baratta, M.V., Yang, A., Lee, D., Boyden, E.S., and Fiorillo, C.D. (2012). Optogenetic mimicry of the transient activation of dopamine neurons by natural reward is sufficient for operant reinforcement. *PLoS One* 7, e33612.
  44. Morrens, J., Aydin, Ç., Janse van Rensburg, A., Esquivelzeta Rabell, J., and Haesler, S. (2020). Cue-evoked dopamine promotes conditioned responding during learning. *Neuron* 106, 142–153.e7.
  45. Lee, K., Claar, L.D., Hachisuka, A., Bakhurin, K.I., Nguyen, J., Trott, J.M., Gill, J.L., and Masmanidis, S.C. (2020). Temporally restricted dopaminergic control of reward-conditioned movements. *Nat. Neurosci.* 23, 209–216.
  46. Iglesias, A.G., Chiu, A.S., Wong, J., Campus, P., Li, F., Bhatti, J.K., Patel, S.A., Deisseroth, K., Akil, H., Burgess, C.R., Flagel, S.B., et al. (2023). Inhibition of dopamine neurons prevents incentive value encoding of a reward cue: With revelations from deep phenotyping. *J. Neurosci.* 43, 7376–7392. <https://doi.org/10.1101/2023.05.03.539324>.
  47. Lee, R.S., Sagiv, Y., Engelhard, B., Witten, I.B., and Daw, N.D. (2024). A feature-specific prediction error model explains dopaminergic heterogeneity. *Nat. Neurosci.* 27, 1574–1586. <https://doi.org/10.1038/s41593-024-01689-1>.
  48. Greenstreet, F., Vergara, H.M., Pati, S., Schwarz, L., Wisdom, M., Marbach, F., Johansson, Y., Rollik, L., Moskovitz, T., Clopath, C., et al. (2022). Action prediction error: a value-free dopaminergic teaching signal that drives stable learning. Preprint at bioRxiv. <https://doi.org/10.1101/2022.09.12.507572>.
  49. Xiong, Q., Znamenskiy, P., and Zador, A.M. (2015). Selective corticostriatal plasticity during acquisition of an auditory discrimination task. *Nature* 521, 348–351.
  50. Znamenskiy, P., and Zador, A.M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature* 497, 482–485.
  51. Ghosh, S., and Zador, A.M. (2021). Corticostriatal plasticity established by initial learning persists after behavioral reversal. *eNeuro* 8, <https://doi.org/10.1523/ENEURO.0209-20.2021>.
  52. Vu, M.T., Brown, E.H., Wen, M.J., Noggle, C.A., Zhang, Z., Monk, K.J., Bouabid, S., Mroz, L., Graham, B.M., Zhuo, Y., et al. (2024). Targeted micro-fiber arrays for measuring and manipulating localized multi-scale neural dynamics over large, deep brain volumes during behavior. *Neuron* 112, 909–923.e9. <https://doi.org/10.1016/j.neuron.2023.12.011>.
  53. Lee, R.S., Mattar, M.G., Parker, N.F., Witten, I.B., and Daw, N.D. (2019). Reward prediction error does not explain movement selectivity in DMS-projecting dopamine neurons. *eLife* 8, <https://doi.org/10.7554/eLife.42992>.
  54. Tsutsui-Kimura, I., Matsumoto, H., Akiti, K., Yamada, M.M., Uchida, N., and Watabe-Uchida, M. (2020). Distinct temporal difference error signals in dopamine axons in three regions of the striatum in a decision-making task. *eLife* 9, e42992. <https://doi.org/10.7554/eLife.62390>.
  55. van Elzelingen, W., Warnaar, P., Matos, J., Bastet, W., Jonkman, R., Smulders, D., Goedhoop, J., Denys, D., Arbab, T., and Willuhn, I. (2022). Striatal dopamine signals are region specific and temporally stable across action-sequence habit formation. *Curr. Biol.* 32, 1163–1174.e6.
  56. Seiler, J.L., Cosme, C.V., Sherathiya, V.N., Schaid, M.D., Bianco, J.M., Bridgemohan, A.S., and Lerner, T.N. (2022). Dopamine signaling in the dorsomedial striatum promotes compulsive behavior. *Curr. Biol.* 32, 1175–1188.e5.
  57. Howe, M.W., and Dombeck, D.A. (2016). Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* 535, 505–510.
  58. Jin, X., and Costa, R.M. (2010). Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* 466, 457–462.
  59. Erlich, J.C., Brunton, B.W., Duan, C.A., Hanks, T.D., and Brody, C.D. (2015). Distinct effects of prefrontal and parietal cortex inactivations on

- an accumulation of evidence task in the rat. *eLife* 4, e05457. <https://doi.org/10.7554/eLife.05457>.
60. Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., and Brody, C.D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* 520, 220–223.
  61. Tai, L.-H., Lee, A.M., Benavidez, N., Bonci, A., and Wilbrecht, L. (2012). Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat. Neurosci.* 15, 1281–1289.
  62. Yartsev, M.M., Hanks, T.D., Yoon, A.M., and Brody, C.D. (2018). Causal contribution and dynamical encoding in the striatum during evidence accumulation. *eLife* 7, e34929. <https://doi.org/10.7554/eLife.34929>.
  63. Bolkan, S.S., Stone, I.R., Pinto, L., Ashwood, Z.C., Iravedra Garcia, J.M., Herman, A.L., Singh, P., Bandi, A., Cox, J., Zimmerman, C.A., et al. (2022). Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. *Nat. Neurosci.* 25, 345–357.
  64. Haber, S.N. (2003). The primate basal ganglia: parallel and integrative networks. *J. Chem. Neuroanat.* 26, 317–330.
  65. Khibnik, L.A., Tritsch, N.X., and Sabatini, B.L. (2014). A direct projection from mouse primary visual cortex to dorsomedial striatum. *PLoS One* 9, e104501.
  66. Peters, A.J., Fabre, J.M.J., Steinmetz, N.A., Harris, K.D., and Carandini, M. (2021). Striatal activity topographically reflects cortical activity. *Nature* 591, 420–425.
  67. Dommett, E., Coizet, V., Blaha, C.D., Martindale, J., Lefebvre, V., Walton, N., Mayhew, J.E.W., Overton, P.G., and Redgrave, P. (2005). How visual stimuli activate dopaminergic neurons at short latency. *Science* 307, 1476–1479.
  68. Watabe-Uchida, M., Zhu, L., Ogawa, S.K., Vamanrao, A., and Uchida, N. (2012). Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74, 858–873.
  69. Solié, C., Contestabile, A., Espinosa, P., Musardo, S., Bariselli, S., Huber, C., Carleton, A., and Bellone, C. (2022). Superior colliculus to VTA pathway controls orienting response and influences social interaction in mice. *Nat. Commun.* 13, 817.
  70. Paxinos, G., and Franklin, K. (2019). *The Mouse Brain in Stereotaxic Coordinates* (Elsevier).
  71. International Brain Laboratory (2020). A standardized and reproducible method to measure decision-making in mice. Preprint at bioRxiv. <https://doi.org/10.6084/m9.figshare.11634732.v1>.
  72. Lopes, G., Bonacchi, N., Frazão, J., Neto, J.P., Atallah, B.V., Soares, S., Moreira, L., Matias, S., Itskov, P.M., Correia, P.A., et al. (2015). Bonsai: an event-based framework for processing and controlling data streams. *Front. Neuroinform.* 9, 7.
  73. Lopes, G., Farrell, K., Horrocks, E.A., Lee, C.-Y., Morimoto, M.M., Muzzu, T., Papanikolaou, A., Rodrigues, F.R., Wheatcroft, T., Zucca, S., et al. (2021). Creating and controlling visual environments using BonVision. *eLife* 10, e65541. <https://doi.org/10.7554/eLife.65541>.
  74. Martianova, E., Aronson, S., and Proulx, C.D. (2019). Multi-fiber photometry to record neural activity in freely-moving animals. *J. Vis. Exp.* 152, e60278. <https://doi.org/10.3791/60278>.
  75. Marshel, J.H., Kim, Y.S., Machado, T.A., Quirin, S., Benson, B., Kadmon, J., Raja, C., Chibukhchyan, A., Ramakrishnan, C., Inoue, M., et al. (2019). Cortical layer-specific critical dynamics triggering perception. *Science* 365, eaaw5202. <https://doi.org/10.1126/science.aaw5202>.
  76. Deisseroth, K. (2012). Predicted irradiance values: model based on direct measurements in mammalian brain tissue. <https://web.stanford.edu/group/dlab/cgi-bin/graph/chart.php>.
  77. Roy, N.A., Bak, J.H., International; Brain Laboratory, Akrami, A., Brody, C.D., and Pillow, J.W. (2021). Extracting the dynamics of behavior in sensory decision-making experiments. *Neuron* 109, 597–610.e6.
  78. Bruijns, S.A., International Brain Laboratory, Bougrova, K., Laranjeira, I.C., Lau, P.Y.P., Meijer, G.T., Miska, N.J., Noel, J.-P., Pan-Vazquez, A., Roth, N., et al. (2023). Dissecting the complexities of learning with infinite hidden Markov models. Preprint at bioRxiv. <https://doi.org/10.1101/2023.12.22.573001>.
  79. Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1701.02434>.
  80. Hoffman, M.D., and Gelman, A. (2011). The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1111.4246>.
  81. Betancourt, M. (2016). Identifying the optimal integration time in Hamiltonian Monte Carlo. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1601.00225>.
  82. Krueger, L.E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behav. Brain Sci.* 12, 251–267.
  83. Nieder, A., and Miller, E.K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 37, 149–157.
  84. Ritz, H., and Shenhav, A. (2024). Humans reconfigure target and distractor processing to address distinct task demands. *Psychol. Rev.* 131, 349–372. <https://doi.org/10.1037/rev0000442>.
  85. Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., and Riddell, A. (2017). Stan: a probabilistic programming language. *J. Stat. Softw.* 76, <https://doi.org/10.18637/jss.v076.i01>.
  86. Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., and Simoncelli, E.P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454, 995–999.
  87. MacKay, D.J.C. (1992). Bayesian interpolation. *Neural Comput.* 4, 415–447.
  88. Maronna, R.A., Douglas Martin, R., Yohai, V.J., and Salibián-Barrera, M. (2019). *Robust Statistics: Theory and Methods (with R)* (John Wiley & Sons).

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Bacterial and virus strains</b>		
AAV2/5-EF1a-DIO-ChRmine-mScarlet-WPRE-hGHpA	Princeton Neuroscience Institute Viral Core	N/A
<b>Deposited data</b>		
Photometry and behavioral data	Generated by this study	<a href="https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/blob/main/alex_figure_1_4_supp_1/download_data.py">https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/blob/main/alex_figure_1_4_supp_1/download_data.py</a> . (This script downloads all raw and processed data from International Brain Lab servers) Documentation for loading the dataset: <a href="https://int-brain-lab.github.io/iblenv/notebooks_external/loading_photometry_data.html">https://int-brain-lab.github.io/iblenv/notebooks_external/loading_photometry_data.html</a>
<b>Experimental models: Organisms/strains</b>		
B6.SJL-Slc6a3 <sup>tm1.1(cre)Bkmm</sup> /J (DAT::IRES-Cre)	The Jackson Laboratory	Stock # 006660; RRID: IMSR_JAX:006660
B6.Cg-Igs7 <sup>tm148.1(tetO-GCaMP6f,CAG-tTA2)Hze</sup> /J (Ai148)	The Jackson Laboratory	Stock # 030328; RRID: IMSR_JAX:030328
<b>Software and algorithms</b>		
Neural Encoding model	Generated by this study	<a href="https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/blob/main/pipeline">https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/blob/main/pipeline</a>
Behavioral model	Generated by this study	<a href="https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/blob/main/pipeline">https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/blob/main/pipeline</a>
Figure visualization	Generated by this study	<a href="https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/tree/main/pipeline/figures">https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/tree/main/pipeline/figures</a>
Statistical analyses	Generated by this study	<a href="https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/blob/main/pipeline">https://github.com/YSanchezAraujo/PRE_VIS_DA_CURR_BIO/blob/main/pipeline</a>
<b>Other</b>		
Fibers for optogenetics	Thor Labs	BFL37-300
Ferrules for optogenetics	Precision Fiber Products	MM-FER-2006SS-330
Fibers for fiber photometry	Neurophotometrics	FOC_BF_200um/1.25mm
IBL Behavior Rig	International Brain Laboratory	<a href="https://figshare.com/articles/preprint/A_standardized_and_reproducible_method_to_measure_decision-making_in_mice_Appendix_3_IBL_protocol_for_setting_up_the_behavioral_training_rig/11634732">https://figshare.com/articles/preprint/A_standardized_and_reproducible_method_to_measure_decision-making_in_mice_Appendix_3_IBL_protocol_for_setting_up_the_behavioral_training_rig/11634732</a>
Multi-fiber photometry rig	Neurophotometrics	FP3002
Branching Bundle Patchcord - Low Autofluorescence	Doric Lenses Inc.	BBP(3)_200/220/900-0.37_2m_SMA-3xMF1.25_LAF

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

#### Mice

For the fiber photometry experiments (Figures 1, 2, and 3), a total of 22 mice ( $n=14$  male and  $n=8$  female) were used from a cross of DAT::IRES-Cre mice (JAX 006660) and the GCaMP6f reporter line Ai148 (JAX 030328). For the optogenetic experiments (Figure 4), we used a total of 13 DAT::IRES-Cre mice ( $n=4$  male and  $n=9$  female). Mice were maintained on a reversed 12 h light cycle and experiments were formed on their dark cycle. All mice used were 3-4 months old at the start of training. All experimental procedures were conducted in accordance with guidelines from the National Institutes of Health and were reviewed by the Institutional and Animal Care Use Committee at Princeton University.

## METHOD DETAILS

### Surgery

Prior to the start of the surgery, mice received a preoperative antibiotic (5 mg/kg Baytril) and analgesic (10 mg/kg Ketofen). Postoperative analgesic (10 mg/kg Ketofen) was administered daily for 3 days from the day of the surgery.

### Headbar implantation

For all stereotaxic surgeries, mice underwent sterile stereotaxic surgery under anesthesia (5% isoflurane for induction, 1.5–2% for maintenance). Briefly, the scalp and underlying periosteum was removed. Bregma and lambda were leveled, and a small steel headbar was centered at -6.9 mm Anterior-Posterior relative to bregma<sup>70</sup> and cemented to the skull with Metabond (Parkell). Headbar implantation was followed by virus infusion and/or optical fiber implantation (see sections below).

### Optical fiber implantation

For fiber photometry experiments (data shown in Figures 1, 2, and 3), low-autofluorescence optical fibers encased in a ferrule (0.37 NA,  $\varnothing$ 200  $\mu$ m core, 1.25mm ferrule, Neurophotometrics) were implanted at each of the following locations (fiber tip location relative to bregma):

- DLS: 2.6 mm (Medio-lateral, M-L), 0 mm (Anterior-Posterior, A-P), -2.8 mm (Dorso-ventral, D-V).
- DMS: 1.5 mm M-L, 0.74 mm A-P, -2.4 mm D-V.
- NAc: 1 mm M-L, 1.45 mm A-P, -4.5 mm D-V.

In each mouse, fibers targeting DLS and NAc were always inserted in the same hemisphere, the DMS fiber was positioned in the opposite hemisphere. The hemisphere allocation was counterbalanced across mice. For DLS, the location above was targeted with a fiber rotated at 10° in the M-L/D-V plane. For the optogenetics experiments (Figure 4), fiber optic fibers ( $\varnothing$ 300  $\mu$ m core/0.39 NA, 2.5 mm ferrule, ThorLabs) were implanted bilaterally to target the DMS at the following coordinates: +/- 1.5 mm M-L, 0.74 mm A-P, -2.4 mm D-V. These locations were reached with a 10° M-L/D-V rotation.

### Virus injections

For the optogenetics experiments (Figure 4), AAV2/5-EF1a-DIO-ChRmine-mScarlet-WPRE-hGHpA (opsin virus, titer:  $9e^{12}$  genome copies/ml, Princeton Neuroscience Institute viral core) or AAV2/5-EF1a-DIO-EYFP-WPRE-hGHpA (control virus, titer:  $1.5e^{14}$  genome copies/ml, Princeton Neuroscience Institute viral core) was infused bilaterally in the VTA-SNc (+/- 1 mm M-L, -3.1 A-P, -4.66 D-V) of ~4–6 weeks old mice. 500 nl were infused in each hemisphere at a speed of 75 nl/min. In order to achieve sufficient terminal expression by the start of training (3/4 months), all viral injections were performed a minimum of 8 weeks in advance of training and prior to the headplate implantation surgery.

### Behavioral task

#### Behavioral apparatus

We used the standardized behavioral apparatus from the International Brain Laboratory. For detailed instructions on the components and operations of behavioral apparatus used please see International Brain Laboratory.<sup>71</sup> Briefly, the rig consisted of an LCD screen (LP097Q  $\times$  1, LG) and a custom 3D-printed mouse holder and head fixation system that held the mouse in front of the screen such that its forepaws rested on a rubber steering wheel (86652 and 32019, LEGO). A spout was positioned in front of the holder, which the mouse could reach it with its tongue but it did not occlude the field of vision. The spout was connected to a water reservoir and water flow was controlled with a solenoid pinch valve (225P011-21, NResearch). The rig was constructed with Thorlabs parts inside a small soundproof cabinet (9U acoustic wall cabinet 600  $\times$  600, Orion). A speaker (HPD-40N16PET00-32, Peerless by Tympany) positioned on top of the screen was used to play task-related sounds, and an ultrasonic microphone (Ultramic UM200K, Dodotronic) was used to record ambient noise from the rig. Wheel position was recorded with a rotary encoder (05.2400.1122.1024, Kubler) controlled by the Bpod Rotary Encoder Module (Sanworks). Video of the mouse was recorded with a USB camera (CM3-U3-13Y3M-CS, Point Grey). All task-related devices were controlled by a Bpod State Machine (Sanworks) and synched with a data acquisition board (USB201, Measurement Computing). The task logic was programmed in Python and the visual stimulus presentation and video capture was handled by Bonsai<sup>72</sup> and the Bonsai package BonVision.<sup>73</sup>

#### Behavioral task and training

We used a standardized visual decision-making task.<sup>7,13</sup> In this task, mice are head fixed in front of a LCD screen. A visual grating (Gabor patches, 0.1 spatial frequency) whose contrast varied across trials (100%, 50%, 25%, 12.5%, 6.25%) appeared on either the right or left side of the screen (+/- 35° azimuth), accompanied by a 0.1 s tone (5 kHz sine wave, 10ms ramp). A steering wheel that could be used to move the visual grating along the horizontal axis was placed under the mouse's paws (4° of visual grating movement/mm of wheel movement). The mouse could obtain a small reward of 10% sucrose water (3  $\mu$ l) by moving the visual grating to the center of the screen. On the contrary, if the mouse steered the grating away from the center (35° from initial position) or failed to center the grating in 60 s, the trial was considered an error. Errors were signaled by the lack of reward delivery and a brief noise (0.5 s, 65 dB, white noise). After a choice was completed (correct or incorrect), wheel movements could no longer move the visual grating for 1 or 2 seconds on correct versus incorrect trials, respectively. After this timeout, all trials were followed by a 0.5 s inter trial interval where no gratings were presented. In order for a new trial to start, the steering wheel had to be still for a "quiescent period", whose duration was randomly sampled on a trial by trial basis from an exponential distribution of mean 0.55 s and truncated from 0.4 to 0.7 s.

To simplify the interpretation of neural activity correlates of behavior across learning, mice experienced the full extent of the task (no shaping or debiasing protocol) from the first session. All mice were trained for a minimum of 18 sessions, 5–7 days a week, for a 1h session each day. In order to motivate the mice to do the task, mice had restricted water access from 1 week before starting training until the end of the experiment. We monitored that their weight never dropped more than 20% from their pre-water restriction weight, and ensured that they consumed a daily minimum of 1 ml of water per 25 g of weight. Most mice were able to obtain their daily allocation of water through the task alone after a few sessions. When this minimum was not achieved, mice were supplemented at the end of the day. Mice did not have a limit on how much water they could obtain in the task (See [Figure S1A](#) for average trials completed across training). Mice were video recorded every session.

### Stimulus pre-exposure

For the experiment in [Figure 3](#), before training on the task (1–10 days prior to the start of training), 18 mice underwent two 1h pre-exposure sessions where we measured neural responses to task features in the absence of reward. In the first pre-exposure session, mice were presented with the same visual gratings used in the task (with the same range of contrasts) on either side of the screen for 250–272 trials with a 10 seconds inter-trial interval. The stimulus contrast and side in each trial was randomly sampled between the 8 possible combinations. As in the standard task, presentation of the visual gratings was accompanied by a brief 0.1 s tone. However, during this pre-training session the wheel was locked and the visual gratings remained static on either side of the screen. In the second session (data not shown), mice were allowed to move the wheel and move the visual grating but no rewards were given.

### Fiber photometry

#### Data acquisition

We simultaneously recorded GCaMP6f signals from DA terminals in DMS, DLS and NAc with a multi-fiber photometry system (FP3002, Neurophotometrics) controlled with the Bonsai Neurophotometrics module.<sup>74</sup> Briefly, the system consists of a CMOS camera acquiring fluorescence emissions and an LED exciting with 470 nm light of 10 ms width pulses at 50Hz (464/476 sessions) or 20 Hz (12/476 sessions). At the tip of the patch cable, the excitation light was ~0.04 mW. The camera acquisition epochs were timed with the emission lights. We used a Low Autofluorescence Patch Cord with 3 branches (BBP(3)\_200/220/900-0.37\_2m\_SMA-3xMF1.25\_LAF, Doric) to be able to image DMS, DLS and NAc simultaneously. Prior to each recording day, we passed 0.5 mW 470 nm light through the patch cord for 1 hour in order to photobleach autofluorescence within the patch cord, and improve recording quality.

#### Signal processing

Fluorescence signals recorded during each session from each location were transformed to dF/F using the following formula:

$$\frac{dF}{F} = \frac{F - F_0}{F_0}$$

$F_0$  was the +/- 30 s rolling average of the raw fluorescence signal. Finally, dF/F signals were z-scored per-session, using a mean and standard deviation calculated based on all the data from each session. To be included for analysis, every recording (i.e. one session from one fiber location) had to have at least one  $\geq 1\%$  dF/F &  $> 3$  standard deviation transient for every 10 min of recording (55/1440 recordings excluded). All data were sampled or resampled at 50 Hz for analysis. Example recording traces can be found in [Figure S3](#).

### Histology

To confirm the locations of the optical fibers and viral expression ([Figure S2](#)), mice were anesthetized with pentobarbital sodium (2 mg/kg, Euthasol) and transcardially perfused first with 10 ml of ice-cold phosphate buffered saline (PBS) followed by 25 ml of 4% paraformaldehyde (PFA) in PBS. Brains were then dissected and post-fixed in 4% PFA overnight at 4°C. After fixation, brains were sliced with a vibrating blade microtome (Vibrotome VT100S, Leica) and mounted with DAPI Fluoromount-G (Southern Biotech). All slices were imaged with an automated slide scanner (NanoZoomer S60, Hamamatsu).

### Optogenetic stimulation

For the optogenetic experiment in [Figure 4](#), fibers were implanted bilaterally in the DMS to avoid potential behavioral biases related to an asymmetrical surgery. Stimulation was delivered unilaterally to DMS terminals expressing the red-shifted opsin ChRmine. The stimulated hemisphere was chosen randomly and kept constant throughout training for any given animal. The group identity of the mice (opsin vs. control) were blinded to the experimenter throughout the duration of training. The stimulation procedure consisted of a 200 ms laser train timed to the onset of any visual stimulus presentation contralateral to the stimulated hemisphere. Each 200 ms train of stimulation consisted of 20 Hz and 5 ms width light pulses at a wavelength of 532 nm (Shanghai Laser and Optics & Co). The light power was adjusted daily to 0.25 mW at the fiber tip (in the brain). Light power was chosen to ensure activation of the terminals immediately below the fiber tip but minimize off-target activation outside DMS. ChRmine can reliably be activated with an irradiance of  $\geq 0.1$  mW/mm<sup>2</sup>.<sup>75</sup> Therefore, we chose a stimulation power that ensured irradiance above this threshold within but not outside of DMS. According to Deisseroth<sup>76</sup> the chosen power and fiber (~0.25 mW 532nm light through  $\varnothing 300$   $\mu$ m core/0.39 NA fiber) yields an irradiance of 0.88 mW/mm<sup>2</sup> (above threshold) just below the fiber tip (DMS) and 0.01 mW/mm<sup>2</sup> (below threshold) at the DMS/NAc border (1.7 mm ventral from the fiber tip).

### Behavioral model

Our approach in modeling behavior aims to descriptively characterize the relatively long time-scale dynamics of learning that would be required to correctly associate stimuli, actions, and outcomes, particularly in the absence of shaping, de-biasing, or other experimental protocols. This relates to previous modeling efforts of similar datasets; however, instead of focusing on trial-to-trial fluctuations in psychophysical weights<sup>77</sup> or the emergence of multi-state behavior,<sup>78</sup> we focus on session-level changes in psychophysical weights. We leveraged advances in MCMC<sup>79–81</sup> to infer a set of parameters and weights for Bernoulli generalized linear models (GLM) that were expressive enough to capture the full set of behaviors that mice in our task explored.

To model the behavioral data, we built a hierarchical Bernoulli GLM to describe the relationship between the animal's choices and a variety of task covariates. The dependent variable was per-trial choice (a Bernoulli variable). The covariates included the stimulus presented to the animal on each trial on the left or right (capturing the classic psychometric curve) together with two additional effects: the animal's exponentially filtered choice history, and a bias term. We parameterized the stimulus using two regressors,  $x_L$  and  $x_R$ , corresponding to the contrast of the left-side and right-side stimulus on each trial; because the stimulus only appeared on a single side in each trial, one of these regressors was zero on each trial. We transformed each contrast regressor using a tanh function:  $x_{side} = \frac{\tanh(\alpha S)}{\tanh(\alpha)}$ , where  $S$  is either  $x_L$  or  $x_R$  and  $\alpha$  is a positive constant governing the shape of the nonlinear transformation. Dividing by  $\tanh(\alpha)$  ensures that at 100% contrast trials  $x_{side} = 1$ . This parametrization allowed the model to saturate at contrast levels below 100%, sidestepping the need to use lapse parameters to account for the flattening of the psychometric function at high contrast levels.<sup>82–84</sup> We generated the choice history regressor  $c_t$  by exponentially filtering previous choices with time constant  $\pi$ :  $c_t = c_{t-1} + \pi (y_{t-1} - c_{t-1})$ , where choice  $y_t$  takes values of -1 and +1 for left and right choices, respectively. We inferred the time constant  $\pi$  using MCMC sampling, along with the other model parameters (see all parameters below). Across animals,  $\alpha$  was fit to 2.5+/-0.5 (mean+/-sem), and  $\pi$  to 0.52+/-0.004.

Bernoulli GLM weights varied across sessions within each mouse. All other parameters were shared across sessions, within each mouse. For each mouse we built a hierarchical model over sessions, instantiating a separate set of Bernoulli GLM weights for each session. We placed broad prior distributions on the means of the weights on the first session. For all subsequent sessions, we assigned the weights a prior centered around the previous session's inferred values.

In particular, weight priors for session  $d$  took the form of a Student's t-distribution:  $\beta_d \sim StudentT(\nu, \beta_{d-1}, \Sigma)$ , where  $\nu$  is the degrees-of-freedom parameter for the Student T prior. This allowed the model to partially pool data across sessions (smoothing the estimates), but to do so adaptively per-animal, taking on larger values of  $\nu$  for animals that steadily and slowly learned the task, and smaller values for  $\nu$  for animals that had sudden and large changes in their learning. We also inferred the covariance of the StudentT priors, allowing parameters to change across sessions in coupled fashion. Unlike the Bernoulli GLM weights, we used a single shared covariance across all sessions. This covariance was parameterized as the quadratic-form product of a diagonal matrix:  $D = \sigma \odot I$  and a correlation matrix  $\Omega$ :  $\Sigma = D\Omega D$ . The diagonal had a truncated Gaussian prior. The correlation matrix  $\Omega$  was constructed from a lower triangular matrix  $L$ , which is a cholesky factor of the correlation matrix. These factors had a prior distribution  $LKJCholesky$ . The  $LKJCholesky$  prior itself has a parameter that tunes the strength of the correlations of the cholesky factor, which we also inferred. Functions and distributions specified here were from the STAN probabilistic programming language, and all model fits were performed in STAN.<sup>85</sup>

We summarize the behavioral model below, (we note that  $x_{side}$  below refers to only the contrast regressors, while  $x$  refers to the vector of all regressors i.e., bias, left and right contrast weight, choice history):

Model Variable	Description
$\nu \sim \text{Gamma}(2, 0.2)$	Student T degrees of freedom
$\eta \sim \text{Normal}(0, 10)^+$	LKJ correlation parameter
$L \sim \text{LKJCholesky}(\eta)$	Lower triangular factor
$\sigma \sim \text{Normal}(0, 1)^+$	Diagonal scale of covariance
$\lambda \sim \text{Normal}(0, 1)$	Pre-transformed choice kernel time constant
$\pi = \text{Phi}(\lambda)$	Choice kernel time constant
$\Omega = LL'$	Correlation Matrix
$D = \sigma \odot I$	Scale Matrix
$\Sigma = D\Omega D$	Covariance Matrix
$\mu \sim \text{StudentT}(\nu, 0, 5)$	Initial prior mean
$\beta_1 \sim \text{StudentT}(\nu, \mu, \Sigma)$	First session choice weight vector
$\beta_d \sim \text{StudentT}(\nu, \beta_{d-1}, \Sigma)$	d'th session choice weight vector
$l \sim \text{Normal}(-2, 0.5)$	Pre-transformed alpha
$\alpha = \text{softplus}(l)$	Scale factor on stimulus

(Continued on next page)

Continued

Model Variable	Description
$x_{side} = \frac{\tanh(\alpha S)}{\tanh(\alpha)}$	Transformed stimulus
$p_{t,d} = \text{logistic}(x_{t,d} \cdot \beta_d)$	Probability of right choice
$y_{t,d} \sim \text{Bernoulli}(p_{t,d})$	Distribution over choices

Neural model

To model the dopaminergic signals across learning, we built a linear-Gaussian regression encoding model to describe the relationship between task events such as the visual stimuli, actions, and reward delivery with the measured dopamine (DA). Since these events can be correlated in time and their effects on dopamine are partly overlapping, estimating such an encoding model helps to tease apart their individual contributions.

The regression model was defined by a set of temporal kernels that describe the DA impulse response to different task-related events, namely “stimulus”, “action”, and “reward”. For stimulus onset events, we used contrast-specific right and left temporal kernels, giving us 4 temporal kernels per side. All kernels were strictly causal, lasting for a period of 1 second (50 Hz).

Similarly, we used contrast-specific action kernels triggered at the onset of the first significant wheel movement for left and right choices (first movement larger than 0.1 radians after the end of the quiescent period). In addition to separating these kernels by contrast and side (right / left choices) we separated them by correct and incorrect trials, resulting in 8 temporal kernels per side. Separating the action kernels in this manner provided an estimate of the DA response to the interacting effects of initial stimulus location and the movement of the stimulus towards or away the center of the screen. Finally, we defined reward kernels corresponding to the moment when the animal received a water reward or a short time-out period in the same fashion as the action kernels, giving us another 8 temporal kernels per side. Thus, in total we had 40 temporal kernels in the encoding model.

We parameterized the temporal kernels in this model using a basis of linearly scaled “raised cosine” functions spanning a 1-second window after each event.<sup>86</sup> The cosine basis significantly reduces the dimensionality of the design matrix  $X$  (compared to a full series of individual lagged event dummies). The effect on estimation of using a cosine basis is regularization. Use of a temporally smooth basis is also justified by the observation that temporally adjacent responses are strongly correlated.

We used ridge regression to estimate the model parameters, with ridge parameter  $\gamma$  and observation noise  $\sigma_f^2$  estimated via evidence optimization.<sup>87</sup> We optimized for the vector of weights  $\beta$ , and the two scalars  $\sigma_f^2$ ,  $\gamma$ , which are related to the vector of neural response  $f$  as follows:  $\Sigma = (\sigma_f^{-2} X'X + \gamma I)^{-1}$ ,  $\beta = \sigma_f^{-2} \Sigma X'f$

All weights that made up the entire set of temporal kernels were denoted by  $\beta$ , and could be indexed by their corresponding event type. For example the vector of weights  $\beta_{1:50}$  contained the weights for the temporal kernel for stimulus appearing on the right at 6.25% contrast (after one transforms them into the standard basis). The vector of weights  $\beta_{51:100}$  contained the weights for the temporal kernel for stimulus appearing on the right at 12.5% contrast, and so on for all remaining contrast levels, and event types. We further computed summary statistics of these temporal kernels, specifically the L2-norm for the stimulus responses. These summary statistics gave us a scalar measure of neural response for each training session that we then related to the estimates from the Bernoulli GLM.

The encoding model for Figure 3 was fit as described above, however we only modeled the stimulus responses, and so the full set of coefficients that made up the kernels was restricted to an intercept, the 4 temporal kernels for stimulus appearing on the right, and 4 temporal kernels for stimulus appearing on the left side of the screen.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistics reporting a correlation coefficient and a p-value on that correlation coefficient were computed using robust regression, in order to reduce the sensitivity of our statistical conclusions to outliers. Robust regression was performed using the *rlm* function from the RobustModels package in the Julia programming language. We used M-Estimators with a HuberLoss Function,<sup>88</sup> the cholesky solver for the method argument, and an initial scale value of 10. For the robust regressions, we computed correlation coefficient-like statistics analogous to Pearson’s R for classic regression. In particular, we computed a pseudo- $R^2$  statistic, and its signed square root  $r$ , using the RobustModels package *deviance* and *nulldeviance* functions:  $\text{pseudo } R^2 = 1 - \frac{\text{deviance}(\text{model})}{\text{nulldeviance}(\text{model})}$ ,  $r = \text{sign}(\beta_1) \times \sqrt{\text{pseudo } R^2}$ . Deviance is a generalization of the residual sum of squares for linear models, and null deviance is a generalization of the total sum of squares.

Statistics in Figure 2F were computed with the *OneSampleTTest* function in the HypothesisTests package from the Julia programming language. Significance was determined at  $p < 0.05$ , and all p-values reported are two-sided unless otherwise noted. See Table S1 for detailed results of these tests.

Statistical tests for group differences in behavioral trajectories in Figures 3E–3H, 4C, S4C–S4F, and S5I–S5L, were carried out with the MixedModels and AnovaMixedModels packages in the Julia programming language. Linear Mixed Models from the MixedModels package were used to test simple effects such as the relationship between session 0 DMS stimulus response on behavioral weight values within each training period (early, middle, late). We further used type-3 F-test ANOVAs from the

AnovaMixedModels package to test the overall effects in the model, such as, across training periods, is there an influence of session 0 DMS strength on behavioral weight trajectories. For all Linear-Mixed Models and ANOVAs,  $a*b*c$  expands into  $a + b + c + a*b + a*c + b*c + a*b*c$ .

### Linear Mixed Models variable coding

Across [Table S2](#); [Data S1](#), the variable “session” is a transformation of sessions 1 to 20. Sessions are split into 3 categories: early, middle, and late. The early category contains sessions 1 to 7, the middle category contains sessions 8 to 14, and the late category contains sessions 15 to 20. This categorical coding of sessions is motivated by the non-linear trajectory of accuracy in [Figure 4C](#). LinearMixedModels package in Julia uses the first session category as the reference category. Thus in these tables “dms” can be interpreted as “session early & dms”.

In [Data S1](#) the variable denoting the striatal region (DMS, DLS, NAc) is the mean-subtracted session 0 striatal region’s contrast dependent stimulus response magnitude. (L2-norm of the difference of the 100% contralateral stimulus contrast response to the 6%.) session & <region> denotes the interaction of the variables session and <region>. The dependent variables:  $\beta_{contra}$ ,  $\beta_{ipsi}$ ,  $\beta_{bias}$ ,  $\beta_{choice}$ ,  $\beta_{history}$  correspond to the behavioral model choice weights.

In [Table S2](#) the variable *cohort* denotes the group identity of each mouse, either Chrmine or YFP. The variable *contra* denotes whether the trial corresponded to a stimulus contralateral from the recording site. The dependent variable *correct* is the side-specific (contra or ipsi) accuracy. Interactions and reference levels are as described above, thus the term: cohort: chrmine & contra:ipsi is the 3-way interaction of the reference level for session (sessions 1–7, e.g. “session early”), cohort, and contra.