nature communications

Article

https://doi.org/10.1038/s41467-024-53176-7

Dopamine release plateau and outcome signals in dorsal striatum contrast with classic reinforcement learning formulations

Received: 26 July 2023

Accepted: 3 October 2024

Published online: 14 October 2024

Check for updates

Min Jung Kim[1,2], Daniel J. Gibson[1], Dan Hu[1], Tomoko Yoshida[1], Emily Hueske[1], Ayano Matsushima[1], Ara Mahar[1], Cynthia J. Schofield[1,3], Patlapa Sompolpong[1,4], Kathy T. Tran[1], Lin Tian[5] & Ann M. Graybiel[1]

We recorded dopamine release signals in centromedial and centrolateral sectors of the striatum as mice learned consecutive versions of visual cue-outcome conditioning tasks. Dopamine release responses differed for the centromedial and centrolateral sites. In neither sector could these be accounted for by classic reinforcement learning alone as classically applied to the activity of nigral dopamine-containing neurons. Medially, cue responses ranged from initial sharp peaks to modulated plateau responses; outcome (reward) responses during cue conditioning were minimal or, initially, negative. At centrolateral sites, by contrast, strong, transient dopamine release responses occurred at both cue and outcome. Prolonged, plateau release responses to cues emerged in both regions when discriminative behavioral responses became required. At most sites, we found no evidence for a transition from outcome signaling to cue signaling, a hallmark of temporal difference reinforcement learning as applied to midbrain dopaminergic neuronal activity. These findings delineate a reshaping of striatal dopamine release activity during learning and suggest that current views of reward prediction error encoding need review to accommodate distinct learning-related spatial and temporal patterns of striatal dopamine release in the dorsal striatum.

Pioneering work has clarified much about dopamine signaling in the brain and about the remarkable relationship between this signaling and predictions of reinforcement learning (RL) algorithms. A canonical view[1] suggests that phasic dopamine signaling acquired during learning represents a reward prediction error (RPE). This view could be formulated in terms of a temporally sequential learning-related process, by which phasic responses originally are elicited by the reward, but these responses then decline as the phasic increase in activity is transferred to the most proximal cue predictive of reward[2]. This theory-based temporal difference (TD) formulation was recognized as having clear parallels to the patterns in electrical activity exhibited by dopamine-containing neuronal cell bodies in the midbrain recorded during learning tasks[3–7].

Subsequent studies building on this pioneering work have demonstrated heterogeneity in dopamine responses in the substantia nigra pars compacta (SNpc) and the striatum, the two poles of the nigrostriatal tract[8–13]. The emergence of chemosensor probes to record dopamine release in the striatum, nucleus accumbens and elsewhere was transformational in opening up detailed analysis of dopamine release[14,15]. Evidence now supports the view that dopamine release in

[1]McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar St., Cambridge, MA 02139, USA. [2]Advanced Imaging Research Center, University of Texas, Southwestern Medical Center, Dallas, TX 75390, USA. [3]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. [4]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [5]Max Planck Florida Institute for Neuroscience, Jupiter, FL 33458, USA. e-mail: graybiel@mit.edu

the striatum can be controlled locally and has suggested novel mechanisms of control. For example, spike activity of nigrostriatal fibers can be triggered within the striatum by cholinergic inputs acting at nicotinic acetylcholine receptors on the dopamine-containing fibers[16–18]. Dopamine release is reported to occur in waves moving across the width of the striatum in ~200 ms[19]. The release can exhibit low frequency oscillations during or even without task engagement, gated by extrinsic striatal afferents[20]. Topographic differences also exist. Striatal dopamine release responses can be different in different striatal sectors, prominently so between the medial and lateral regions of the dorsal striatum in mice[11,21–23]. Such differences have been reported for other topographic dimensions as well[12,24–26]. The dopamine release signals can be principally related to negative as well as positive reinforcement[5,27–29] or to non-reward parameters of movement[21,30,31], can occur as prolonged ramping signals[32], and can be compartmentally selective for striosome and matrix compartments of the striatum[18,33–36]. Especially for the nucleus accumbens (ventral striatum), but also for the dorsal striatum, the relation of the release patterns to RPE-TD learning algorithms has been strongly questioned[37–39] and strongly defended[10,23,30,37,39–47].

We took up this issue for the central part of the dorsal striatum (caudoputamen) by training mice consecutively on a series of cue-association tasks and recording dopamine release population-level responses with dopamine sensors throughout the time that the mice were learning the tasks. Mindful of the complexities of dopamine signaling, we nevertheless looked for patterns of activity similar to those recorded in classic work on the activity of dopamine-containing cell bodies in the SNpc[3]. We found shifts in the patterning of dopamine release signals as successive versions of the cue-association tasks were acquired, and sharp differences in the dopamine release patterns between the centromedial and centrolateral striatal sites from which we recorded in the 67 mice sampled.

Notably, outcome did not evoke transient dopamine increases in the centromedial sites. Over time, the cue responses declined, rather than increasing. The centrolateral sites did exhibit both cue and outcome responses, but they failed to exhibit a shift from primarily signaling outcome to primarily signaling cue, a canonical feature of RPE algorithms applied to the nigral dopamine system. Finally, prolonged plateau release responses to cues predicting reward emerged when the mice shifted from simple cue-association conditioning to more cognitively demanding cue discrimination conditioning, and these plateau responses appeared both medially and laterally and were evident in somewhat modified form through the cue reversal and probabilistic reward training sessions. To verify our expectations of what RPE signals should look like in these tasks, we constructed a simple Q-learning model, which prominently features a RPE signal. The discrepancies that we found between the striatal dopamine release responses recorded and the expectations based on cell body recordings in the dopamine-containing midbrain encourage further review of these classic algorithms.

Results

Experiment design

We recorded real-time dopamine release by photometry with D1 or D2 dopamine receptor-based sensors[14,48] placed in the centromedial or centrolateral sites of 67 mice (41 male and 26 female) that learned and performed a series of consecutively presented tasks with visual cues to instruct reward availability (Fig. 1a–c and Supplementary Data 1–3). These included, first, random reward presentation, and then, in succession, single-cue association conditioning, cue discrimination for two cues, reversal learning, probabilistic reward leaning, and extinction. In each task, only a single cue was presented at a time, either to the left or to the right of the mouse. For simple cue-association learning, the right or left cue, randomized across mice, was associated with reward. For the cue discrimination tasks, again only one was shown in any given trial, but two cues could be presented, one at a time. The same cue (left or right) that had predicted reward during the cue-association task was still the cue predicting reward, but its presentation alternated semi-randomly with the presentation of another cue on the other side, and it was not associated with reward. These contingencies were reversed during reversal discrimination. In the probabilistic reward task, one cue was associated with reward on 100, 75 or 50% of trials.

Each mouse was implanted with a single optic probe in either the centromedial sector (36 mice) or centrolateral sector (31 mice) of the dorsal striatum in the right hemisphere (Fig. 1b and Supplementary Data 4). This unilateral, single-probe recording protocol was chosen to minimize potential damage to the striatum and damage to the overlying neocortex due to the insertion of the probe, important given the extended periods of chronic recording required for the mice to complete the six different tasks.

Training

The sequence of training on different task variations was chosen to facilitate rapid learning, compatible with the maintenance of high signal quality across the many task versions. Given that the training protocol was similar for all mice, initial random reward sessions, usually 2-3 in number, were given to accustom the animal to receiving reward at the delivery spout (Fig. 1c). Random reward probe sessions were also included to allow assessment of dopamine release responses to unpredicted rewards both early and late in training.

After the initial random reward training, a visual cue was introduced to signal upcoming reward delivery. Next, a different visual cue was added that was not rewarded, to constitute the basic cue discrimination task. As the weeks of training accumulated, some sensors in some mice suffered from bleaching and/or fouling of the sensor tips. Our yields were accordingly reduced. This temporal limit on probe life meant that in any given mouse it was not possible to test different sequential training protocols. We are aware of the possible effects of our training regimen with a fixed sequence of paradigms across animals, and that our data might partly reflect such order effects.

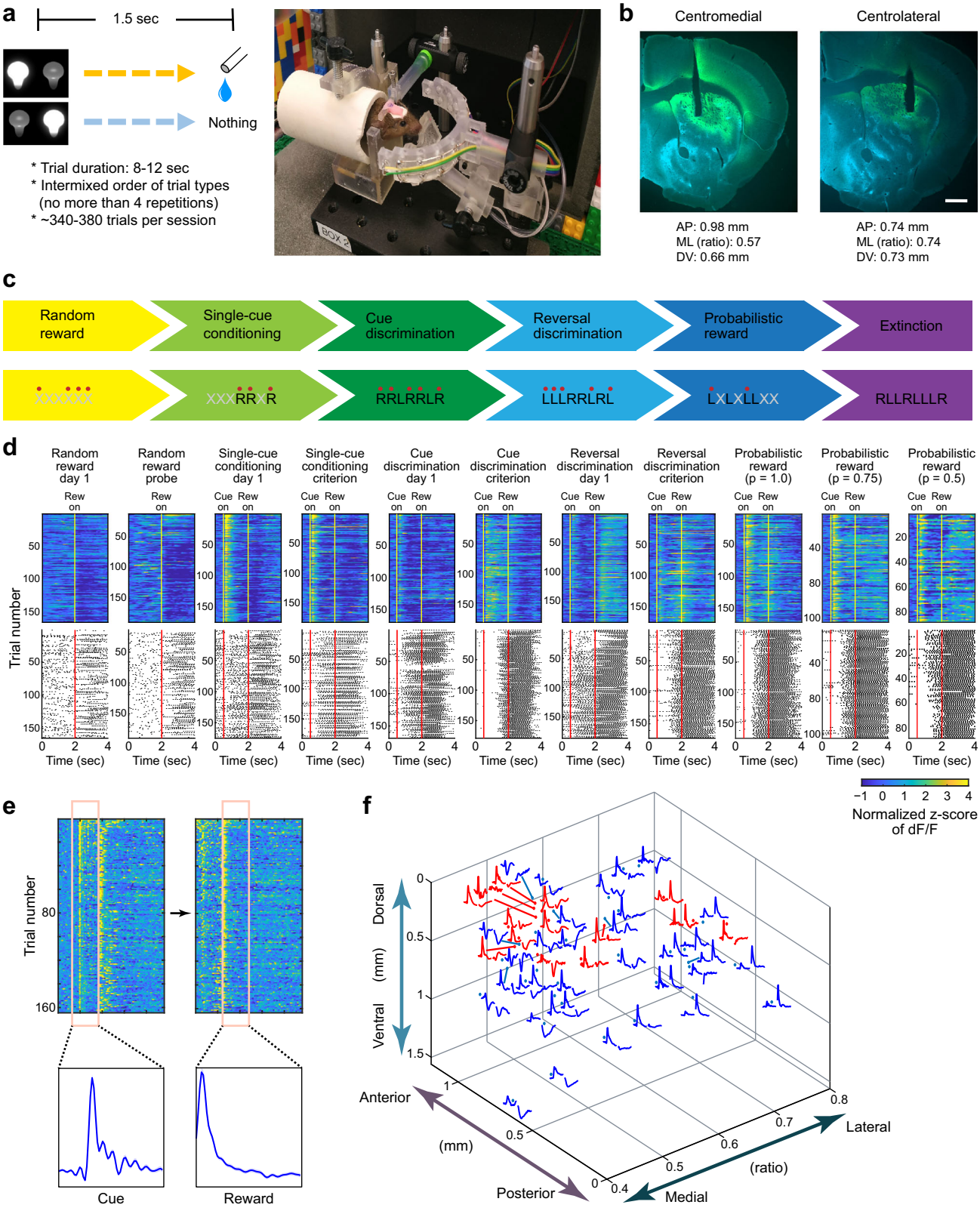
This phase was followed by sequential training stages, beginning with reversal training, in which the rewarded and non-rewarded cues were switched, and then probabilistic reward delivery following the most recently (i.e., reversed) rewarded cue without any presentations of the most recently non-rewarded cue. Finally, the animals received extinction training, but we do not present those data here due to frequent signal quality issues. Throughout the full sequence of task versions, random reward sessions were inserted every 9 or 10 sessions to determine whether the release responses were affected by the progressive learning stages (Methods).

Fig. 1d illustrates for one mouse the dopamine release patterns at the centromedial sites and corresponding licking activity during key sessions throughout learning. In sessions selected for intensive analysis (the first, last, and middle sessions for each task variation for each mouse), a whole-session average response was computed by averaging across trials as shown in Fig. 1e (see Methods). Cue responses and reward responses were averaged separately to permit the alignment of reward responses on the animal's first lick following reward delivery. Such session averages are shown in Fig. 1f for each mouse for the location of the mouse's probe tip. We did not observe different results across the data sets acquired with the D1- or D2-based probes, and we merged these for the analyses. We also found no differences between ipsilaterally and contralaterally presented cues (Supplementary Fig. 2) and merged these as well.

Dopamine responses varied with striatal sub-region

Reward-evoked dopamine in the centromedial and centrolateral sites (Fig. 2a) exhibited marked differences. In the centromedial sites, the responses to randomly delivered rewards at first dipped below

https://doi.org/10.1038/s41467-024-53176-7



baseline release levels, then flattened to low positive levels. By contrast, the dopamine release signals evoked in the centrolateral sites were at first small negative-positive blips that soon developed into strong positive, transient responses.

In the first cue conditioning sessions, in which the mice learned to associate the cue with reward, the centromedial sites and centrolateral sites again exhibited clearly different release dynamics (Figs. 1f and 2). In the centromedial sites, strong positive transient dopamine release responses occurred at cue, but responses at outcome (registered as the first lick after the reward delivery) were at first negative (below baseline) and then nil or weakly positive. The centrolateral sites, by contrast, exhibited strong phasic responses at outcome as well as at cue onset, and these responses remained stable throughout the sessions. Outcome signaling thus was exhibited in the centrolateral sites

Fig. 1 | Experiment design and data analysis. a Head-fixed apparatus with right and left visual cues and a drop of sucrose solution as a reward (right), and trial structure of an experiment session (left). **b** Examples of histological verification of dopamine sensor injection sites and optic probe locations in centromedial (left) and centrolateral (right) sites. Scale bar: 500 μm. Similar results were obtained for all 67 animals included in this study (see Supplementary Data 4). **c** Experimental design showing different phases of training (each with different task contingencies, top) and examples of task events in each phase for a mouse that initially received reward following the right (R) cue (bottom). Black letters indicate cue presentation on the specified side, red dots show reward delivery, and gray Xs indicate trials in which neither cue was presented. **d** Trial-by-trial data from centromedial sites in a single mouse, illustrating dopamine traces (top) and corresponding lick activity (bottom) for reward-predicting cues and rewards across learning sessions. See also Fig. 2b. Data from random reward sessions are aligned to the first lick following reward delivery (Rew on), and data from all other sessions are aligned to the cue onset (Cue on) with the 0.5 s precue, 1.5 s cue and 2 s reward periods. The "Random reward probe" session is one of those inserted during later phases of training (see Methods). **e** Construction of cue and reward response traces. For each session, cue data are aligned to cue onset, and reward data are aligned to the first lick after reward delivery. Note that the examples show two different sessions. **f** 3D reconstruction of dopamine responses to cues and rewards according to histological confirmations of probe locations in standardized coordinates (see Methods). Dopamine release was recorded with D1R-based (red) and D2R-based (blue) sensors during the first session of the single-cue conditioning. Illustration and panel arrangement by Johnny Loftus.

but was weak or negative-going in the centromedial sites. In both the centromedial and centrolateral regions, the cue responses were diminished in amplitude by the late training sessions, not enhanced as expected from classic RL-RPE accounts based on recordings of nigral dopamine-containing neurons. Most notably, we were unable to identify a so-called transfer of the dopamine signaling from outcome to reward-predictive cues as suggested by classical work based on electrophysiological recordings from the dopamine-containing cells of the SNpc[1,3,49].

Dopamine plateau responses emerged during discrimination training

We searched for such dynamics by continuing the training to require the mice to discriminate which of two cues was associated with reward. The originally rewarded cue was now randomly alternated with a second cue, which appeared on the opposite side and did not predict reward. Again, the dopamine responses recorded at outcome were nearly absent in the centromedial and centrolateral sites, outcome release transients did not decline with training, in sharp contrast to the predictions of classical RPE (Fig. 2a)[49]. We did see in the centromedial sites a positive response to cue onset at the very beginning of discrimination training, and an absence of response at reward delivery, both of which would be expected from the RPE model as a result of cue conditioning training; but we note that these did not evolve during cue conditioning as expected for RPE. The evolution of these responses, with discrimination training evoking the same or nearly the same responses but with a plateau phase added, cannot readily be accounted for in terms of RPE.

Remarkably, both centromedial and centrolateral cue responses became sustained. They persisted during most or all of the 1.5 s long cue period as learning proceeded, then fell at outcome in the centromedial sites and rose transiently at this cue-off/outcome-on time in the centrolateral sites. These plateaus emerged during the first days of discrimination training (Fig. 2b, c).

The development of the plateau-like response to the reward-predicting cue was not the product of averaging across mice; it could be seen in individual mice (Fig. 2d; also see below). In the example shown in Fig. 2d, at the start of cue discrimination training, both cues produced a slowly decaying dopamine response that extended well past the initial peak. As training proceeded, the response to the reward-predicting cue became larger and more sustained, whereas the response to the non-reward-predicting cue was nearly abolished by the third session. A similar process occurred during cue reversal training (see following), and by the time the mouse reached criterion, the long latency component of the dopamine response to the non-rewarded cue showed an anti-plateau, i.e., a sustained drop below baseline.

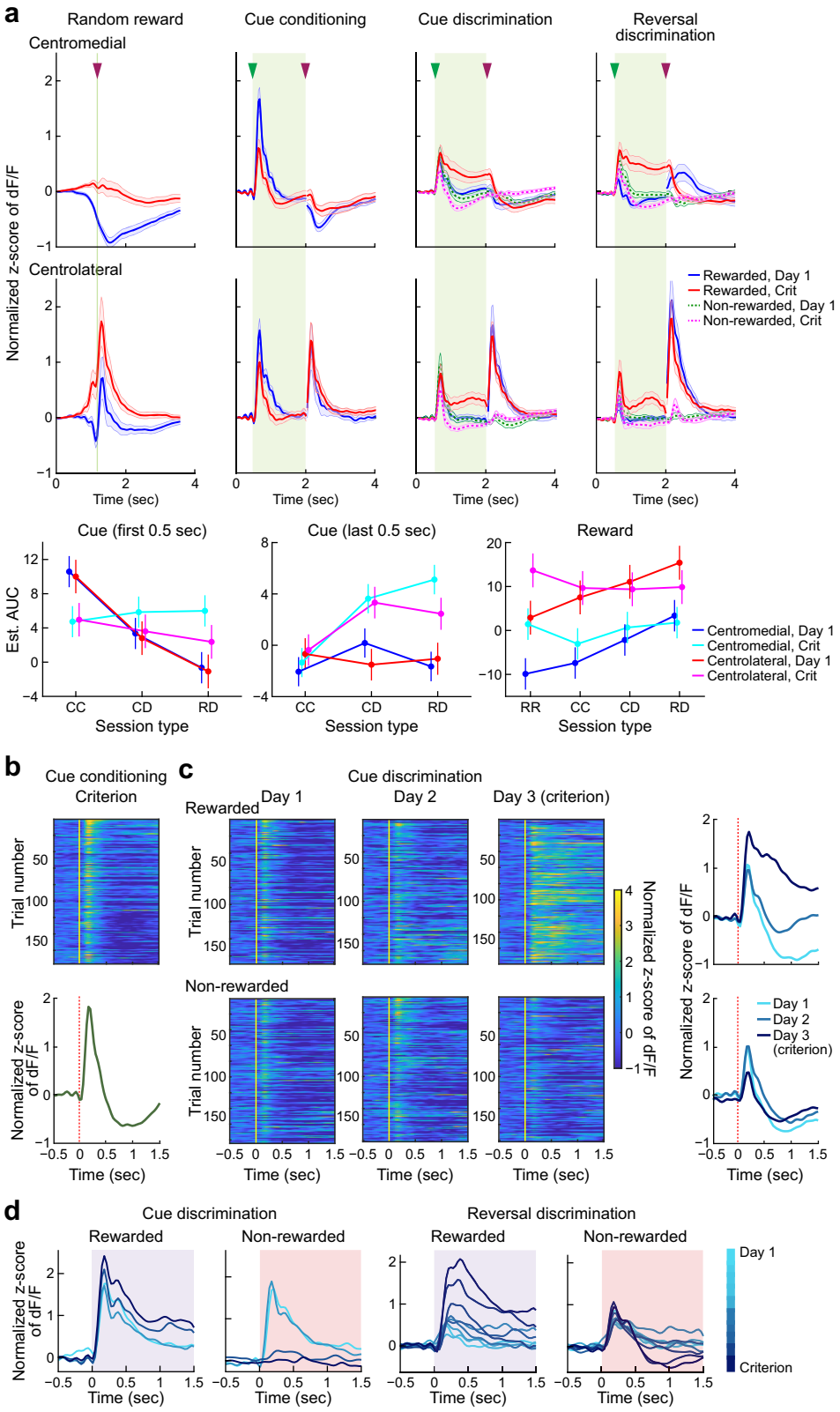
This emergence of plateau-like responses also was not a product of averaging across trials within single sessions. Fig. 3a illustrates single trial responses from a mouse (animal pa38, which had the fifth largest PC1 amplitude across all mice) in groups of ten trials per plot. The heavy black line in each plot indicates the average of the trials in that same plot. There is considerable volatility in the dopamine signal within each trial. Nonetheless, a relatively consistent pattern of higher maxima and higher minima during the cue, as compared to before the cue, can be seen for every set of ten trials. To get a better view of this pattern, we developed a method for fitting the centerline of the peaks and valleys of a signal by finding the upper envelope determined by the peaks and the lower envelope determined by the valleys, and then averaging the upper and lower envelopes. This method is illustrated in Fig. 3b. First, all 3-point local maxima and minima of the raw dopamine response waveform were found. Then upper and lower envelopes were constructed by, respectively, linear interpolation of the maxima and minima. Finally, the mean of the upper and lower envelopes was calculated, which we refer to as the "midline" of the waveform. Fig. 3c shows the same sets of trials as Fig. 3a but showing individual waveform center lines instead of the raw waveforms. The majority of waveform center lines echo the shape of the waveform averaged across trials.

After each mouse reached criterion for cue discrimination, the mouse was trained on a reversal discrimination task that required the mouse to learn that the formerly rewarded cue was now the non-rewarded cue. This task version again required cue discrimination learning for success in obtaining reward. Prolonged plateau dopamine release responses to cue occurred at both centromedial sites and centrolateral sites (Fig. 2a). In the centrolateral sites, outcome signaling was strong and steady. There was a substantial increase in the cue response with learning, consistent with an RPE signal, but only a minor decrease in outcome response, in contrast to the large decrease expected on the basis of RPE algorithms. More medially, a positive outcome signal emerged for the first time, but then waned with exposure across trials. This brief positive signal, unique to the beginning of reversal training, suggests that in some centromedial sites, outcome responses might be associated with reversing the previously learned association, a pattern compatible with dopamine serving as an RPE signal.

Single-cue probabilistic reward exhibited some RPE-like features

To characterize further the relationships between dopamine, learning, and RPE, we introduced a task version with probabilistic reward sessions, in which one cue (the same one that was rewarded in the preceding reversal training) always signaled potential reward, but with varying probabilities (Fig. 4a, b). The responses in the centromedial sites, instead of being dominated by a single sustained plateau, now had two components: an early strong transient followed by a much lower amplitude sustained plateau response that was greater than baseline for the high-probability conditions, but nearly zero for the 50% rewarded condition. There were dips at the no-reward outcomes, but little or no response to reward outcomes. The centrolateral sites, by contrast, again exhibited stability in their strong dopamine release

https://doi.org/10.1038/s41467-024-53176-7



transients both at cue and at positive outcomes, with dips for non-rewarded trials. Notably, the magnitude of the centrolateral outcome responses clearly scaled with probability of reward, systematically increasing with lower probabilities of reward during the probabilistic reward sessions, as though scaling with uncertainty (largest response at 50% probability). These response patterns, along with the more medial region's reward response early but not late in cue reversal training, were unique in the present data set in being consistent with an RPE interpretation.

At the transition from cue reversal learning to 100% probabilistic reward sessions, the average dopamine response changed in shape (Fig. 4c). The initial peak increased in height by nearly twofold, and the plateau height dropped by a similar amount. Based on the GLMM interaction model and the pairwise comparisons, responses to the

Fig. 2 | Distinctive characteristics of dopamine release signals in dorsal striatal subregions evoked during a series of conditioning tasks. a Learning-related effects on the dopamine responses (mean ± 2 SEM) to the cue (green arrowheads) and reward (purple arrowheads) in centromedial (top) and centrolateral (middle) sites. Averaged traces for all mice from the first (Day 1) and criterion (Crit) sessions for each session type are shown. Data from random reward sessions, which included those inserted late in training, are aligned with reward delivery at t = 1.0 s. Plots for other session types span 0.5 s precue, 1.5 s cue (light green shade) and 2 s reward periods. For discrimination learning, average traces of non-rewarded trials are shown with dotted lines. The small oscillations preceding cue onset in some plots are artifacts from low-pass filtering. Summary graphs of post-estimation by GLMM shows the estimated fit (EF) and 95% confidence limits (CL) for dopamine response in early (first 0.5 s) and late (last 0.5 s) cue periods and reward period of random reward (RR), cue conditioning (CC), cue discrimination (CD) and reversal discrimination (RD) sessions. Dopamine responses were quantified as AUC of the *dF/F* data (see Methods) for the rewarded trials. *N* = (top row) 36, 36, 35, 31; (second row) 31, 31, 28, 23; (bottom row, blue/cyan) RR: 36, CC: 36, CD: 35, RD: 31; (bottom row, red/magenta) RR: 31, CC: 31, CD: 28, RD: 23. **b**, **c** Transition from initial cue-association (**b**) to cue discrimination (**c**) training shown for a single mouse (pa96, centromedial). In **c**, dopamine response in rewarded (top) and non-rewarded trials are shown. Vertical lines indicate cue onset. Dopamine traces aligned as in Fig. 1d, e illustrate responses in consecutive sessions from last cue conditioning (**b**, bottom) to first three cue discrimination sessions (**c**, right), showing the gradual emergence of prolonged plateau dopamine release during the cue presentation period in rewarded trials of cue discrimination sessions. Color scale shows z-scores of *dF/F*, ranging from −1 to +4. Color-coded line plots are shown on the right. **d** Superimposed session-averaged dopamine release in response to rewarded and non-rewarded cue onsets recorded in a mouse (animal pb43, centromedial) during all training sessions, from Day 1 (light blue) to criterion (dark blue), of cue (2 left panels) and reversal (2 right panels) discrimination training. Shaded purple and pink boxes indicate 1.5 s cue period.

rewarded cue changed significantly from reversal discrimination to probabilistic reward sessions in both the initial and later cue periods ($p < 0.0001$).

Dopamine release signals had higher amplitudes in the sample of female mice

To look for other possible factors contributing to the cue and outcome responses, we examined signal differences based on the animal's sex (Fig. 4d) and on the basis of their performance levels in the task versions (Fig. 4e). Across task versions, dopamine release level changes (both positive and negative) were larger for the females than for the males, but were similar in pattern, thus exaggerating the contrast between the centromedial and centrolateral dopamine release patterns in females (Fig. 4d).

Dopamine signals at centromedial striatal sites exhibited higher amplitudes in better performers

We also plotted the trial-averaged dopamine responses for the first and last sessions of cue conditioning and cue discrimination, averaged over the group of mice that performed better than median (better performers), and separately averaged over those that performed poorly (worse performers) (Fig. 4e). In centromedial sites, dopamine responses were strikingly different for the better and poorer performers in both the cue conditioning and cue discrimination tasks; they exhibited a strong difference between rewarded and non-rewarded trials in cue discrimination that was absent from the worse performers' patterns. In sharp contrast, the dopamine responses at centrolateral sites were similar regardless of performance level across all task conditions. Thus, the learning-related remodeling of dopamine release was a property of centromedial but was not detectable in the centrolateral striatal sites.

Dopamine release response amplitudes were positively correlated with levels of performance

The amplitudes of the initial peak at cue onset and the plateau response during cue presentation generally followed the levels of cue discrimination task acquisition for many mice (Fig. 5a), high for most of the good learners, less prominent for the middling learners, and not detectable in the mice that did not learn well or at all. The same trends were present for reversal discrimination responses to the previously non-rewarded cue (Fig. 5b). Each of these data sets was accompanied by analysis of the responses that the mice made to the non-rewarded cue. Many of the proficient and moderately good learners developed brief transient responses to the non-reward-predicting cue, but those peaks were not followed by plateaus. Correlation analysis showed that among mice that reached the learning criterion in the cue discrimination task, there was a highly significant ($p = 0.002$) correlation, when data from centromedial and centrolateral sites were combined, between learning index and the difference in area under the curve (AUC) of the dopamine responses to reward-predicting and non-reward-predicting cues (Fig. 5c). The centromedial and centrolateral sites both showed positive correlations, but these were not statistically significant laterally. A similar result was obtained for the grouped data for the cue reversal task ($p = 0.02$; Fig. 5d). Thus, in this comparison, positive but non-significant correlations were found for both sites, but an overall correlation for the aggregated data was significant.

Principal component analysis supported the presence of plateaus

The existence of a variable plateau-like component in the dopamine responses was further confirmed by performing principal component analysis (PCA) on the cue discrimination data (Fig. 6) to identify correlated components of variance across waveforms, sorted in order of decreasing variance explained. Each mouse was represented in the PCA by the response waveform during the cue period averaged over all rewarded trials in that mouse's first (Day 1) and final (Criterion) sessions of discrimination training, and an analogous averaged waveform was calculated for the reward period. Fig. 6a shows the waveforms for the cue period, with the grand average across all mice shown in black, and the first three principal components of the variance (PCs) in different colors. The average waveform had a prominent plateau at about half the amplitude of the initial peak. The shape of the first PC, in red, indicated that the height of the plateau could vary independently from the height of the initial peak, and the higher the plateau was, the more it tended to decay slowly over time. The first PC for this data set accounted for most of the variance across mice during both the cue and outcome periods, and therefore across recording locations (Fig. 6a, right). For the reward period (Fig. 6b), PC1 again accounted for more than half the variance in the waveforms and demonstrated that a slowly decaying plateau was the chief type of variation across different waveforms. There was a slight correlation between plateau height and peak height, but the prominence of the peak in both PC2 and PC3 indicated that there was also considerable independence between peak height and plateau height.

Because PCs account for correlated components of variance across all dimensions (i.e., all time points, in this case), it is tempting to assume that each PC ultimately identifies a single causal source of variance. This is not a logically necessary inference, but it is generally difficult to find an alternative explanation for how two or more sources of variance can be correlated unless they do in fact share some ultimate common cause. We thus tentatively interpret PC1 as reflecting an input to dopamine release that is relevant to the majority of mice, and the other PCs as representing factors that may be relevant in smaller numbers of cases.

https://doi.org/10.1038/s41467-024-53176-7

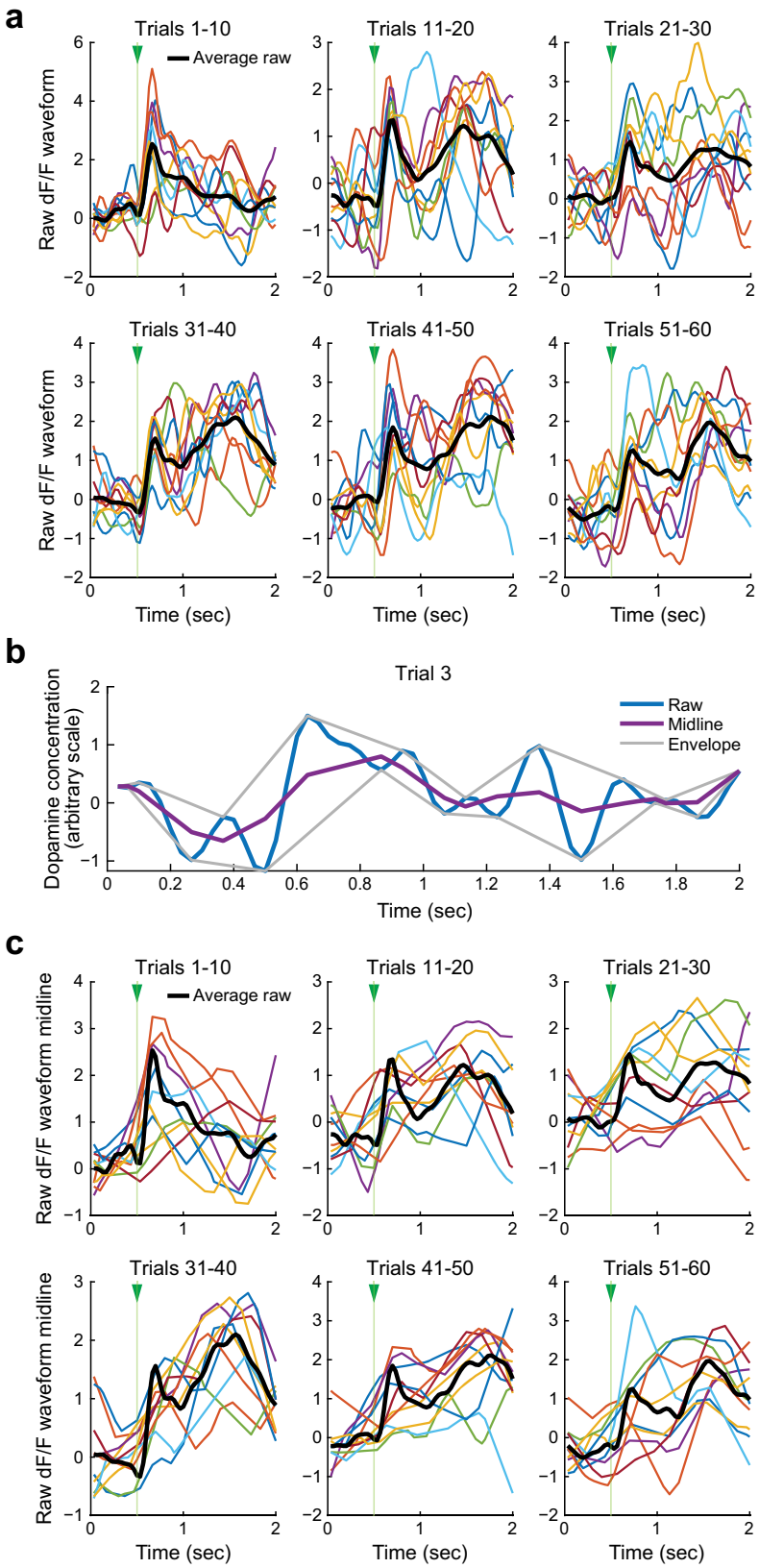


Fig. 3 | Single-trial dopamine signals (animal pa38, centromedial). a Raw single trial recordings (colored traces) from the criterion session of cue discrimination training. Each panel shows a different group of ten trials. Heavy black traces represent the average of all single trial traces shown in the same panel. **b** Method for finding the centerline of a raw single trial waveform. **c** Centerlines (colored traces) of the same single trial recordings shown in **a**. Heavy black traces are reproduced from **a** for ease of comparison. Note that the vertical scales may differ between **c** and **a**.

https://doi.org/10.1038/s41467-024-53176-7

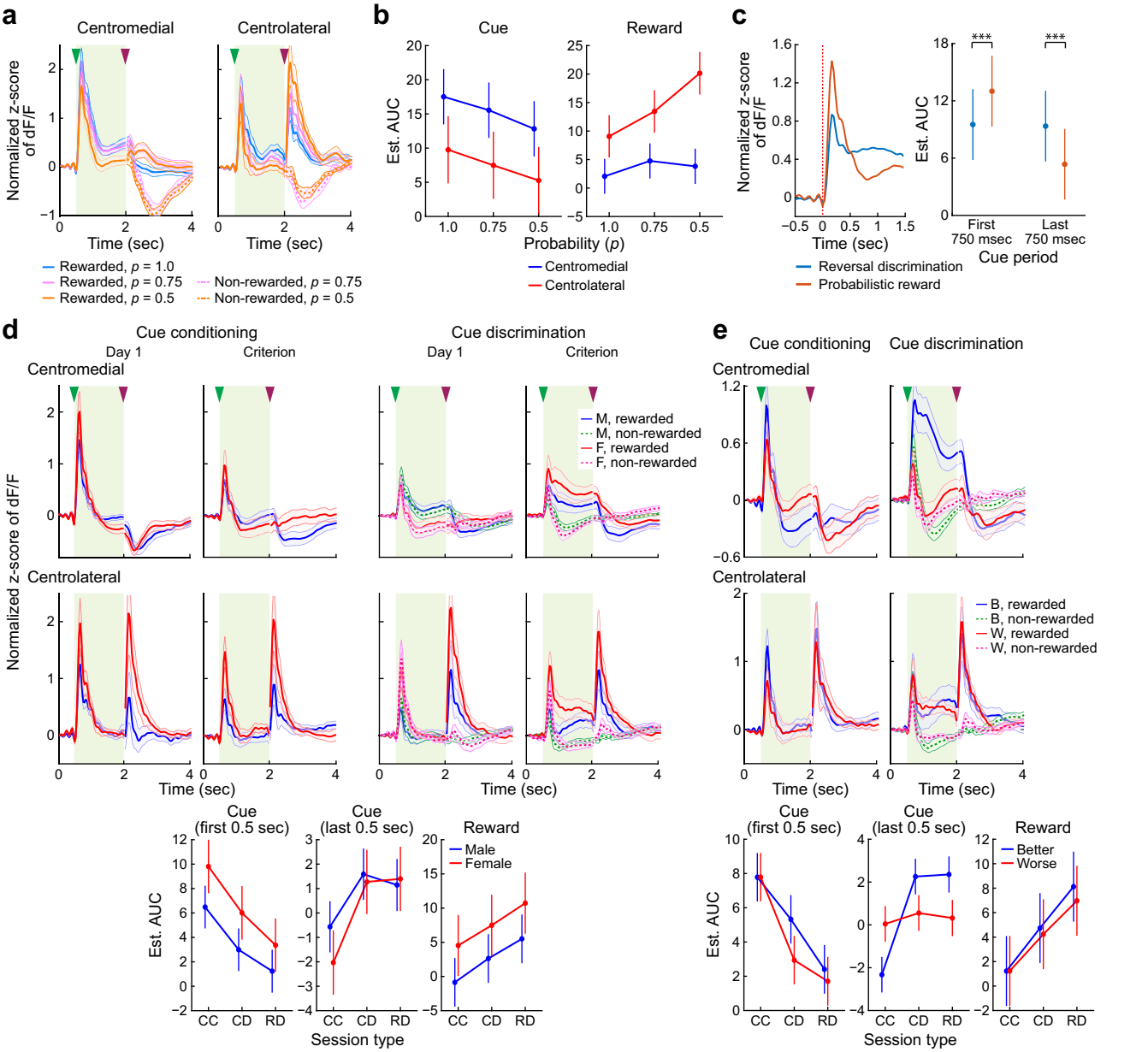


Fig. 4 | Effects on dopamine response of reward probability, biological sex, and discrimination performance. a Dopamine response (mean ± 2 SEM) to cue and reward during the probabilistic reward sessions recorded in the centromedial and centrolateral regions, shown separately for rewarded (R) and non-rewarded (NR) trials with three different reward probabilities: 0.5, 0.75 and 1.0. Green and purple arrowheads represent, respectively, cue and reward onsets. Ns for reward probabilities of 1.0, 0.75 and 0.5 were, respectively, 25, 22 and 24 for centromedial; and 17, 16 and 16 for centrolateral. **b** Dopamine response AUC values (EF ± 95% CL) for cue and reward periods in sessions with different reward probabilities, in relation to the recorded sites. Ns are identical to those in **a**. **c** Left: Average dopamine response to reward cue during the last reversal discrimination session and the first probabilistic reward (PR) session. Right: The post-estimation of GLMM of dopamine signal (EF ± 95% CL) for the reward cue showing significant differences in early (the first half) and late (the second half) cue periods between reward discrimination and probabilistic reward sessions. $N = 54$. ***$p < 0.0001$, computed by R's glmmTMB package. **d** Average dopamine traces (mean ± 2 SEM) recorded during the first (Day 1) and last (Criterion) sessions of cue conditioning (CC) and cue discrimination (CD) learning at centromedial (top) and centrolateral (middle) sites. Traces are color-coded for male (M) and female (F) mice, and rewarded and non-rewarded trials. $N = 41$ (M) and 26 (F) for cue conditioning; and 37 (M) and 26 (F) for cue discrimination. Bottom row depicts post-hoc estimation of dopamine response AUC (EF ± 95% CL) for first and last third of cue period, and entire reward period, calculated from the rewarded trial data. Day 1 and Criterion sessions were combined for CC, and separately combined for CD, for each AUC period. N (CC, CD, RD) = 41, 37, 28 (M); 37, 26, 26 (F). **e** Top two rows: average dopamine responses (mean ± 2 SEM) recorded in the last (Criterion) sessions for cue conditioning and discrimination learning plotted for better than median (B) and worse than median (W) performers, and for rewarded and non-rewarded trials. Left (CC) N = B-medial: 16, B-lateral: 17, W-medial: 19, W-lateral: 14. Right (CD) N = B-medial: 19, B-lateral: 13, W-medial: 16, W-lateral: 15. Bottom row: Post-estimation of response (EF ± 95% CL). AUC for cue and reward periods from the rewarded trial data, as in **d**. N for CC, CD as above. N (RD) = B-medial: 18, B-lateral: 8, W-medial: 11, W-lateral: 14.

Spatial maps delineated regions of greater and lesser plateau responsivity

It was clear by eye that the prominence of the plateau components following both cue onset and reward delivery differed across different subregions of the striatum. We therefore constructed spatially smoothed maps illustrating the variation in average amplitude from multiple probes that were in the same vicinity (see Methods). Such maps are shown for PC1 amplitude projected onto the three

https://doi.org/10.1038/s41467-024-53176-7

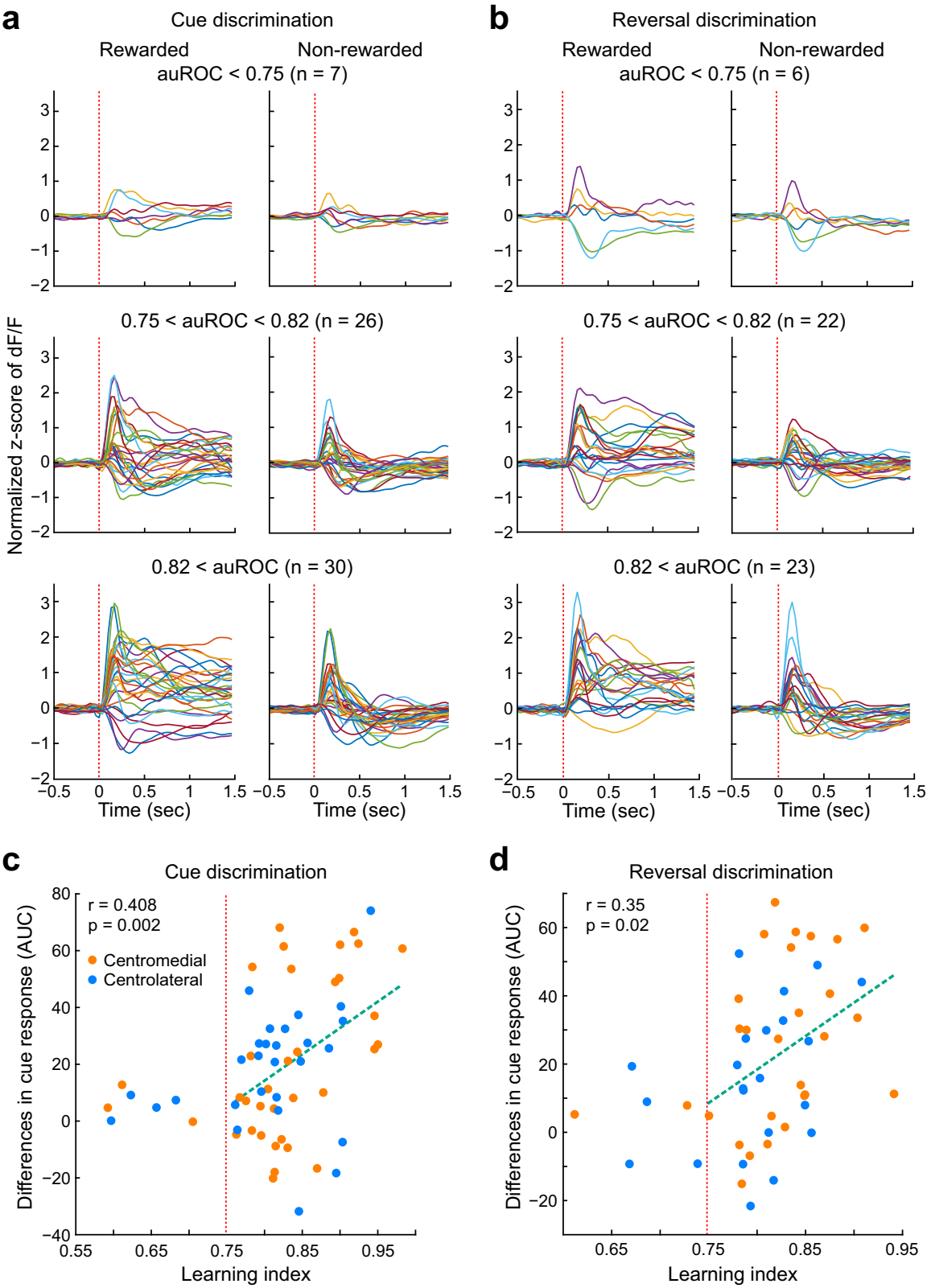


Fig. 5 | Dopamine plateau and discrimination learning. a, **b** Dopamine responses to cues predictive of reward (left) or no reward (right) shown for mice (both centromedial and centrolateral) according to their final performance rates (top to bottom, poor to best learners) during cue discrimination (**a**) and reversal discrimination (**b**) sessions. Each trace represents one mouse. Vertical line shows time of cue onset. **c**, **d** Scatter plots of learning index and differential cue responses between rewarded and non-rewarded trials during the last session of cue discrimination (**c**) and reversal discrimination (**d**). Each point represents one mouse (orange for centromedial sites, blue for centrolateral) with Pearson's correlation coefficient (r) and the corresponding *p*-value. The vertical red line indicates learning criterion.

https://doi.org/10.1038/s41467-024-53176-7

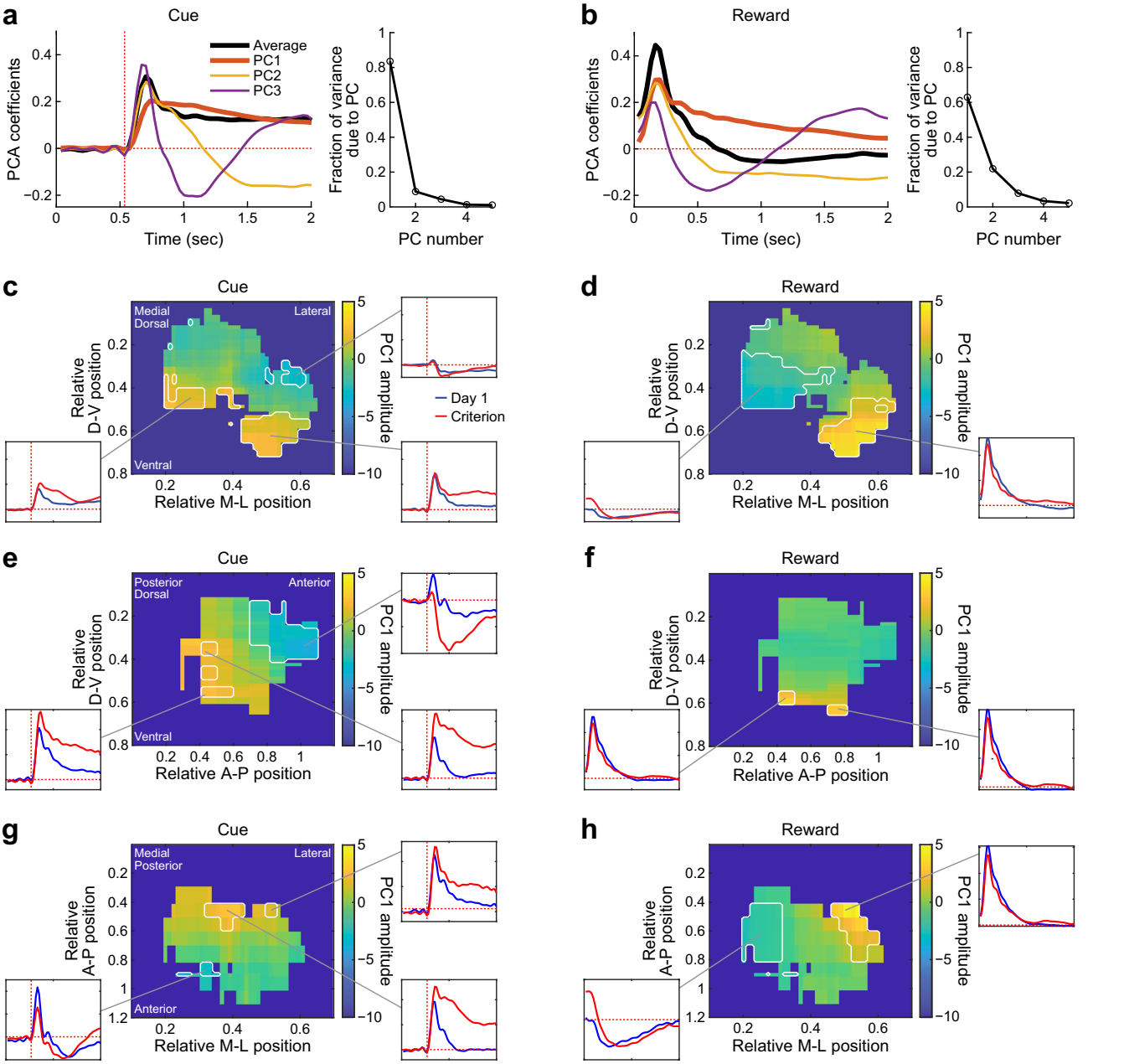


Fig. 6 | Amplitude of first principal component of fluorescence waveform as a function of probe position in striatum after discrimination learning. a Left: Average response waveform across all mice in the last session of discrimination training (i.e., the "learned" state), and first three principal component waveforms, during the cue period. Average waveform has been normalized to unit magnitude to match the scale of the principal component waveforms. Vertical dotted line indicates cue onset. Right: Fraction of waveform variance across mice explained by each of the first five principal components. **b** Same as **a**, but for reward period, starting at reward onset. **c** Anatomical distribution in coronal plane of first principal component amplitudes during cue period. Boundary between centromedial sites and centrolateral sites was in the 0.4 to 0.5 range, depending on the A-P position of each section. Color scale shows average amplitude of PC1 across all recording sites that were within 8 spatial bins of each point. Only points that had at least 3 recording sites contributing to the average are plotted; other points are shown in dark blue. White outlines show regions that were significantly different from median ($p < 0.05$, two-tailed bootstrap). Inset plots show fluorescence waveforms recorded in the first (Day 1) and last (Criterion) sessions of discrimination training, and averaged across probes that contributed to the average in the middle of the indicated region. All insets are shown with z-score normalized dF/F ranging from −0.5 to 2.7 on the vertical axis, and time spanning 2 s on the horizontal axis. Dotted vertical line indicates cue onset, dotted horizontal line marks dF/F = 0. **d** Same as **c**, but for responses in reward period. Insets do not include the time of cue onset, but start at reward onset as in **b**. **e**, **f** Same as **c** and **d**, but projected onto sagittal plane instead of coronal. **g, h** Same as **c** and **d**, but projected onto horizontal plane instead of coronal.

cardinal anatomical planes, aligned at cue onset (Fig. 6c, e, g) and reward delivery (Fig. 6d, f, h), with shades of yellow representing high values of PC1 amplitude, and shades of blue or green representing low or negative amplitudes. The changes across learning stages from early training to acquisition were equally striking (Supplementary Fig. 1). The maps in all three anatomical planes exhibited substantial spatial variation, so that across the 3D extent of the central dorsal striatum, there were districts with strong changes in plateau levels and learning-related development and others where they were not so prominent. The PC1 component of the response waveforms was significantly higher in our more ventrolateral sites for both cue and reward periods, but in the more ventromedial sites, it was

https://doi.org/10.1038/s41467-024-53176-7

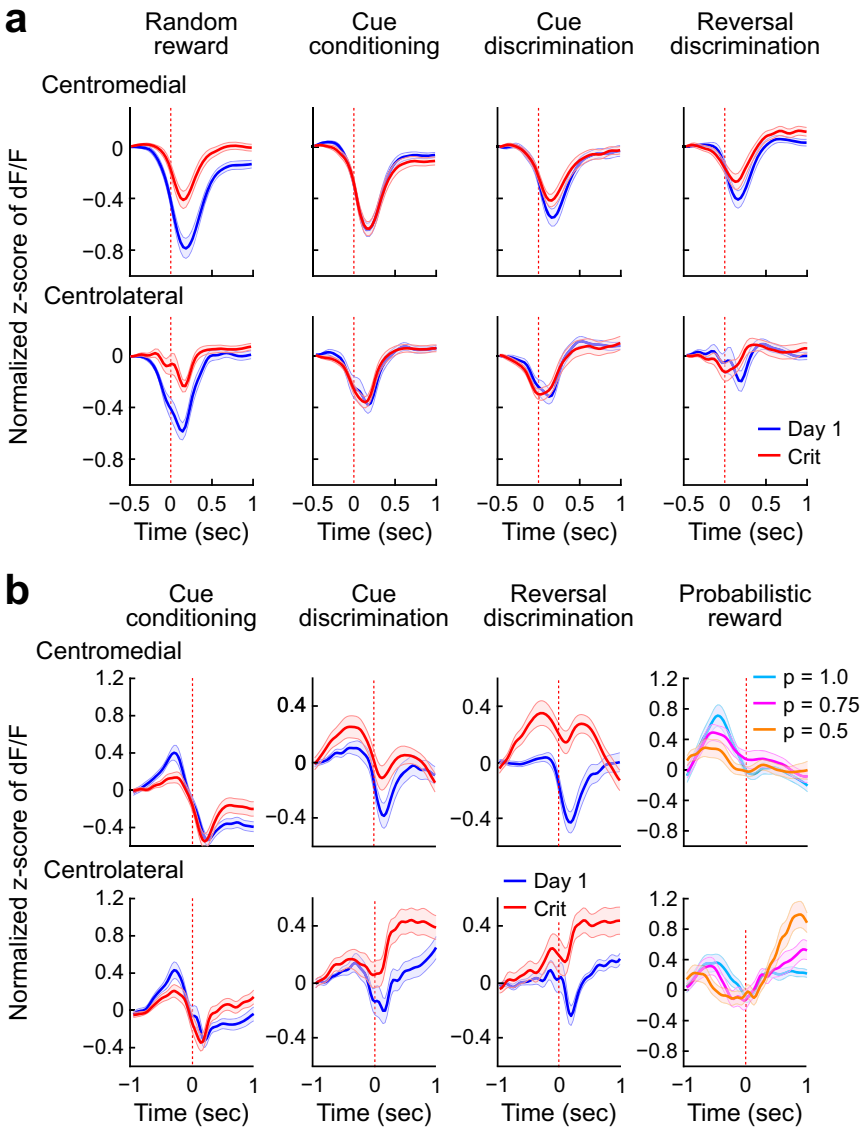


Fig. 7 | Average dopamine release in relation to licks. a Averaged *dF/F* dopamine traces (±2 SEM) aligned to spontaneous, out-of-task licks (red vertical lines) recorded in the first (Day 1) and last (Crit) sessions of each phase of training. *N* = (top row) 36, 36, 35, 31; (bottom row) 31, 31, 28, 23. **b** Averaged *dF/F* traces (±2 SEM) aligned to first lick after cue onset. Color code for last column represents the probability of receiving reward after presentation of the reward cue. *N* = (top row, first 3 col) 36, 35, 31; (bottom row, first 3 col) 31, 28, 23. *N* = (top row, last col; $p = 1.0$, 0.75, 0.5) 25, 22, 24; (bottom row, last col; $p = 1.0$, 0.75, 0.5) 17, 16, 16.

significantly lower in the reward period and significantly higher in the cue period (Fig. 6c, d). The distributions of dopamine response waveforms were thus quite different following cue onset and reward delivery. Given that there were 67 probes, and that some mice did not complete training on the entire series of task variants, dividing the probe population into halves along all three dimensions would have produced 8 octants each containing fewer than 8 probes. We therefore did not attempt a more refined spatial analysis of response properties after completing the comparison of the centromedial probes with the centrolateral probes.

Dopamine decreased in response to first licks

Dips in dopamine release going below baseline levels occurred early on in cue discrimination training sessions at the end of the cue period (beginning of the availability of reward; Fig. 2a). These were also present in the random reward sessions. We therefore asked whether the licking patterns themselves could have been important in shaping the release response profiles as the mice adjusted these patterns and formed stereotyped licking patterns toward cues and reward. We aligned the dopamine signals relative not only to the first lick after reward availability (end of cue), as in Fig. 2, but also to the spontaneous, un-cued licks that occurred during the inter-trial intervals, identified as having at least a 0.5 s period without licks prior to the spontaneous lick (Fig. 7a); and also to the first lick after cue-onset (anticipatory licking; Fig. 7b). In all instances, the dopamine responses were negative, occurred at both centrolateral and centromedial sites and were generally greater medially. The magnitude of this negativity waned for both types of licks as sessions continued (Fig. 7a, b). In the cue conditioning sessions, there was very high dopamine release in the early trials, followed later by very large reductions at the first lick after cue onset, both medially and laterally. The high dopamine release diminished during single-cue association training, as did the decrease at first lick. In both the subsequent cue discrimination and cue reversal discrimination tasks, there was little change in dopamine release (Fig. 7b; note differences in vertical scales).

Simple Q-learning RL model cannot account for our observations

It was clear that the absence of transfer of transient dopamine from outcome to predictive cue in the centromedial sector and the absence there of a positive outcome response were not in accord with classic RPE models[49]. But this was not so clear for the plateau responses. To assess this issue, we constructed a simple Q-learning model (see Supplementary Fig. 3 and Supplementary Note 1). As expected, this simple model soundly replicated the well-known RPE signals. However, it did not replicate plateau responses, especially the fact that they emerged when the second, unrewarded cue light was added (Supplementary Fig. 3). Kim et al.[50] have found that the prolonged ramping dopamine signals reported experimentally[32] can represent RPE signals when temporal discounting (discount rate in ref. 2) is a factor. In our data, plateau responses varied in shape from trial to trial (Supplementary Fig. 3), but in the aggregate and in most individual mice they were more flattened than ramp-like, and in some animals, even resembled reversed (decreasing) ramps, indicating that temporal discounting was not responsible. Our RL model also did not account for why plateau responses were absent in the cue association task and abruptly emerged when the alternative but never simultaneously presented cue was presented in the sessions (Supplementary Fig. 3). As a test for whether the order of task paradigms could be important for these responses, we switched tasks in the Q-learning model so that the cue discrimination preceded rather than followed the cue association sessions. The model did not behave as the mice did: it gave transients at both cue and outcome. More stringently, any RPE model is strongly challenged to explain the decrease of cue-associated phasic response as mice learn the task, or the increase of reward-evoked responses in random reward paradigm, even if they incorporate high eligibility trace (lambda) and sensory uncertainty[50,51]. Future studies should address how these issues could be addressed in more complex models that might account for the behavior and dopamine signals reported here.

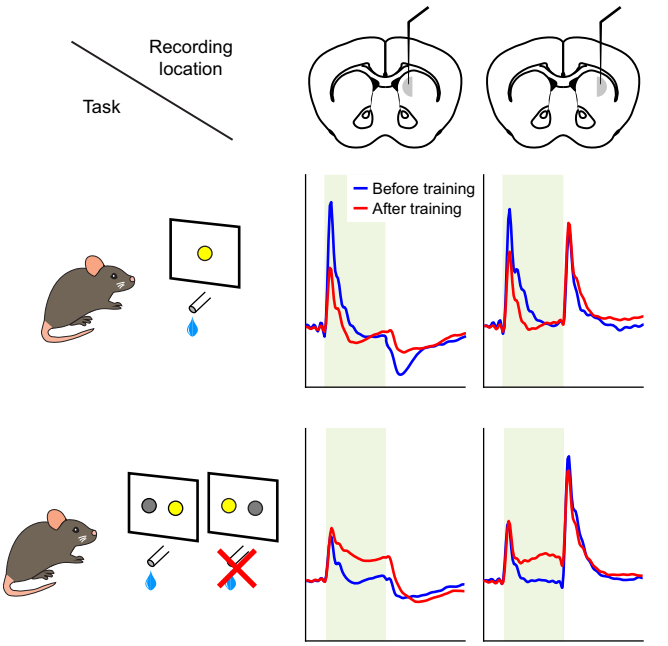


Fig. 8 | Summary of main results. Cartoons at left show two different tasks (top: cue conditioning; bottom: cue discrimination). Coronal section drawings at top show the general recording locations for centromedial (left) and centrolateral (right) sites. Waveform plots show time courses of dopamine responses on rewarded trials before (blue) and after (red) training for all four combinations of task and recording location. Plateau responses develop with training on cue discrimination but not on simple cue conditioning. Centromedial sites do not show dopamine release to reward delivery. Reward delivery in both tasks elicits the same dopamine release laterally after training as before training. Cue presentation in cue conditioning elicits less dopamine release after training than before training, whereas in cue discrimination the response remains unchanged.

Discussion

Our experiments with simple Pavlovian tasks lead to three major findings that suggest the need to review current RL-RPE models of dopamine's functions in the striatum (Fig. 8). First, in the centromedial dorsal striatum, dopamine exhibited no or negative reward-associated outcome responses, in contrast to what was expected based on the RPE interpretation of dopamine activity. Also, in both centromedial and centrolateral striatum, the dopamine response to random reward increased with training, whereas an RPE signal would decrease with training. The RPE model is thus not sufficient to account for these data. Second, phasic dopamine release responses did occur to both the conditioned cue and to reward outcome in the centrolateral striatum, but with training the outcome response did not decline, and the cue response did not increase, also in contrast to the RPE interpretation of dopamine activity. Third, with discrimination learning, plateau-like responses, which tended to bridge the cue and reward associated responses, emerged and were strongest in the best performers, but almost absent in non-learners. Simple RL models, prima facie, do not predict the emergence of such responses, though with complex RL models they might appear (see below). We conclude that, at least at the population level that can be imaged by fiber photometry, dorsal striatal dopamine release responses do not fully follow RPE formulations in either more medial or more lateral regions, but exhibit instead unpredicted heterogeneities that we have shown in detail along the mediolateral dimension, and indicated briefly in Fig. 6 in the other two dimensions. These findings encourage further work on how the multitudes of striatal circuits are coordinated to instruct learning and to modulate behavior under the influence of dopamine.

Dynamic shaping of striatal dopamine release responses during learning

We recorded dopamine responses daily during learning and found across samples of these recordings strong evidence that dopamine release profiles in the dorsal striatum undergo learning-related changes with transient increases and dips, as are well known, but also prolonged plateau release responses. These different release profiles were differentiable both by their striatal region and, even for single sites, by their differential responses during the different versions of associative cue-outcome conditioning. We chose to use these simple conditioning tasks to connect with classic evidence for RPE reinforcement learning profiles of neurons recorded electrophysiologically in nigral dopamine-containing neurons[1]. During single cue association sessions, large transient increases in dopamine release occurred in the centromedial sites in response to the cue predictive of reward, but these sites lacked increased release responses at outcome (reward). Centrolateral sites, by contrast, exhibited strong phasic increases in release both at cue and at positive outcome. The response to the cue decreased with learning, whereas RPE to cue increases with learning[49], and their response to reward delivery remained about the same across training, whereas RPE to reward delivery decreases[49]. Remarkably, when the mice proceeded from single cue association to cue discrimination training, prolonged plateau responses to the cues appeared in both centromedial sites and centrolateral sites. These plateaus extended throughout the cue period during cue discrimination, reversal discrimination, and probabilistic reward training, with slightly different forms suggesting that a two-component initial phasic increase carried into a plateau release of dopamine largely continuing

through the cue presentation period. Examples of the heterogeneity of these plateau-like responses are given in Figs. 3 and 5.

We found that these plateaus were most pronounced in mice with the most proficient performance and were nearly absent in the slow learners and non-learners. This learning-related remodeling of the responses was most evident at centromedial sites. One possibility to raise here is that the especially emphasized plateau in the centromedial sites is related to the lower concentration of dopamine transporter (uptake) than in the centrolateral sites, and so the centromedial sites have longer (slower) plateaus. The sources influencing these learning-related plateaus were not identified; but it is possible that in some way the presence of an alternative cue, even if not immediately visible at the time of a response and not indicative of reward availability, could induce a network state change leading to the tendency for an extended cue response. We cannot account for the mediolateral differences in the non-learners (Fig. 4e); one possibility is that this difference is related to motor learning. By contrast, it was only for outcome signals recorded at centrolateral sites during the probabilistic reward sessions that we could detect systematic changes related to reward probability as predicted by RL models. These observations, taken together, introduce dopamine plateau responses as learning-related features to add to transient and ramping responses formerly reported, and raise new questions about RPE encoding by the striatum during learning[39].

Absence of transfer of responses from outcome to outcome-predictive cue

In a notable deviation from classic RPE models, and from our own RPE model, we did not observe, as learning proceeded, a transfer of the dopamine responses from the time of outcome to the time of the outcome-predictive cues. Instead, in all versions of the task, cues were signaled by dopamine release in both centromedial sites and centrolateral sites throughout training, and outcome was accompanied throughout training at the centrolateral sites by dopamine release. Temporal transfer and the development of RPE signals have been questioned before (reviewed in refs. 37,39,40), but in recent work by Watabe-Uchida and colleagues[40,41], both RPE phenomena have been shown for the ventral striatum/nucleus accumbens and its ventral tegmental area afferent dopamine-containing neurons both at a population level and at the level of single cells. Our findings for the dorsal striatum suggest that further refinement and extension of reinforcement learning algorithms is needed to account for the spatiotemporal dynamics of dopamine release in the dorsal striatum, as here represented by recordings in centromedial sites and centrolateral sites, and also for the different release dynamics recorded during multiple phases of cue-association conditioning.

The transfer of dopamine signaling from outcome reinforcer to the most proximal predictor of that outcome is a central feature of RPE algorithms as applied to neural activity in the dopamine system[49]. How could the lack of such transfer in our mice be accounted for? One possibility is that we missed this outcome-to-cue transfer in our dopamine release recordings because these were population measurements with incomplete coverage of the dorsal striatum, hiding sub-populations conforming to the RPE predictions or sub-populations not in the range of our probes. This possibility is clearly high on the list of issues needing further testing, but it does nothing to account for the behavior of other aspects of the findings that do conform to RPE. An alternative possibility is that local circuits in the striatum affected by top-down signals from the thalamus and neocortex or elsewhere could modify the striatal firing of nigral dopamine-containing axons/terminals to block outcome signaling in the centromedial sites in our experiments. Intrastriatal modulation of dopamine release by cholinergic interneurons, proposed for many years based on striatal pharmacology[52–54], can change their activity during learning[4]. For example, some cholinergic inputs, some likely from these interneurons, generate action potentials in intrastriatal dopamine fibers far from their cell bodies[16]. Further, oscillatory local field potentials can accompany and even modulate activity[39,55–57]. We did not monitor this activity. Yet another possibly is suggested by the report by Hamid et al.[19] that dopamine release signals in the striatum occur in mediolateral and lateromedial waves, moving during Pavlovian conditioning from lateral to medial at rates of about 200 ms per transit. This activity was mainly recorded in the context of heavy damage to the overlying neocortex, which we tried to avoid here by using a single probe per mouse, but it could potentially comprise a scanning mechanism sensitive to such signals as we report here, imposed by yet unknown afferent or intrastriatal circuit elements. Further dynamics of intraneuronal networks in the striatum, such as the activation of dopaminergic fibers by acetylcholine released from cholinergic interneurons[20,58], surely must contribute. Across all these possibilities, at the population level, the patterns of both transient and plateau release of dopamine in the dorsal striatum, as measured here with two different sensor types, were quite distinct from, and difficult to align with, pure RPE signals.

It is theoretically possible that a sufficiently complex RL model could be tuned to show a plateau-like component of RPE at intermediate levels of training. Kim et al.[50] showed how temporal discounting can produce upward ramping RPE responses that resemble ramping dopamine responses that have been reported[10,32,43,44,59–61]. However, such upward ramps were rarely observed in our data. In RL models that endow the agent with a fine sense of the passage of time, such that each time point can be represented as a distinct state, it is also possible to find a small hump in the RPE signal in between the cue and reward delivery that becomes progressively earlier in every trial[1,40]. This hump only occurred in the middle of training, not at the start or end. If the right range of mid-training trials were analyzed, a combination of an upward ramp due to temporal discounting and a moving hump due to progressive transfer of the RPE signal to earlier states in trials could potentially add up to a roughly plateau-like response. We observed plateaus at the end of training, but because we did not keep training our mice after they reached criterion, it is impossible to say based on the present experiments whether the plateaus we observed would persist indefinitely with additional training. Such humps occur at the expense of the reward response, and we did not observe a diminution of the reward response when the plateau components arose. Another potential mechanism that can add a hump to the RPE signal is uncertainty as to exactly when the reward will be delivered[51,62]. In our task, the animal received a very clear reward delivery signal (i.e., the extinguishment of the rewarded cue), so this effect is unlikely to play a role in the present study. Also, a hump due to uncertainty will necessarily be close to the actual reward delivery time, unless the animal systematically overestimates the passage of time between cue and reward. Extensive additional modeling work will be required to determine whether sufficiently complex RL models actually can produce plateau-like RPE signals between cue and reward, and such models might have so many free parameters that they could be fitted to arbitrary data. We thus did not pursue these questions here. In rare occasions in our dataset, the dopamine responses exhibited RPE-like patterns, i.e., responses to the probabilistic reward and reward response early in cue reversal training recorded from centromedial dorsal striatum. The factors to shape dopamine response to be RPE-like are unknown, and possibly include cognitive demands or effects of overtraining. We await future studies to identify these factors.

Another phenomenon that is difficult to explain in terms of classic reinforcement learning theory is anticipatory licking. In any state in which reward is not available, licking produces a net loss, and so the model will learn to wait instead of to lick in those states. Thus, anticipatory licking seems to go outside the basic reinforcement learning framework. Additional innovative modeling work will be required to find an appropriate way to deal with it.

Contrasting dopamine release patterns in different sectors of the dorsal striatum

Each district of the dorsal striatum, as each area of the neocortex, likely uses and encodes different aspects of reinforcement along a nuanced scale from appetitive to aversive reinforcement options and action options. The striatal processing surely must involve much higher-dimensional algorithms than one just dealing with expected appetitive to aversive value, or only RPE; and different sectors of the dorsal striatum and corresponding corticostriatal circuits are surely engaged by different components of task execution[63]. Our recordings were limited to centromedial and centrolateral sites in the dorsal striatum, and thus did not fully span the striatum as now can be done with emerging methods[64]. Dopamine waves have been reported to travel in a lateral to medial direction when mice learn Pavlovian tasks rather than instrumental ones[19]. Also, lateral and medial parts of the dorsal striatum have been found to receive projections from different molecular-subtypes of dopaminergic neurons; the calbindin-positive type signals RPE, whereas the Anxa1-positive type encodes the acceleration of locomotion[65]. Thus, the differential dopamine dynamics observed in this study could be attributed to the directional dopamine wave or the differential contribution of dopaminergic cell subtypes to the ambient dopamine content of the different sub-regions. Delineating the functions of these neuromodulatory and neurochemical gradients awaits future study.

For dopamine release recorded in the nucleus accumbens, Jeong et al.[37], with a series of conditioning tasks, have found inconsistencies between dopamine release signals there and the predictions of RPE formulations. These favor what the authors term as a retrospective causal learning algorithm. The recordings by Jeong et al., like our recordings, were made with the aid of dopamine sensors, not with microelectrodes recording the spike activity of dopamine-containing cell bodies as in the original studies linking dopamine dynamics to RPE. Such discrepancies could be accounted for by findings of the Uchida and Watabe-Uchida groups (e.g., ref. 40; see also ref. 39).

Summary and caveats related to the findings

Here, we have shown discrepancies in both space and time between dopamine release patterns and patterns predicted by RPE formulations. These results corroborate the idea that the dorsal striatum is a composite of zones participating in multiple functional circuits. Cells involved in these circuits compute information in unique ways not necessarily equivalent to those of RPE formulations. Striatal microcircuitry is complex and spatially heterogeneous. It is possible that all or many regions of the striatum perform a similar core computation, but that single regions deal with different input-output and local circuit modulation according to requirements of given contexts and circumstances. Detailed study of the full range of variation in striatal dopamine response profiles could help to uncover the remarkable functional range of dopamine-based systems in modulating adaptive behavior.

We are aware of caveats that should accompany our conclusions. The tasks were variants of Pavlovian tasks and lacked the richness of much behavioral learning, decision-making and response variety. We used a fixed sequence of paradigms across animals as representative of the many switches that can occur in daily experience, but we are aware that the results could be constrained by this training sequence. We used both D1R-based (i.e., dLight1.3b and $GRAB_{DA3m}$) and D2R-based (i.e., $GRAB_{DA2m}$) dopamine sensors. Decay time constants for $GRAB_{DA2m}$ and $GRAB_{DA3m}$ are, respectively, 1.3 s[48] and ~600 ms[66], but that for dLight1.3b has not been determined. This imposes a lower temporal resolution on our data as compared to electrical recordings. We only sampled relatively restricted sites within more medial and lateral parts of caudoputamen, favoring centromedial and centrolateral sites, and we did not consider the compartmentalization of the striatum, in which striosome and matrix compartments have different relationships to dopamine-containing neuruons[33,35,67]. We used photometry, a recording method that measures the local sum of extracellular dopamine, whereas dopamine likely works both at individual synapses[68] and as an ambient non-synaptic modulator. Our findings cannot address the synaptic actions of dopamine because our measurements are probably dominated by extrasynaptic dopamine. RPE-observing dopamine signaling might instruct reinforcement plasticity only in a small subset of synapses that convey the relevant information, as suggested by reports of multiple, multiplexed responses of dopamine and dopamine neuron firing[10,43,44,59–61]. Despite these uncertainties, the surprises that emerged in our experiments open new opportunities to probe and to model mechanisms underlying striatum-based learning and its modulation by dopamine.

Methods

Animals

All experimental procedures were performed on 3–6 month-old wild-type mice and F1 hybrids on C57BL/6 J (Jackson Laboratory, strain ID #: 000664) and FVB (Taconic, model #FVB) background with the approval of the Committee on Animal Care at the Massachusetts Institute of Technology (MIT). F1 hybrids were produced from FVB mice in which *Pde6brd1* and *Disc1* were bred out ('corrected FVB'). Mice were group-housed separated by sex at 25 °C, 50% humidity with a 12:12 h light/dark cycle until the intracranial injection of sensors and optic fiber and head bar implantation. Subsequently, mice were single-housed in home-cage environment enriched by addition of eco-bedding, Nestlets and a PVC tube matching their body length as a play tunnel and then a body case during subsequent experimental sessions. Training sessions were conducted during the light cycle, 3–6 h after the daylight cycle switch. Mice were placed for at least 20 min in the experimental room after transport from the vivarium before testing.

Mouse preparation

Prior to daily training, mice ($N = 84$ prepared in total, 67 included, 17 mice later excluded due to 3 misplaced probes, 5 implant detachments, 2 mouse illnesses, and 7 unidentifiable probe locations) underwent stereotaxic surgery twice for virally mediated injection of a single dopamine sensor, either $GRAB_{DA2m}$ (AAV9-hSyn-DA2m, Addgene), $GRAB_{DA3m}$ (AAV9-hSyn-DA3m, WZ Biosciences), or dLight1.3b (AAV5-CAG-dLight1.3b, Addgene), followed by optic fiber and head-bar installation one week to 10 days later. Mice deeply anaesthetized with isoflurane (1-2% on oxygen flow rate of 0.4 L/min) were mounted on a stereotaxic apparatus and were injected subcutaneously with buprenorphine (2 mg/kg) and meloxicam (2 mg/kg) as pre-surgical analgesics, and for 3 days post-surgically as needed. The skin covering the skull was incised and a burr hole was made to place an injection needle to carry the viral construct to the target site in the right hemisphere (AP: +1.0 mm, ML: +1.7 mm, DV: −3.1 mm from bregma). A 0.5 μL aliquot of viral construct was administered at a rate of 0.05 μL/min. The injection needle was left in place for 5–10 min after completion of the injection and then was slowly removed from the brain. The burr hole was filled with bone wax and the overlying skin was sutured shut. One week to 10 days later, mice were anesthetized as before, mounted in the stereotaxic device, and the burr hole exposed and enlarged medially. The exposed skull was cleaned with cotton swabs dampened in 3% hydrogen peroxide, scarified using the tip of a surgical scalpel to enhance subsequent bonding of bone cement. The optic fiber was inserted to the target position (AP: +1.0 mm, ML: +1.5 mm, DV: −2.7 mm from bregma for centromedial sites; AP: +1.0 mm, ML: +1.9 mm, DV: −3.0 mm from bregma for centrolateral sites), the burr hole opening was filled with a small amount of petroleum jelly, and a thin layer of Metabond was applied to the exposed skull and to the bottom face of the optic fiber ferrule. It should be noted that there was considerable scatter in the final positions of the probe tips (see Figs. 1f and 6c, d). A 3D-printed head bar (3.2 mm (W)

x 3.2 mm (H) x 25.5 mm (L); weight 0.25 g) was positioned −2 mm posterior from the lambda and securely cemented onto the Metabond treated skull. Mice were allowed to recover for at least 3 weeks and left undisturbed except for regular animal husbandry care.

Thereafter, each mouse was mounted in the head-fixed recording chamber, and spontaneous dopamine signals were collected for an hour to check signal quality. Mice having an acceptable signal-to-noise ratio were placed on a water regulation schedule, with provision of 99% hydrogel (HydroGel®, ClearH_2O, USA) substituting for water intake. The daily amount of hydrogel was gradually decreased over a week from 2 g to the targeted amount. With this water restriction protocol, daily amounts were regulated to maintain body weight up to 85% of age- and sex-matched ad-lib group for the entire recording period. Based on weekly assessment of body condition by veterinarian staff, an additional amount of hydrogel was added if necessary.

All mice had a one-day per week break from daily sessions. On the day before a break day, mice were provided with double their daily amount, and on the break day, they resumed water restriction. In this manner, performance fluctuation often observed in ad-lib provision of water during a break could be minimized, and their overall health could be sustained.

Apparatus

To maximize efficiency and throughput, we constructed 9 identical training apparatuses, each equipped with a fiber-photometry recording setup. Each apparatus housed a mounting plate and posts (Thorlabs) attached with small devices (head bar holder, reward delivery tube, photobeam sensor, light emitting diode (LED) panel, mouse body case holder, etc.), supported by 3D-printed support frames. Each apparatus was shielded with 0.5" thick soundproof sheets to minimize noise distractions during training. To minimize any mechanical sound generated by the solenoid controlling reward delivery, we hung the device outside of each training apparatus to minimize the possibility that the mice use the activating sound as an additional cue for reward delivery. All electronic devices in each apparatus were controlled by Arduino and Raspberry Pi systems, which generated timestamps of training events and TTL pulses that synchronized with time events for behavioral analysis. Each mouse was mounted on the apparatus by screws attached to their surgically attached head bar, and the small PVC tube otherwise kept in their home-cage was placed so as to encase their body during the training sessions. Each recording apparatus was threaded with a patch cord that delivered excitation and received emission signals via an optic fiber with a 200 μm circular cross-section diameter, connecting individually to integrated fluorescence Mini-cubes (Doric Lenses). Each fiber-coupled LED (405 and 470 nm, Thorlabs) was activated by an LED driver (Thorlabs) according to triggering pulses generated by a dual-channel multi- function waveform generator (Owon). The square pulses (1.5 ms) from each channel triggered 405 nm (isosbestic control) and 470 nm (green fluorescent protein (GFP) signal) LED drivers continuously at a rate of 30 Hz with the two pulses shifted 120° relative to each other. The emission signals were detected and amplified with a fluorescence detector (Doric Lenses). The excitation power of the LED driver was adjusted individually to achieve a peak emission signal intensity of 0.3 V for each excitation. This intensity was measured using a photodetector that converts fluorescent intensity to voltage. The detected signals as well as LED driving pulses and Arduino trial start pulse were acquired with a sampling rate of 10 kHz with T7-pro DAQ and LJStreamM software (LabJack).

Training procedures

Sucrose solution habituation. On the first day of habituation, the reward spout delivering a 4% sucrose solution was placed close to the mouth of the head-fixed mouse, and drops of solution were provided frequently until the mouse drank actively from the spout. The spout position was then moved gradually away from its mouth, requiring the mouse to protrude its tongue to lick the sucrose. Reward retrieval of a drop of solution (4 μL) then occurred by licking. Each tongue protrusion was detected as a lick event by a photobeam sensor installed at the side of the spout.

Lick-activity-dependent and random reward habituation. After a mouse learned to retrieve the reward comfortably and actively, a lick-activity-dependent (LAD) reward habituation session began. During this habituation session, when a lick event was detected, a drop of solution was given, followed by various intervals from 6 to 8 s.

The 1 h LAD habituation (Operant) sessions continued daily until mice actively consumed more than 150 droplets per session. Most of the mice required 1-2 habituation sessions, but some mice required more. Once mice exhibited active licking for reward consumption, they underwent random reward habituation sessions in which they received unexpected drops of reward with varying intervals of 8–48 s, given for two or three sessions (Fig. 7a). Therefore, all mice underwent LAD reward habituation followed by random reward habituation before cue training sessions began. Random-reward probe sessions were also inserted every 9 or 10 sessions during training on subsequent tasks to determine whether dopamine responses to random rewards would change over the longitudinal training sessions.

Single-cue and reward learning. Following random reward habituation, all mice began daily training sessions on visual cue and reward association. Water regulated mice were placed in the head-fixed apparatus, and a blue LED (intensity setting at 3 lux) was placed to present visual cues on the right and left side at eye level (Fig. 1a), and with the reward delivery/lick detection device placed near the mouth. Each trial started with an LED lit (cue) on one side. The cue was lit for 1.5 s, the LED was turned off, and a reward was concurrently delivered. Each mouse had one of the two LEDs (right or left) designated as the cue predicting reward throughout training sessions. The locations of the reward cue were counterbalanced among subjects. One daily session typically consisted of ~175 trials with uniformly distributed random durations of 8–48 s. The daily session continued until a mouse showed stable performance defined by greater than 0.75 in area under the receiver operating characteristic (auROC; see below) comparing lick counts during pre-cue and cue period for at least 2 consecutive sessions.

Cue discrimination and reversal discrimination training. After completion of single-cue and reward conditioning, daily cue discrimination training began by inserting, into the original schedule of rewarded trials, trials with the opposite LED presented without reward (non-rewarded trials). Therefore, for each training session, two trial types (rewarded and non-rewarded trials) were intermixed pseudo-randomly (with no more than a 4x sequential repetition of one trial type). Each trial started with either the left- or right-side LED lit (cue), and a reward was delivered 1.5 s later at cue off only for the reward cue. Daily sessions, each typically consisting of around 350 trials with randomly varying trial duration of 8–12 s, continued until the mouse exhibited a stable discrimination level defined by greater than 0.75 in auROC value between cue presentations for at least 2 consecutive sessions. Daily training sessions on reversal learning then began. The cue and reward contingencies were reversed. Mice learned that the previously rewarded cue was no longer rewarded, but that the previously non-rewarded cue now predicted an upcoming reward. Daily sessions of reversal learning continued until a mouse exhibited a stable discrimination defined by greater than 0.75 in auROC value between cue presentations for at least 2 consecutive sessions.

Matched reward and cue rates. All tasks in this study were variants of the same basic discrimination task and were controlled by the same

task management code. Reward was given at the same temporal schedule across all tasks, which was determined entirely by the task software irrespective of mice's behavior. The only differences among tasks lay in the contingencies for delivering or withholding reward and illuminating the cue LEDs. As compared to the common basis, i.e., cue discrimination task, the single-cue task was implemented by simply disabling the non-rewarded cue LED; thus the average rate and time intervals between cue and reward deliveries were also the same. Similarly, the random reward task was implemented by disabling both cue LEDs, the probabilistic reward was implemented by disabling the non-rewarded cue LED and withholding reward on a certain fraction of trials where the nominally rewarded cue was illuminated, and the extinction task was implemented by illuminating the cue LEDs as in the discrimination task but withholding reward on all trials (see lower row of Fig. 1c).

Probabilistic reward learning. After completion of reversal learning, probabilistic reward training proceeded. The rewarded cue given during the prior reversal learning was the only cue presented. For the daily probabilistic reward sessions, reward was provided partially according to the target probability, and omitted reward trials were randomly selected for each session. Two sessions of reward probabilities of 0.75 and 0.5 were run, and for each reward probability, the session with the better recording quality was selected for analysis shown in Fig. 4a, b. A session with reward probability of 1.0 was always given before and after each probabilistic reward session. For some mice, two additional sessions having a block design of reward probabilities were performed, consisting of 4 blocks with different reward probabilities (0.75, 0.25, 0.5 and 1.0 reward probability with 44 trials per block) in a session.

Performance levels and calculation of auROC. To evaluate a performance level for each session, lick numbers for the duration of interest for all trials were used to compute auROC as a learning index. For cue discrimination, lick numbers during 1.5 s from cue onset were used and sorted according to trial types. For cue-evoked licks, lick numbers were documented for the period from 1.5 s before cue onset until cue onset (pre-cue period), and from cue onset until 1.5 s after cue onset (cue period) for all cue presentations. To examine lick behavior for the reward cue, lick numbers from 1.5 s before reward cue onset until reward cue onset (pre-reward cue period) and from 0 s to 1.5 s after reward cue onset (reward cue period) were separately tallied. The empirical auROC was calculated to represent a differential index of two datasets. The thresholds for constructing an auROC curve were taken for every middle point of all data sample differences of each given two data sets. Based on each threshold, true positive (*TP*; the number of incidents of one kind greater or equal to threshold; i.e., lick rate for rewarded trials or lick rate for cue period) and false negative (*FN*; the number of incidents of the same kind less than threshold) samples were computed to calculate *TPR* (true positive rate) as $TP / (TP + FN)$ for each threshold. Similarly, false positive (FP; the number of incidents of the other kind greater or equal to threshold; i.e., lick rate for non-rewarded trials or lick rate for pre-cue period) and true negative (*TN*; the number of incidents of the same kind less than threshold) were computed to calculate *FPR* (false positive rate) as $FP / (FP + TN)$ for each threshold. The auROCs for cue discrimination ($auROC_{disc}$) or for cue evoked lick ($auROC_{evk}$) were taken from the trapezoidal values of the ROC curves generated by FPR and TPR. The session auROC was mainly used to determine when to advance to the next learning schedule for each mouse. Each mouse was given up to 20 daily sessions, and if it failed to reach the learning criterion of at least 0.75 session auROC, the mouse was excluded from the following daily session schedules. Mice that reached the learning criterion were able to advance to the next phase of learning.

Possible effects of task order. Every mouse was trained on the same set of tasks in the same order, and so it is possible that some of the differences reported across tasks might depend on the history of the training rather than on intrinsic differences between the tasks and their corresponding evoked release signaling characteristics. The order of the tasks through reversal discrimination training was chosen partly to minimize the amount of time it took for the mouse to learn each task, and thus to maximize the variety of tasks that we were able to record before the signal quality started to degrade. Testing the effects of task order would have required additional sets of mice beyond the 67 successfully trained here to be trained for each permutation of the task order. We therefore did not attempt to disambiguate this potential confound.

Fiber-photometry recording

Each fiber-coupled LED (405 and 470 nm, Thorlabs) was activated by an LED driver (Thorlabs) according to triggering pulses generated by a dual-channel multi-function waveform generator (Owon). The square pulses (1.5 ms) from each channel triggered 405 nm (isosbestic control) and 470 nm (GFP signal) LED drivers continuously at a rate of 30 Hz, with the 470 nm pulse train lagging 120° behind the 405 nm control pulse train. The emission signals were detected and amplified with a fluorescence detector (Doric Lenses). The emission signals from the fluorescence detector, the TTL pulses to drive two LED drivers (Thorlab), and the trial start TTL from the Arduino were acquired with a sampling rate of 10 kHz with a T7-pro DAQ and LJStreamM software (LabJack). To acquire emission signals from six mice in six separate apparatuses at one time, data acquisition of 6 analog inputs of the emission signal and 8 channels of TTL inputs (405 LED driver, 470 LED driver, and trial start TTLs of 6 Arduinos) was arranged through a CB37 terminal board (LabJack). Before each recording session, the excitation intensity was set to produce emission photodetector output around 0.3 V. Because the emission signal from 470 nm excitation fluctuated due to active GFP, the minimal amplitude was set to around 0.3 V. The digital inputs were then separated offline based on the connecting ports. The TTL for triggering two LED drivers was used to extract emission signals during the excitation period, and TTLs from each Arduino were matched to the corresponding analog channels. The last sample of each 1.5 ms pulse was taken as a reading value for the entire excitation pulse, and the emission samples were separated according to the excitation wavelength. The time of trial start TTL served as an event timestamp for the corresponding analog channels.

The data acquired from each session were prepared with several preprocessing steps. The raw data of a session for GFP (470 nm) and control (405 nm) signals were extracted and separated by excitation pulses and band-passed with a low pass filter (5 Hz). The max values for each GFP pulse were converted to $(F_{GFP} - F_{CNT})/F_{CNT}$, where F_{CNT} denoted the max value obtained from the nearest control pulse (dF/F). Then, dF/F of low pass filtered data of each session data were z-score transformed. Z-scores were computed using the mean and standard deviation calculated for the whole data recorded during the entire session, including both inter-trial interval and task activity. Alternative methods for calculating z-scores are shown in Supplementary Figs. 6 and 7, and discussed in Supplementary Note 1. Each trial was then "baseline calibrated" by subtracting the mean dF/F during the baseline period (the last 1.5 s before cue onset) to make the calibrated mean dF/F during the baseline period zero. To verify that these procedures adequately suppressed motion artifacts, we performed a number of control experiments and a targeted analysis of the nominally isosbestic (405 nm) signal (see Supplementary Figs. 9 and 10, and Supplementary Notes 1 and 2).

Statistical test with generalized linear mixed model

For statistical inferences, we used a generalized linear mixed model (GLMM) in R package (glmmTMB). Our main interest was to confirm

the effects of recording region, learning type, and performance achieved on the dopamine release responses to cue and reward that were recorded. To quantify and build a model, we used AUC (trapezoidal method) of the dopamine trace (z transformed *dF/F* data) for the rewarded trials. The trial trace was prepared by calibrating with its own baseline (subtracting mean of 1.5 s pre-cue value from each data point), and trial AUC of early cue (0.5 s period from cue onset), late cue (0.5 s period to cue offset), and reward (1 s period from the first lick after reward), and these calibrated traces were used as response variables. For testing the effects of sex on dopamine response, we tested the rewarded trial data from cue conditioning and cue discrimination (with regressors learning kind × sex × region on early and late cue and reward). For estimating the effect of performance on the dopamine responses, we examined the rewarded trial data from the last session of cue conditioning, cue discrimination, and reversal discrimination (learning kind × performance level × region).

Discrimination performance level

The performance level was determined by the auROC value at the last session and its group median for discrimination between rewarded and non-rewarded trials based on lick counts during the cue presentation period. A mouse having a higher auROC value than group median in the last session was assigned as a “better” (higher) performer, otherwise the mouse was assigned as a "worse" (lower) performer. The interaction model was chosen over an additive model based on the ANOVA likelihood ratio test of two models. The post-estimation process of GLMM results then was performed with the emmeans package in R to check the effect of learning kind over other levels (learning, region, sex, or performance).

The initiation of spontaneous licks was detected if a lick occurred during the inter-trial interval (period starting 6.5 s after cue onset and ending with the next cue onset) and was preceded by a period without licks of at least 0.5 s. Lick-aligned fluorescence trace data were prepared based on each detected lick (0.5 s pre-lick and 1 s post-lick data), and each lick trace was calibrated using the pre-lick data as baseline. The AUC of lick traces was obtained for a 0.5 s period after each lick. These lick AUCs were used as response variables on learning kinds, learning and region in GLMM model and post-estimation.

Histology

Mice were perfused with a 0.9% saline solution followed by 4% paraformaldehyde (PFA) in 0.1 M phosphate buffer (PB). Brains were harvested, post-fixed in 4% PFA overnight in 4 °C, cryoprotected in 25% glycerol in 0.1 M PB at least one night, and then stored at 4 °C until sectioning. The brains were frozen in powdered dry ice, and 30 µm thick coronal sections were cut using a sliding microtome. Sections were stored at 4 °C in 0.02% sodium azide in 0.1 M PB until use.

Striatal sections containing the implant were selected for immunofluorescence staining. The sections were rinsed three times for 5 min in 0.01 M phosphate-buffered saline (PBS) with 0.2% Triton X-100 (Tx), blocked with tyramide signal amplification (TSA) blocking solution for 1 h at room temperature, and then incubated with primary antibody solution containing chicken anti-GFP antibody (Abcam, ab13970, 1:2000) and rabbit anti-mu opioid receptor antibody (Abcam, ab134054, 1:500) in TSA blocking solution for two nights at 4 °C. After primary antibody incubation, the sections were rinsed three times for 5 min in PBS-Tx, then incubated for 2 h in a secondary antibody solution containing goat anti-chicken Alexa Fluor (AF) 488 (ThermoFisher, A-11039, 1:300) and goat anti-rabbit AF647 (ThermoFisher, A-21245, 1:300) in TSA blocking solution.

The sections were rinsed for 2 min in 0.1 M PB, incubated for 2 min in DAPI (Life Technologies, D1306, 1:1000) solution in PBS, rinsed 3 times for 2 min in 0.1 M PB, and then mounted onto glass slides and coverslipped with ProLong Gold Antifade Reagent (Life Technologies, P36930). Images were taken by AxioZoom V16 (Ziess).

Histological localization of recording sites

The dorsal striatum contains many semi-independent histochemical and physiological gradients in all three cardinal dimensions. The boundaries of striosomes can be sharply delineated with some histochemical stains, but there are no known similarly clear boundaries subdividing the dorsal striatum at maturity into districts at a larger scale. In placing our probes, we avoided the medial and lateral extremes of the caudoputamen and thus populated roughly the central half of the volume of the striatum with probe tips. We defined a standardized coordinate system within the striatum to accommodate its slightly irregular shape, as follows.

We defined four reference points in each coronal section: the most medial point of striatum (*x1*, *y1*; usually about 1 mm from midline), the most lateral point of striatum (*x2*, *y2*), the most dorsal point of striatum (*x3*, *y3*), and the most dorsal point of anterior commissure (*x4*, *y4*). Designating the tip of probe as (*x5*, *y5*), the standardized medial-lateral coordinate of the probe tip was calculated as $(x5 - x1)/(x2 - x1)$, and the standardized depth coordinate of the probe tip was calculated as $(y5 - y3)/(y4 - y3)$. We refer to this coordinate system as relative position or standardized coordinate in the coronal plane (shown in Fig. 6c, d and Supplementary Fig. 8). A-P coordinates (specified in mm) were determined by comparing histological sections to the atlas[69] and are given relative to bregma as in the atlas.

A slightly different coordinate system was used by a second person to classify probes as centromedial or centrolateral. Recording sites were classified as centromedial if the distance from the midline to the tip of the probe was less than 0.6 (60%) of the mediolateral distance from the midline to the lateral edge of the striatum in the coronal section containing the site. In Fig. 1b, f, M-L coordinates are calculated in this way. The D-V coordinates shown in these figures are given by the distance (mm) from dorsal surface of the striatum as it appears in the same section containing the probe track. These measurements differ from the standardized coordinate system described above. Depending on the A-P plane of section, the division between centromedial and centrolateral corresponded to around 0.4 to 0.5 M-L in the standardized coordinate system.

Anatomical maps of first principal component

Principal components analysis calculations were performed by the Matlab ‘pca’ function, using each mouse’s waveform of dF/F, averaged over all rewarded trials in that mouse’s final (“acquisition”) session of discrimination training, as input. Spatial smoothing of the PC1 amplitude values across mice was done in each cardinal anatomical plane as follows. First a coordinate grid of 50 bins was constructed to span the relative position values in each direction, resulting in bins whose widths depended on the 3D axes (0.012 wide in the M-L direction, 0.016 in the D-V direction, 0.012 in the A-P direction). For each mouse, the value of the amplitude (“score”) for PC1 was assigned to the bin containing the recording site’s coordinates and was copied throughout a square of 17 × 17 neighboring bins extending eight bins to each side of the bin containing the recording site, truncated if necessary to stay within the bounds of the 50 × 50 coordinate grid. All other bins were assigned the value NaN and thus excluded from statistical calculations. Each mouse produced a single set of 50 × 50 bins. Two computations were then performed across the 67 sets of bins corresponding to the 67 mice: the total number of non-NaN values was counted for each bin position, and the NaN-tolerant average value was computed for each bin position (i.e., the mean obtained strictly from the mice that had non-NaN-valued bins at a given position, or NaN if there were no non-NaN values). The average values of any bins for which the total number of non-NaN values in the bin was less than 3 were then reset to NaN, resulting in a 50 × 50 matrix in which every bin

contained either NaN or the average of values from at least 3 mice. The 50 × 50 matrix was then plotted as a pseudo-color image, with the color scale chosen so that its limit in the negative direction, representing NaN, was well below the most negative value in any bin.

To assess statistical significance of the spatially smoothed average values, we performed bootstraps on the set of mice included in the entire calculation. Two-hundred bootstraps were performed by randomly selecting 63 mice at a time, with replacement (i.e., a given mouse could be repeated) from the actual set of 63 mice recorded. The median value across bootstraps was used to create the final pseudo-color plots, and the median of those median values was used as a reference value for statistical significance. If the 2.5th percentile of values across bootstraps was greater than the reference value, that bin was marked as significantly high. If the 97.5th percentile of values was less than the reference value, the bin was marked as significantly low.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The source data for the figures are provided with the paper as a Source Data File. Examples of raw data are provided as Supplementary Data files (see Description of Additional Supplementary Files for contents of each Supplementary Data file). Source data are provided with this paper.

Code availability

The codes generated for this study have been deposited to Code Ocean. Capsule slugs: 8994384, 0659284, 8953637, 2418919, 1192434, 8202933, 0294351, 8081510, 9452945.

References

1. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
2. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction*. Second edn, (MIT Press, 2018).
3. Romo, R. & Schultz, W. Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.* **63**, 592–606 (1990).
4. Joshua, M., Adler, A., Mitelman, R., Vaadia, E. & Bergman, H. Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. *J. Neurosci.* **28**, 11673–11684 (2008).
5. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
6. Puryear, C. B., Kim, M. J. & Mizumori, S. J. Conjunctive encoding of movement and reward by ventral tegmental area neurons in the freely navigating rodent. *Behav. Neurosci.* **124**, 234–247 (2010).
7. Eshel, N. et al. Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* **525**, 243–246 (2015).
8. Robinson, S., Sandstrom, S. M., Denenberg, V. H. & Palmiter, R. D. Distinguishing whether dopamine regulates liking, wanting, and/or learning about rewards. *Behav. Neurosci.* **119**, 5–15 (2005).
9. Starkweather, C. K., Babayan, B. M., Uchida, N. & Gershman, S. J. Dopamine reward prediction errors reflect hidden-state inference across time. *Nat. Neurosci.* **20**, 581–589 (2017).
10. Berke, J. D. What does dopamine mean? *Nat. Neurosci.* **21**, 787–793 (2018).
11. Lerner, T. N. et al. Intact-brain analyses reveal distinct information carried by SNc dopamine subcircuits. *Cell* **162**, 635–647 (2015).
12. Howe, M. W. & Dombeck, D. A. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* **535**, 505–510 (2016).
13. Parker, N. F. et al. Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* **19**, 845–854 (2016).
14. Patriarchi, T. et al. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science* **360**, eaat4422 (2018).
15. Sun, F. et al. A genetically encoded fluorescent sensor enables rapid and specific detection of dopamine in flies, fish, and mice. *Cell* **174**, 481–496 e419 (2018).
16. Liu, C. et al. An action potential initiation mechanism in distal axons for the control of dopamine release. *Science* **375**, 1378–1385 (2022).
17. Threlfell, S. et al. Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron* **75**, 58–64 (2012).
18. Brimblecombe, K. R. & Cragg, S. J. The striosome and matrix compartments of the striatum: a path through the labyrinth from neurochemistry toward function. *ACS Chem. Neurosci.* **8**, 235–242 (2017).
19. Hamid, A. A., Frank, M. J. & Moore, C. I. Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. *Cell* **184**, 2733–2749 e2716 (2021).
20. Krok, A. C. et al. Intrinsic dopamine and acetylcholine dynamics in the striatum of mice. *Nature* **621**, 543–549 (2023).
21. Cox, J. & Witten, I. B. Striatal circuits for reward learning and decision-making. *Nat. Rev. Neurosci.* **20**, 482–494 (2019).
22. Saunders, B. T., Richard, J. M., Margolis, E. B. & Janak, P. H. Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nat. Neurosci.* **21**, 1072–1083 (2018).
23. Tsutsui-Kimura, I. et al. Distinct temporal difference error signals in dopamine axons in three regions of the striatum in a decision-making task. *Elife* **9**, e62390 (2020).
24. Hikosaka, O., Kim, H. F., Yasuda, M. & Yamamoto, S. Basal ganglia circuits for reward value-guided behavior. *Annu Rev. Neurosci.* **37**, 289–306 (2014).
25. Choi, K. et al. Distributed processing for action control by prelimbic circuits targeting anterior-posterior dorsal striatal subregions. *bioRxiv* https://doi.org/10.1101/2021.12.01.469698 (2021).
26. Choi, K., Holly, E. N., Davatolhagh, M. F., Beier, K. T. & Fuccillo, M. V. Integrated anatomical and physiological mapping of striatal afferent projections. *Eur. J. Neurosci.* **49**, 623–636 (2019).
27. Matsumoto, M. & Hikosaka, O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* **459**, 837–841 (2009).
28. Bromberg-Martin, E. S., Matsumoto, M. & Hikosaka, O. Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron* **68**, 815–834 (2010).
29. Brischoux, F., Chakraborty, S., Brierley, D. I. & Ungless, M. A. Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proc. Natl. Acad. Sci. USA* **106**, 4894–4899 (2009).
30. Markowitz, J. E. et al. Spontaneous behaviour is structured by reinforcement without explicit reward. *Nature* **614**, 108–117 (2023).
31. Dai, B. et al. Responses and functions of dopamine in nucleus accumbens core during social behaviors. *Cell Rep.* **40**, 111246 (2022).
32. Howe, M. W., Tierney, P. L., Sandberg, S. G., Phillips, P. E. & Graybiel, A. M. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* **500**, 575–579 (2013).
33. Prager, E. M. et al. Dopamine oppositely modulates state transitions in striosome and matrix direct pathway striatal spiny neurons. *Neuron* **108**, 1091–1102 e1095 (2020).

34. Nadel, J. A. et al. Optogenetic stimulation of striatal patches modifies habit formation and inhibits dopamine release. *Sci. Rep.* **11**, 19847 (2021).
35. Sgobio, C. et al. Aldehyde dehydrogenase 1-positive nigrostriatal dopaminergic fibers exhibit distinct projection pattern and dopamine release dynamics at mouse dorsal striatum. *Sci. Rep.* **7**, 5283 (2017).
36. Graybiel, A. M. & Matsushima, A. The ups and downs of the striatum: Dopamine biases upstate balance of striosomes and matrix. *Neuron* **108**, 1013–1015 (2020).
37. Jeong, H. et al. Mesolimbic dopamine release conveys causal associations. *Science* **378**, eabq6740 (2022).
38. Coddington, L. T., Lindo, S. E. & Dudman, J. T. Mesolimbic dopamine adapts the rate of learning from action. *Nature* **614**, 294–302 (2023).
39. Cone, I., Clopath, C. & Shouval, H. Z. Learning to express reward prediction error-like dopaminergic activity requires plastic representations of time. *Res. Sq* rs.3.rs-3289985 (2023).
40. Amo, R. et al. A gradual temporal shift of dopamine responses mirrors the progression of temporal difference error in machine learning. *Nat. Neurosci.* **25**, 1082–1092 (2022).
41. Akiti, K. et al. Striatal dopamine explains novelty-induced behavioral dynamics and individual variability in threat prediction. *Neuron* **110**, 3789–3804 e3789 (2022).
42. Takahashi, Y. K. et al. Dopaminergic prediction errors in the ventral tegmental area reflect a multithreaded predictive model. *Nat. Neurosci.* **26**, 830–839 (2023).
43. Hamid, A. A. et al. Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016).
44. Mohebi, A. et al. Dissociable dopamine dynamics for learning and motivation. *Nature* **570**, 65–70 (2019).
45. Lee, R. S., Sagiv, Y., Engelhard, B., Witten, I. B. & Daw, N. D. A feature-specific prediction error model explains dopaminergic heterogeneity. *Nat. Neurosci.* **27**, 1574–1586 (2024). Online ahead of print.
46. Berridge, K. C. & Robinson, T. E. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.* **28**, 309–369 (1998).
47. Lee, R. S., Mattar, M. G., Parker, N. F., Witten, I. B. & Daw, N. D. Reward prediction error does not explain movement selectivity in DMS-projecting dopamine neurons. *Elife* **8**, e42992 (2019).
48. Sun, F. et al. Next-generation GRAB sensors for monitoring dopaminergic activity in vivo. *Nat. Methods* **17**, 1156–1166 (2020).
49. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
50. Kim, H. R. et al. A Unified Framework for Dopamine Signals across Timescales. *Cell* **183**, 1600–1616 e1625 (2020).
51. Mikhael, J. G., Kim, H. R., Uchida, N. & Gershman, S. J. The role of state uncertainty in the dynamics of dopamine. *Curr. Biol.* **32**, 1077–1087 e1079 (2022).
52. Glowinski, J., Cheramy, A., Romo, R. & Barbeito, L. Presynaptic regulation of dopaminergic transmission in the striatum. *Cell Mol. Neurobiol.* **8**, 7–17 (1988).
53. Cragg, S. J. & Greenfield, S. A. Differential autoreceptor control of somatodendritic and axon terminal dopamine release in substantia nigra, ventral tegmental area, and striatum. *J. Neurosci.* **17**, 5738–5746 (1997).
54. Nelson, A. B. et al. Striatal cholinergic interneurons drive GABA release from dopamine terminals. *Neuron* **82**, 63–70 (2014).
55. Beatty, J. A., Song, S. C. & Wilson, C. J. Cell-type-specific resonances shape the responses of striatal neurons to synaptic input. *J. Neurophysiol.* **113**, 688–700 (2015).
56. Thorn, C. A. & Graybiel, A. M. Differential entrainment and learning-related dynamics of spike and local field potential activity in the sensorimotor and associative striatum. *J. Neurosci.* **34**, 2845–2859 (2014).
57. Wilson, C. J. Predicting the response of striatal spiny neurons to sinusoidal input. *J. Neurophysiol.* **118**, 855–873 (2017).
58. Chantranupong, L. et al. Dopamine and glutamate regulate striatal acetylcholine in decision-making. *Nature* **621**, 577–585 (2023).
59. Phillips, P. E., Stuber, G. D., Heien, M. L., Wightman, R. M. & Carelli, R. M. Subsecond dopamine release promotes cocaine seeking. *Nature* **422**, 614–618 (2003).
60. Roitman, M. F., Stuber, G. D., Phillips, P. E., Wightman, R. M. & Carelli, R. M. Dopamine operates as a subsecond modulator of food seeking. *J. Neurosci.* **24**, 1265–1271 (2004).
61. Engelhard, B. et al. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature* **570**, 509–513 (2019).
62. Gershman, S. J. & Uchida, N. Believing in dopamine. *Nat. Rev. Neurosci.* **20**, 703–714 (2019).
63. Graybiel, A. M. & Matsushima, A. Striosomes and Matrisomes: Scaffolds for Dynamic Coupling of Volition and Action. *Annu Rev. Neurosci.* **46**, 359–380 (2023).
64. Vu, M. T. et al. in *International Basal Ganglia Society Meeting*.
65. Azcorra, M. et al. Unique functional responses differentially map onto genetic subtypes of dopamine neurons. *Nat. Neurosci.* **26**, 1762–1774 (2023).
66. Zhou, Y. et al. Improved green and red GRAB sensors for monitoring dopaminergic activity in vivo. *Nat. Methods* **21**, 680–691 (2023).
67. Salinas, A. G., Davis, M. I., Lovinger, D. M. & Mateo, Y. Dopamine dynamics and cocaine sensitivity differ between striosome and matrix compartments of the striatum. *Neuropharmacology* **108**, 275–283 (2016).
68. Yagishita, S. et al. A critical time window for dopamine actions on the structural plasticity of dendritic spines. *Science* **345**, 1616–1620 (2014).
69. Franklin, K. B. J. & Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates, Compact*. 3rd edn, (Elsevier, 2008).

Acknowledgements

We thank Dr. Steven Worthington for consulting and helping on statistical approach at the Institute for Quantitative Social Science, Harvard University, Dr. Yulong Li for early contribution of $GRAB_{DA}$-encoding viruses, Henry F. Hall for constructing recording apparatus, Dr. Sebastien Delcasso for designing the apparatus system, Dr. Yasuo Kubota for help with manuscript preparation, and Johnny Loftus for help with figure preparation. This work was funded by the National Institutes of Health (R01 MH060379 to A.M.G.), the William N. & Bernice E. Bumpus Foundation (RRDA Pilot: 2013.1 to A.M.G.; Postdoctoral Fellowships to M.J.K. and A.M.), the Saks Kavanaugh Foundation (to A.M.G.), the CHDI Foundation (A-5552 to A.M.G.), and Dr. Lisa Yang (to A.M.G.).

Author contributions

A.M.G. and M.J.K. designed and initiated the research; M.J.K. and D.H. performed the surgeries; M.J.K., C.S., P.S. and K.T. performed the recordings; D.H., M.J.K. and C.S. performed perfusion; A.M. and T.Y. performed the histology and imaging; D.H., C.S., P.S. and K.T. provided animal handling and care; M.J.K., D.J.G. and A.M.G. analyzed data and T.Y. and A.M.G. analyzed histological material; E.H. supported set-up of recording and behavioral chambers; A.M. advised on photometry aligned to licking and final editing; L.T. for early contribution of dLight-encoding virus; A.M.G., D.J.G., and M.J.K. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53176-7.

Correspondence and requests for materials should be addressed to Ann M. Graybiel.

Peer review information *Nature Communications* thanks Mitsuko Watabe-Uchida and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.