HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2020 April 10.

Published in final edited form as:

Nature. 2019 June ; 570(7762): 509–513. doi:10.1038/s41586-019-1261-9.

Specialized coding of sensory, motor, and cognitive variables in VTA dopamine neurons

Ben Engelhard[1], **Joel Finkelstein**[1,2], **Julia Cox**[1], **Weston Fleming**[1], **Hee Jae Jang**[1], **Sharon Ornelas**[1], **Sue Ann Koay**[1], **Stephan Y. Thiberge**[1], **Nathaniel Daw**[1,2], **David W. Tank**[1,3], **Ilana B. Witten**[1,2,*]

[1]Princeton Neuroscience Institute, Princeton University, Princeton NJ 08544

[2]Department of Psychology, Princeton University, Princeton NJ 08544

[3]Department of Molecular Biology, Princeton University, Princeton NJ 08544

Abstract

There is increased appreciation that dopamine (DA) neurons in the midbrain respond not only to reward[1] and reward-predicting cues[1,2], but also to other variables such as distance to reward[3], movements[4–9], and behavioral choices[10,11]. Based on these findings, a major open question is how the responses to these diverse variables are organized across the population of DA neurons. In other words, do individual DA neurons multiplex multiple variables, or are subsets of neurons specialized in encoding specific behavioral variables? The reason that this fundamental question has been difficult to resolve is that recordings from large populations of individual DA neurons have not been performed in a behavioral task with sufficient complexity to examine these diverse variables simultaneously. To address this gap, we used 2-photon calcium imaging through an implanted lens to record activity of >300 midbrain DA neurons in the ventral tegmental area (VTA) during a complex decision-making task. As mice navigated in a virtual reality (VR) environment, DA neurons encoded an array of sensory, motor, and cognitive variables. These responses were functionally clustered, such that subpopulations of neurons transmitted information about a subset of behavioral variables, in addition to encoding reward. These functional clusters were spatially organized, such that neighboring neurons were more likely to be part of the same cluster. Taken together with the topography between DA neurons and their projections, this specialization and anatomical organization may aid downstream circuits in correctly interpreting the wide range of signals transmitted by DA neurons.

*Correspondence: iwitten@princeton.edu.

Contributions

B.E., D.W.T. and I.B.W. conceived the project. B.E., J.F., J.C., W.F., S.O. collected data. B.E., W.F., J.C., S.O., and H.J. analyzed data. S.Y.T. and S.A.K. provided technical training. N.D., D.W.T. and I.B.W. advised on the data analysis. B.E. and I.B.W. wrote the paper.

Competing interests

The authors declare no competing interests.

Code and data availability statements

The code for the encoding model and the motion correction are available on github (https://github.com/benengx). All other code and data are available upon reasonable request.

To determine how responses are organized across the population of VTA DA neurons, we sought to record at cellular resolution from ensembles of identified DA neurons in a behavioral task with sufficient complexity to engage many of the behavioral variables that are now thought to be of relevance to DA neurons. These variables include reward[1,12], reward-predicting cues[1,2], reward history[11,13], spatial position[3], kinematics (velocity, acceleration, view angle)[4–7], and behavioral choices[10,11,14].

Towards this end, we trained 20 mice on a decision-making task in a VR environment that encompassed this wide range of behavioral variables ("Accumulating Towers" task[15]; Fig. 1a,b; visual snapshots of maze in Extended Data Fig. 1a; Supplementary Video 1). As mice navigated the central stem of the virtual T-maze, they observed transient reward-predicting cues on the left and right of the maze stem that signaled which maze arm was most likely to be rewarded ("cue period"; Fig. 1b; cues consisted of white towers, see Methods). By turning to the side with more cues, the mice received a water reward, while turning to the other side resulted in a tone and a 3s time out. The 2s period after reward delivery or tone presentation was termed the "outcome period" (Fig. 1b). As expected, after training, mice tended to turn to the maze arm associated with more cues (Fig. 1c; average percent correct is 77.6±0.9%).

To perform 2-photon activity imaging from ensembles of DA neurons during this task, we implanted a gradient index (GRIN) lens above the VTA[16]. GCaMP expression was achieved either by injecting a Cre-dependent GCaMP virus in the VTA of DAT::Cre mice, or by crossing a GCaMP reporter line with DAT::Cre mice (Fig. 1d; Supplementary Video 2 for sample imaging video; also see Extended Data Fig. 2 for relationship between spikes and fluorescence in DA neurons). In either case, an mCherry virus was injected into the VTA to facilitate motion correction (Extended Data Fig. 3, see Methods). Using this approach, we recorded activity of ~10–30 DA neurons simultaneously in each of 20 mice during performance of the VR task (Fig. 1e,f; n=303 DA neurons from 20 mice; 292 neurons were estimated to be in the VTA and 11 in the SNc, see Extended Data Fig. 4b for reconstructed locations).

Responses of 284 out of 303 DA neurons were significantly modulated by one or more of the following variables (Fig. 2a): spatial position (n=91, 30%), kinematics (n=137, 45%), reward-predicting cues (n=77, 25%), choice accuracy (whether the trial resulted in reward; n=69, 23%), reward history (whether the previous trial was rewarded; n=95, 31%), and reward (n=232, 77%; significance was assessed based on nested comparisons of the encoding model described below, see Methods). The first five variables were quantified during the cue period, and the final variable (reward) was quantified during the outcome period.

During the cue period, individual neurons exhibited diverse responses to most of these variables (Fig. 2a). For example, neurons that were modulated by spatial position most often exhibited upward ramps, although some displayed downward ramps, consistent with ramps previously identified with fast-scan cyclic voltammetry in the striatum[3,17,18] (example single trials in Extended Data Fig. 1b). Neurons that were selective to kinematics were tuned to a range of velocities, acceleration or view angles. Neurons that responded to reward-predicting

cues often, but not always, displayed stronger responses to contralateral versus ipsilateral cues[19]. Neurons that were modulated by accuracy universally displayed higher activity to error (as opposed to correct) trials, while neurons that were modulated by previous trial outcome were modulated in either direction.

In contrast to the diverse responses to many of the variables during the cue period (e.g. upward versus downward spatial ramps), most neurons responded consistently during the outcome period, with stronger responses to reward than lack of reward (Fig. 2a).

Thus, for the first time, we have access to many of the behavioral variables that are thought to be relevant to DA neurons within a single behavioral paradigm. This puts us in a position to achieve our goal of understanding how the responses to these variables are organized across the DA population. To do this, we need a method to accurately quantify how much of the variance of the neural responses can be attributed to each behavioral variable individually, despite the presence of multiple behavioral variables.

Towards this end, we quantitatively predicted the GCaMP signal based on the measured behavioral variables with an encoding model (Fig. 2b; see Methods). To derive the predictors for the model, each variable was considered either as a discrete "event" variable, a "whole-trial" variable, or a "continuous" variable. In the case of "event" variables (left cues, right cues, reward), predictors were generated by convolving the event's time series with a spline basis set, in order to allow flexibility in the temporal influence of cues on GCaMP. In the case of "whole-trial" variables (previous reward, accuracy), the value of the binary predictor throughout the trial indicated reward on the previous (or current) trial. In the case of "continuous" variables (position, kinematics [velocity / acceleration / view angle]), predictors included the variables raised to the first, second and third power, in order to enable flexibility in the relationship between the variable and GCaMP. This model was chosen to include behavioral variables that significantly improved predictions of neural activity, after comparing several models (model comparisons in Extended Data Fig. 1d,e).

Using this encoding model, we quantified the relative contribution of each behavioral variable to the response of each neuron by determining how much the explained variance declined when that variable was removed from the model (see Methods; relative contributions for example neurons in Extended Data Fig. 5). Averaged across the population, the highest relative contribution during the cue period was attributed to kinematics (32.4±1.9% of the total variance explained during the cue period), followed in descending order by spatial position (22±1.7%), previous reward (17.7±1.5%), cues (14.6±1.4%), and accuracy (13.5±1.5%; Fig. 2c,d). During the outcome period, reward contributed strongly to the response (74.7±1.8%), consistent with the large number of neurons that responded to reward (Fig 2a).

How is the relative contribution of these behavioral variables to neural responses distributed across the population? During the cue period, most behavioral variables had a small contribution to the response of each neuron, while a small subset had a large contribution. In contrast, during the outcome period, reward contributed to a large fraction of the response of most neurons (Fig. 2d). This raises the possibility that during the cue period, subsets of DA

neurons are specialized to encode specific behavioral variables, while during the outcome period, most DA neurons encode reward.

To more systematically examine this idea, we performed clustering of the neurons based on the relative contributions of each behavioral variable to each neuron, using a Gaussian Mixture Model (GMM; Fig. 3a; see Methods). We found that 5 clusters of neurons gave the best (lowest) Bayesian Information Criterion (BIC) score for this data (Fig. 3a; see Methods for details on BIC score calculation). These 5 clusters explained the data better than expected by chance ($p<0.0001$, comparing the likelihood of the data given the clustering model to that of shuffled data, for null distributions generated by shuffling across behavioral variables, as well as by shuffling across neurons; Extended Data Fig. 6a). Thus, we can conclude that VTA DA neurons display a statistically significant degree of functional clustering.

Each cluster was composed of DA neurons that responded most strongly to a specific behavioral variable during the cue period. Note that this specialization does not mean that DA neurons only encoded a single variable during the cue period; in fact, many neurons also significantly encoded a 2nd variable, but not as strongly (Fig. 3b). In contrast to the specialization during the cue period, all clusters were composed of neurons that had reward responses (Fig. 3a; Extended Data Fig. 7a). Thus, this clustering analysis provided further evidence that VTA DA neurons are specialized during the cue period, while they share a response to reward during the outcome period. Consistent with the idea that cue period activity differed across clusters, neural activity predicted choice and accuracy to different extents in different clusters (Extended Data Fig. 6b). Supporting the robustness of these clusters, similar cluster assignment was obtained when the procedure was implemented independently on random halves of the trials of each neuron, or with clustering based on a different clustering procedure[20] (Extended Data Fig. 7; see Methods).

We next sought to determine if the functional clusters of DA neurons were anatomically organized within the VTA. The location of each neuron was estimated based on combining histological reconstruction of the lens tract with the position of the neuron within the imaging field[21] (Extended Data Fig. 4a,b; see Methods). We observed significant dependence of cluster identity on A/P location for 3 of the 5 clusters, and on M/L location for 4 of the 5 clusters (Fig. 3c,d; $p<0.01$, comparing STD of the relative concentration of neurons within a cluster to a shuffled distribution obtained by randomly permuting the A/P or M/L location of all neurons relative to cluster identity, Holm-Bonferroni correction; see Methods). Specifically, neurons belonging to the cluster associated with kinematics were located more laterally and posteriorly (cluster 1), those associated with accuracy were located more medially and anteriorly (cluster 5), and neurons associated with previous reward were located more laterally (cluster 3).

Directly correlating the A/P and M/L location of the neurons with the relative contributions of each behavioral variable led to similar findings (Extended Data Fig. 4c,d). To ascertain that this anatomical organization cannot be explained by differences between individual mice rather than by a true dependence on location, we considered a multinomial mixed effect regression using the cluster identity of the neurons as the dependent variable, the A/P

and M/L locations as fixed effects, and mouse identity as a random effect for the intercepts. This confirmed that anatomical location significantly predicted cluster identity ($p<0.002$, Wald test on the set of null hypotheses that all A/P coefficients in the model are equal to each other and all M/L coefficients are equal to each other; n=190, $\chi^2=20.82$, deg. freedom=6).

A complementary approach to examine spatial organization in our data is to examine the spatial organization of pairwise correlations between neurons. This allows us to separately consider the spatial organization of the "signal" correlation (i.e. correlations that can be explained by responses to behavioral variables; conceptually related to functional clustering in Fig. 3), and also of the "noise" correlation (i.e. neural correlations that cannot be explained by the behavioral variables). DA neurons are thought to have high noise correlations[22–24], but the spatial organization of these correlations has not been described.

To first confirm that DA neurons in our experiment indeed have high noise correlations, we added an additional predictor to the encoding model from Fig. 2b: a "network" predictor that reflects the activity of other simultaneously imaged neurons (for each neuron, the new predictor was the 1st PCA of the ΔF/F from all other simultaneously recorded neurons; Fig. 4a). Consistent with DA neurons having high noise correlations, the performance of this new model explained a substantially higher variance of neural activity (R^2 from behavioral + "network" model: 50.7±1%; behavior-only model: 25.7±0.9%; Fig 4b).

We examined the spatial structure of the signal and noise correlations by considering all simultaneously recorded pairs of neurons (n=1492; Fig. 4c). Signal correlation was defined as the pairwise correlation between the predictions of the behavior-only encoding model for each neuron; noise correlation was defined as the pairwise correlation between the residuals of the same model. The signal correlation decreased with distance between neurons during the cue period ($\rho=-0.1$, $p<6\times10^{-5}$), but not the outcome period ($\rho=-0.03$, $p<0.23$). This is consistent with the results from the previous analyses, which had suggested specialized and spatially organized responses during the cue period (Fig. 3d) in contrast to widespread reward responses during the outcome period (Fig. 2a,c). On the other hand, the noise correlations decreased similarly with distance during both the cue period ($\rho=-0.19$, $p<4\times10^{-13}$) and the outcome period ($\rho=-0.14$, $p<4\times10^{-8}$), suggesting that noise correlations arise from electrical synapses or shared inputs between neighboring neurons not accounted for in the model. These findings were confirmed using an alternative method for calculating noise correlations[25], and were robust to the level of neuropil correction (Extended Data Fig. 6c,d).

Are the widespread reward responses in VTA DA neurons during the outcome period consistent with reward prediction error (RPE)? We first confirmed that we can replicate classic RPE during Pavlovian conditioning with 2-photon imaging (Fig. 5a–d). We then sought to determine to what extent reward expectation modulates reward responses in our decision-making task. In this regard, a strength of our task is that it engages two separable dimensions of reward expectation: previous trial outcome, and trial difficulty (Fig. 5e). If DA neurons reflect RPE, we would expect reward responses to be higher whenever reward expectation is low, for both dimensions of reward expectation. Indeed, across the population,

reward responses were modulated by expectation in a manner that was consistent with RPE (Fig. 5f–h; median d'=0.1 comparing reward responses based on median splitting trial difficulty, $p<3\times10^{-12}$; median d'=0.094 comparing reward responses across both previous trial outcomes, $p<6\times10^{-5}$; two-sided Wilcoxon signed rank test and n=232 in both cases). Interestingly, across neurons, the extent of modulation by each dimension of reward expectation was (weakly) correlated, suggesting that neurons are modulated similarly by each type of RPE (Fig. 5i; ρ=.21, p<0.002, Pearson correlation between the RPE d' values for previous trial outcome and trial difficulty for all reward responsive neurons, n=232). In addition, reward responses in all but one of the functionally defined clusters are significantly modulated by RPE (Fig. 5j). In further support of the modulation of reward responses by reward expectation, we found that modulation of the reward response depends on task performance in a manner that is consistent with RPE (performance across individuals: Extended Data Fig. 6e; performance during the shaping protocol: Extended Data Fig. 8, 9).

In summary, we have described organizational principles of the DA system: neurons display specialized and anatomically organized responses to non-reward variables, while the same neurons convey a less specialized reward response. These conclusions depended on combining, for the first time, a high-dimensional behavioral task (6 quantified behavioral variables) with high-dimensional neural recordings (>300 identified VTA DA neurons).

Considering the functional and anatomical organization reported here, alongside the established topography between DA neurons and their downstream targets[19,26,27], we can predict that specific downstream targets are likely to receive information from DA neurons about reward and only a subset of non-reward variables. Thus, this organizational structure may greatly simplify the question of how downstream circuits correctly interpret the wide range of non-reward signals encoded by midbrain DA neurons. A major open question is how downstream targets utilize these specialized non-reward signals. One possibility is that these signals reinforce downstream activity patterns related to the encoded variable, altering the probability that the behavior is repeated (in analogy to the established reinforcement function of reward responses[28,12]). Alternatively, or in addition, they may serve to enhance ongoing activity patterns[29], influencing the vigor of the ongoing behavior[30], but not necessarily the probability of it being repeated in the future. New experiments will likely be designed to address these important hypotheses.

METHODS

Animals and surgery

All experimental procedures were conducted in accordance with the National Institutes of Health guidelines and were reviewed by the Princeton University Institutional Animal Care and Use Committee (IACUC). A total of 31 mice were used in this study. For the virtual reality experiments, we used either male DAT::IRES-Cre mice (n=14, The Jackson Laboratory strain 006660; extensively characterized in [31]) or male mice resulting from the cross of $DAT^{IRES\,cre}$ mice and the GCaMP6f reporter line Ai148 mice [32] (n=6, Ai148×DAT::cre, The Jackson Laboratory strain 030328; see Extended Data Fig. 10 for validation of co-localization of GCaMP and TH in this line). For the Pavlovian conditioning experiments, we used male and female Ai148×DAT::cre mice (n=8). For the slice recording

experiments, we used male and female Ai148×DAT::cre mice (n=3). Mice were maintained on a 12-hour light on – 12-hour light off schedule. All procedures were conducted during their light off period. Mice were 2–6 months old.

Mice between 8–12 weeks underwent sterile stereotaxic surgery under isoflurane anesthesia (3–4% for induction, .75–1.5% for maintenance). The skull was exposed and the periosteum removed using a delicate bone scraper (Fine Science Tools). The edges of the skin were affixed to the skull using a small amount of Vetbond (3M). We injected 800 nl of a viral combination of AAV5-CAG-FLEX-GCaMP6m-WPRE-SV40 (n=12) or AAV5-CAG-FLEX-GCaMP6f-WPRE-SV40 (n=2; U Penn Vector Core) with 1.6×10^{12}/mL titer and AAV9-CB7-CI-mCherry-WPRE-rBG (U Penn Vector Core) with 2.3×10^{12}/mL titer. Two such injections were made at stereotactic coordinates: 0.5 mm lateral, 2.6 or 3.8 mm posterior, 4.7 mm in depth (relative to bregma). After the injections, we implanted a 0.6 mm diameter GRIN lens (GLP-0673, Inscopix or NEM-060–25-10–920-S-1.5p, GrinTech) in the VTA (coordinates shown in Extended Data Fig. 4) using a 3D printed custom lens holder. After implantation, a small amount of diluted metabond cement (Parkell) was applied to affix the lens to the skull using a 1 ml syringe and 18 gauge needle. After 20 minutes, the lens holder grip on the lens was loosened while the lens was observed through the microscope used for surgery to ascertain there was no movement of the lens. Then, a previously described titanium headplate was positioned over the skull using a custom tool and aligned parallel to the stereotax using an angle meter [33]. The headplate was then affixed to the skull using metabond. A titanium ring was then glued to the headplate using dental cement blackened with carbon.

Virtual reality behavioral system

In order to enable a navigation-based decision-making task under head-fixed conditions, we used a virtual reality (VR) system similar to that described previously [34,35] (Fig. 1a). Mice were held head-fixed under a two-photon microscope using two custom headplate holders and ran on an air-supported, Styrofoam spherical treadmill that was 8-inch in diameter. We found that the precise alignment of the mouse on top of the sphere was important for maintaining good behavioral performance; therefore, we used a custom alignment tool for this purpose. The sphere's movement were measured using an optical flow sensor (ADNS3080) located underneath the sphere and controlled by an Arduino Due; this information was sent to the VR computer, running the ViRMEn software engine [36] (https://pni.princeton.edu/pni-software-tools/virmen) under Matlab, which displayed and controlled the VR environment. The measured sphere displacements (dX and dY, where Y is parallel to the long stem of the T-maze) resulted in translational displacements in the virtual environment of equal length in the corresponding axis. The speed of the mouse was given by $\sqrt{\frac{dX^2}{dt}+\frac{dY^2}{dt}}$, where dt was the time elapsed from the previous sampling of the sensor. The mouse acceleration was the moment-by-moment change in speed. The mouse view angle in the virtual world was calculated as follows: first, we calculated the current displacement angle as: $\omega = atan2(-dX \cdot sign(dY), |dY|)$. Then, the rate of change of the view angle (θ) was given by:

$$\frac{d\theta}{dt} = sign(\omega) \cdot min\left(e^{\left(1.4|\omega|^{1.2}\right)} - 1, \frac{\pi}{2}\right) - \theta$$

This exponential function was tuned to stabilize trajectories during the long stem of the maze, while allowing sharp turns into the maze arms (see [15] for more details).

The display was projected using a DLP projector (Mitsubishi HD4000) running at 85 Hz onto a custom toroidal screen with a 270° horizontal field of view. Reward delivery was accomplished by sending by a TTL pulse from the VR computer to a solenoid valve (NResearch) which released a drop of a water to a lick tube located slightly in front and below the mice's mouth. The tone signifying trial failure was played through conventional computer speakers (Logitech). The setup was enclosed in a custom-designed cabinet built from optical rails (Thorlabs) and lined with sound-absorbing foam sheeting (McMaster-Carr).

Optical imaging and data acquisition

Imaging was performed using a custom-built, VR-compatible two-photon microscope [35]. The microscope was equipped with a pulsed Ti:sapphire laser (Chameleon Vision, Coherent) tuned to 920 nm. The scanning unit used a 5 mm Galvanometer and an 8 kHz resonant scanning mirror (Cambridge Technologies). The collected photons were split into two channels by a dichroic mirror (FF562-Di03, Semrock). The light for the green and red channels respectively were filtered using bandpass filters (FF01–520/60 and FF01–607/70, Semrock), and then detected using GaAsP photomultiplier tubes (pmts, 1077PA–40, Hamamatsu). The signal from the pmts was amplified using a high speed current amplifier (59–179, Edmund). Black rubber tubing was attached to the objective (Zeiss 20×, 0.5 NA) as a light shield covering the space from the objective to the titanium ring surrounding the GRIN lens. Double distilled water was used as the immersion medium. The microscope could be rotated along the medial-lateral axis of the mice which allowed alignment of that optical axes of the microscope objective and GRIN lens as described previously for microprism imaging [35]. Control of the microscope and image acquisition were performed using the ScanImage software (Vidrio Technologies; [37]) that was run on a separate (scanning) computer. Images were acquired at 30 Hz at a resolution of 512 × 512 pixels. Average beam power measured at the front of the objective was 40–60 mW. Synchronization between the behavioral logs and acquired images was achieved by sending behavioral information each time the VR environment was refreshed from the VR computer to the scanning computer via an I2C serial bus; behavioral information was then stored in the header of the image files.

Behavioral training

Seven days after the surgery, mice were started on a water restriction protocol, with a daily allotment of water of 1 – 1.5 ml. Mice were monitored for signs of dehydration or drops in body mass below 80% of the initial value. If any of these conditions occurred, mice were given *ad libitum* access to water until recovering. The animals were handled daily from the start of water restriction. 5 days after starting water restriction and handling, mice began

training in the behavioral setup. Training consisted of a shaping procedure with 9 levels of T-mazes with progressively longer stem length and cognitive difficulty (Extended Data Fig. 8). After shaping concluded, in each session the first few trials (5–30) were warm-up trials drawn from mazes 5–8, and then trials from the final maze (#9) were used for the remainder of the session; Warm-up trials were excluded from all analyses in the paper. The mice typically received their daily allotment of water during task performance; if not, the remainder was provided to them at the end of the day.

Details of the behavioral task

At the beginning of each trial, mice were presented with the start of a virtual T-maze. After 30 cm (Start region) the cue region began, in which cues randomly appeared on either side of the corridor. The number of cues presented were sampled from a Poisson distribution, with means of 6.4 to one of the sides, and 1.3 to the other. In order to obtain better estimation of the psychometric curves, we additionally oversampled easy trials by having 5% of trials with a difference in # cues between the sides of 12 or more (using the same probability distributions). The identity of the high-cue-probability and low-cue-probability sides (left or right) were recalculated each trial to randomize the task and avoid side bias [15]. The locations of the cues were randomly assigned along the cue region using a uniform distribution, with the added constraint of a minimum spatial distance of 14 cm between cues (regardless of their side). Each cue was presented when the mouse arrived 10 cm from its location, and disappeared once it was 4 cm behind the mouse. Thus, presentation of multiple cues did not overlap in time. The portion of the maze where cues were presented (cue region) was 220 cm long, and after it the stem of the T-maze continued for another 80 cm where no cues were presented (delay region). At the end of the T-maze the mouse had to enter one of the arms, and full entry constituted a choice. Turning into the correct (more cues) side would elicit a water reward (6.4 uL), while an incorrect choice elicited a tone (pulsing 6 to 12 KHz tone for 1 s). At the time of reward or tone delivery, the visual environment froze for 1 s, and then disappeared for 2 s (after a successful trial) or 5 s (after a failed trial) before another trial was started.

Pavlovian conditioning

After water restriction and handling, mice were habituated to head fixation for 2–3 sessions. Training consisted of 5 sessions (1 session/day); each session consisted of 50 reward deliveries (8 ul of water/reward). During training, each reward was preceded by a 2 s tone that ended at the time of reward delivery. The time between a reward and the next tone delivery was sampled from an exponential distribution with a mean of 40 s. The tone consisted of a sum of multiple sine waves with frequencies of 2, 4, 6, 8 and 16 Khz, and an amplitude of 70dB. All of the mice exhibited anticipatory licking by the end of the 5 days (increase in lick rate after tone presentation but before reward delivery). Some of the mice were previously trained for several days in a similar protocol where the tone amplitude was 60dB and the time between reward and subsequent tone was sampled from a uniform distribution between 5 and 15 s; these mice did not exhibit anticipatory licking until trained in the final protocol. After training, RPE was assessed in a single test session that consisted of 64 trials; 50 of those trials were identical to the training trials (tone followed by reward), 7 trials were unexpected reward trials (reward delivery with no preceding tone) and 7 trials

were unexpected omissions (tone not followed by reward). In all cases the intertrial interval was sampled from an exponential distribution with a mean of 40 s. Trial identity was sampled randomly with the following exceptions: 1- the first 5 trials were standard trials (tone+reward). 2- The first 2 non-standard trials were unexpected reward trials.

Session and trial selection

We selected sessions and trials such that each recorded neuron would only appear in one session, and during which mice were engaged in the task. Our dataset contained one main imaging field/mouse, with the exception of three mice, in which we obtained two separate imaging fields at different depths. Thus, we analyzed 23 sessions from 20 mice (one session per imaging field). Sessions had at least 100 trials and mice performed at least 65% correct. Mice were between 3–6 months old during imaging and were trained for an average of 30 sessions before data collection (a range of 18–51 training sessions).

We removed a small fraction of trials in which mice were not engaged in the task, based on the following criteria: i) We calculated a smoothed performance measure by processing the binary trials success vector through a zero-phase filter composed of a 21 point centered Gaussian with std. dev.=3. Trials where this measure was less than 0.5 were removed. ii) A sequence of 5 or more trials with the same choice and success rate equal or less than 20% was removed. iii) A sequence of 10 or more trials with the same choice was removed. The removed trials comprised 15% of trials per session on average. Most of these trials occurred close to the end of the session when the animals tended to exhibit decreased performance. These trials were not removed for consideration of the mice performance when dividing the mice into two groups based on performance, or from the dataset used when dividing blocks of trials in a session based on performance (Extended Data Fig. 6). Average performance across sessions on all trials was 73.3±1.1%, average performance after removal of these trials was 77.6±0.9%, average performance on the easiest 20% of trials (based on the absolute difference in cues) after removal was 87%±1.7%.

Motion correction procedure

Deep brain imaging can be associated with spatially nonuniform fast motion (frame to frame), as well as spatially nonuniform slow drift of the field of view (over several minutes). To perform accurate motion correction despite the spatial non-uniformity, we divided the video into small regions (‘patches’) that had relatively uniform motion, and separately corrected the motion within each patch, as described below (schematic of procedure in Extended Data Fig. 3; example video before and after motion correction in Supplementary Video 2). Motion correction was performed on the red channel of the recording when available, otherwise it was performed on the green channel (n=9).

Before dividing the video into patches, we first performed rigid motion correction using a standard normalized cross-correlation method, to eliminate any spatially uniform motion (‘matchTemplate’ function in the openCV package in Python). This correction was performed on non-overlapping 50 s video clips to eliminate concerns that slow drift over the course of minutes would degrade performance. The template for the cross-correlation was calculated by dividing each clip into non-overlapping sections of 100 frames, calculating the

mean image of each section, and obtaining the median of the mean images. Before these motion correction steps, the video was pre-processed as follows: i- thresholded by subtracting a constant number and setting negative values to 0, such that the lower ~50% of pixels were 0, ii- used the openCV function 'erode' (with a scalar '1' kernel), iii- convolved with a Gaussian (std. dev. = 2 pixels). Motion correction and template calculation were performed iteratively 10 times or until all absolute shifts were less than 1 pixel in both axes. Finally, the 50 s clips had to be aligned to each other. This required generating a 'master template' for the entire video, and then using the same normalized cross-correlation procedure as before ('matchTemplate' function). The master template was calculated by taking the median of the templates of all clips.

The next step of motion correction involved compensating for spatially nonuniform, slow drift by estimating the drift in local patches. Patches were defined manually around neurons of interest to contain objects that drifted coherently (patch width ~80–160 pixels). In order to estimate the drift of each patch over time, we used a non-rigid image registration algorithm (demons algorithm, 'imregdemons' function in matlab). This algorithm outputs a pixel by pixel correction. However, directly applying this correction risks distorting the shape of the neurons or the amplitude of signals. Therefore, we applied a uniform correction for each patch, based on the average shift of all pixels in the patch (based on the demons output). We implemented the demons algorithm on the templates from the 50 s clips described in the previous paragraph, again using the median of these templates as the 'master template'. The registration and master template was computed iteratively 20 times, or until the increase in the average correlation between each corrected template and the overall template was less than the s.e.m. of these correlations. We found that the performance of the non-rigid registration improved if the templates were first processed through a local normalization procedure [38].

Finally, we performed standard rigid motion correction using the normalized cross-correlation method on each patch and each clip. We then repeated the rigid motion correction after taking a rolling mean of every two frames and downsampling the video by a factor of two. This increased signal strength; we used this downsampled video for subsequent analysis. After correcting for motion within clips, we had to correct across clips. To this end, we performed rigid motion correction on the clip templates. The motion correction code can be found in: https://github.com/benengx/Deep-Brain-Motion-Corr.

Calculation of ΔF/F from the motion-corrected images

The first step in calculating ΔF/F for each neuron was to define the neuron's ROI, as well as the annulus around that ROI that would be used for neuropil correction [39,40]. Each neuron's ROI was defined manually using the mean and std projections of the movie as well as inspecting a movie that was downsampled by a factor of 5. An initial automatic annulus was generated by enlarging the borders of the ROI twice (by 5 um and 10 um); the annulus was the shape contained between the two enlarged borders, where we expect that observed activity would be due to neuropil but not the cell itself. Next, we manually reshaped the annulus region to avoid any visible dendrites, processes or cell bodies, while approximately maintaining its original area.

In order to correct for neuropil contamination, we subtracted a scaled version of the annulus fluorescence from the raw trace ($F_{corr}(t) = F_{raw}(t) - \Upsilon \cdot F_{annulus}(t)$), where $F_{raw}(t)$ is the mean fluorescence in the neuron's ROI at time t, $F_{annulus}(t)$ is the mean fluorescence in the corresponding annulus ROI at time t, and Υ is the correction factor [21,39]). The correction factor is intended to reflect the fraction of the z-section that is generated by neuropil versus the cell that is being imaged. The correction factor used was 0.58, which is in line with previously reported correction factors in GRIN lens imaging [21,41] and resulted in positive corrected traces. After neuropil subtraction, smoothing was performed by processing the corrected trace through a zero-phase filter using a 25 point centered Gaussian with 1.5 samples points std.

ΔF/F at time t was defined as $(F(t)-F_0(t))/F_0(t)$, where $F_0(t)$ is the 8th percentile of the smoothed and neuropil corrected trace based on the preceding 60 seconds of recording.

Selection of neurons in the dataset

Neurons were selected for analysis based on visual inspection of recording stability, using both the images as well as ΔF/F traces. Only neurons that were stable for at least 50 trials were included in the dataset. The full dataset comprised of n=303 neurons from n=20 mice. Of these, n=233 were considered to have a good fit by the encoding model described in the next section (>5% variance explained by the model during the cue period; reduced dataset). The full dataset was used in Fig. 2a, Fig. 3b, Fig. 4b, and Extended Data Fig. 1. For analyses where the specific output values of the encoding model were important, we used the reduced dataset composed of neurons for which the encoding model had a good fit (Fig 2c,d, Fig. 3a,c,d, Fig. 4c, Extended Data Fig. 4, Extended Data Fig. 6, Extended Data Fig. 7). With regards to the dataset collected throughout learning, neurons that had >5% variance explained by the model during the cue period were used in Extended Data Fig. 8b (except for the panel titled "Model Fit", for which all neurons were used). The full learning dataset was used in Extended Data Fig. 8c and Extended Data Fig.9. When analyzing modulation of outcome activity in rewarded trials (Fig. 5f–j,), we used all neurons that had significant reward responses (n=232; see Fig. 2a).

Encoding model

In order to quantify the contribution of behavioral variables to neural activity, we employed an encoding model, which was a multiple linear regression with the ΔF/F trace of each neuron as the dependent variable, and predictors derived from the behavioral variables as the independent variables (Fig. 2b). To derive the predictors, we divided the behavioral variables into 3 classes: "event" variables, "whole trial" variables, and "continuous" variables. "Event" variables (left and right cues, reward) were variables that occurred in discrete points in time. To derive the predictors for these variables, each event was convolved with a 7 degrees-of-freedom regression spline basis set with a 2 s duration, generated using the 'bs' package in R. "Whole-trial" variables (accuracy, previous reward) were variables whose value remained constant for an entire trial. These were coded as binary predictors, with a value of '1' in all time points of trials where the animals received a reward (accuracy) or trials after receiving a reward (previous reward) and '0' elsewhere. "Continuous" variables (position and kinematic variables) could change their value at every time point. In the case

of kinematics, we included 3 "sub-variables" that were closely related to each other: velocity, acceleration, and view angle. Up to 3 predictors were generated per continuous variable (or sub-variable), by raising each variable to the 1st, 2nd and 3rd powers. The optimal number of predictors to use per continuous variable (for each neuron) was assessed by 5-fold cross-validation over trials. (The reason that we used position along the maze as a continuous variable, rather than time in trial, was a previous study [3] which found that on a T-maze in which rats occasionally paused, DA activity seemed to be more closely related to position than time.)

The encoding model thus was:

$$F = \beta_0 + \sum_{k=1}^{K_E} \sum_{j=1}^{N_{sp}} \beta_{jk}^{E} e_j^k + \sum_{k=1}^{K_W} \beta_k^W w_k + \sum_{k=1}^{K_C} \sum_{j=1}^{d_k} \beta_{jk}^{C} (c_k)^j + \varepsilon$$

Where F is ΔF/F of a neuron, e^k_j is the jth spline basis function convolved with the kth event variable, w_k is the predictor for the kth whole-trial variable, c_k is the kth continuous variable, K_E, K_W, K_C are the numbers of Event, Whole-trial, and Continuous variables correspondingly. N_{sp} is the number of splines (7 in all cases), d_k is the maximal polynomial degree used for each kth continuous variable, the β values are the regression coefficients for the different predictors, and ε is a Gaussian noise term. The β values were calculated using the least squares criterion after z-scoring the predictors ('glmfit' matlab function). The code can be found in: https://github.com/benengx/encodingmodel. Example single-trial fits for several cells are shown in Extended Data Fig. 1c.

Model comparison

We tested several behavioral variables on order to optimize the encoding model. The behavioral variables used in the final model (position, cues, kinematics, accuracy, previous reward) were those whose removal resulted in a significant degradation of the fit of the model prediction to the data across the population (Extended Data Fig. 1d). Improved fits were assessed by comparing the R^2 for each model (obtained with 5-fold crossvalidation) with a paired t-test across the population of neurons. We also considered other behavioral variables that did not improve the fit and therefore were not included in the final model (see Extended Data Fig. 1d,e). The other variables that we considered are: *early and late cues*: a separate set of predictors was calculated for cues appearing in the 1st half of the cue region and cues appearing in the 2nd half. *#L - #R*: a predictor that at each timepoint takes the value of the current difference between left-and right-side cues that had appeared in the trial. *|#L - #R|*: a predictor that at each timepoint takes the absolute value of the current difference between left- and right-side cues that had appeared in the trial. *#L, #R*: two predictors that at each timepoint take the value of the current number of either left- or right-side cues that had appeared in the trial. *P(Reward on right) (nominal)*: a predictor that takes the current probability of the right side being rewarded based on the number of left- and right-side cues that had appeared in the trial and the sampling statistics of the cues. Given the Poisson distributions from which the cues were sampled (and ignoring the constraint of minimum distance between cues) this probability is given by the following logistic function:

$\frac{1}{1+(4.92)^{\#L-\#R}}$ where *#L*, *#R* are the current counts of left- and right-sided cues respectively. The value of 4.92 is the ratio of Poisson means for high- and low-cue probability sides. *P(Reward) (nominal)*: a predictor that takes the current probability of being rewarded (i.e. making the correct choice) based on the number of left- and right-side cues that had appeared in the trial and the sampling statistics of the cues. Equivalent to max(*P(Reward on right)*,1-*P(Reward on right)*). *P(Reward on right) (empirical)*: a predictor that takes the current probability of the right side being rewarded based on the number of left- and right-side cues that had appeared in the trial, but instead of using the actual statistics of the cues, this probability was calculated using the psychometric curve of each mouse as the function that related the cue appearances to the probability of each side to be rewarded. Thus, this probability is given by: $\frac{1}{1+a^{\#L-\#R}}$ where the parameter *a* is estimated by fitting a logistic function to the psychometric curve of each mouse. *P(Reward) (empirical)*: a predictor that takes the current probability of being rewarded (i.e. making the correct choice) based on the number of left- and right-side cues that had appeared in the trial and calculated using the psychometric curve of each mouse as the function that related the cue appearances to the probability of each side to be rewarded. Equivalent to max(*P(Reward on right)*,1-*P(Reward on right)*). *Difficulty of previous trial*: a predictor that is the final value of $|\#L - \#R|$ from the previous trial. *Confirmatory/disconfirmatory cues*: Instead of dividing cues in left- and right-sided, cues are divided depending on whether they are confirming or disconfirming the current best estimate of the rewarded side. e.g. if the current count is 3 left-side cues and 1 right-side cue, if the next cue is a left-side cue it is confirmatory, and if it is a right-side cue it is disconfirmatory (in case of an even count the next cue is considered confirmatory).

Calculation of the relative contributions of behavioral variables to neural activity

We quantified the relative contribution of each behavioral variable to neural activity (Fig. 2c,d) by determining how the performance of the encoding model declined when each variable was excluded from the model. We predicted neural activity with all variables ("full model") or by excluding one of the variables ("partial model"), in either case with 5-fold cross-validation (over trials; meaning that in each fold 80% of trials were used for training the model and the remainder of trials were used for testing the model performance). The relative contribution of each behavioral variable was calculated by comparing the variance explained of the partial model to the variance explained of the full model. In the case of the cue period, in which five behavioral variables, relative contribution of each variable was defined as $\left(1-\frac{R^2_{p,i}}{R^2_f}\right)/\sum_{j=1}^{5}\left(1-\frac{R^2_{p,j}}{R^2_f}\right)$ where $R^2_{p,i}$ is the variance explained of the partial model that excludes the i^{th} variable and R^2_f is that of the full model. In the case of the outcome period, two event variables were considered: time of reward and time of outcome (reward or tone delivery). The relative contribution of reward was calculated by comparing the variance explained of a partial model with only the time of outcome, compared to a full model that had both time of reward and time of outcome as event predictors, $1-\frac{R^2_p}{R^2_f}$. This

allowed us to identify variance in the neural activity that could be attributed to reward rather than simply reaching the end of the maze. Negative relative contributions were set to 0 (this occurs when the R^2 of the full model is lower than that of the partial model, due to introduction of noise by the excluded variable).

We used two approaches to exclude variables from the full model and calculate variance explained by the partial model. In the first approach, the partial model was equivalent to the full model, except that the β values of the predictors of the excluded variable were set to zero ("no refitting"). In the second approach, we calculated new β values by re-running the regression without the predictors of the excluded variable ("refitting"). Both approaches to exclude variables produced comparable results; the "no refitting" approach was used to generate the main figures, while comparison with the "refitting" approach is shown in Extended Data Fig. 7b,c,g.

To determine if the contribution of a behavioral variable was statistically significant for each neuron (Fig. 2a; Fig. 3b; Extended Data Fig. 8c; Extended Data Fig. 9), we first calculated the F-statistic of the nested model comparison test where the reduced model was the model without that behavioral variable included. We then proceeded to calculate the same statistic on 1000 instances of shuffled data, where shuffling was performed on non-overlapping 3s bins (to maintain the autocorrelation of the signal). The p-value used for significance was obtained by comparing the value of the original F-statistic to the shuffle distribution, using the Bonferroni correction to account for the number of behavioral variables tested for each neuron; the threshold for significance was a p-value of 0.01 after correction.

To visualize the average responses for all significant neurons for each behavioral variable (Fig. 2a) averaging was performed as follows: In the case of position, accuracy and previous reward, the averaging is over trials. In the case of kinematics, the averaging is over timepoints. In the case of cues and reward, the averaging was across event occurrences. For the event variables (cues and reward), the average baseline activity was subtracted (in the second preceding the event).

Weighted Regression

When calculating the relative contribution of reward (Fig. 2c,d, Fig. 3a, Extended Data Fig. 5, Extended Data Fig. 6e,f, Extended Data Fig. 7, Extended Data Fig. 8b) and the decoding performance of choice and accuracy (Extended Data Fig. 6b), we used weighted regression to control for the different number of trials of each type (correct/incorrect trials or left/right choices). Assuming n_a trials of type a and n_b trials of type b the weights of type a trials are given by: $\frac{n_b}{n_a + n_b}$ and the weights of type b trials are given by: $\frac{n_a}{n_a + n_b}$.

Clustering analysis

To identify functional clusters of neurons (Fig. 3a), we used a clustering procedure based on a Gaussian mixture model (GMM) that was applied on the matrix of contributions of behavioral variables to the neural activity. To do that, we used the 'fitgmdist' function in Matlab (Mathworks, Inc) with 1000 maximum iterations, 0.35 regularization value, 100 replicates, and the covariance matrix constrained to diagonal. This produces a Gaussian

mixture model where the major axes of the Gaussians are parallel to the axes of the feature space, which enables flexibility beyond that of the k-means algorithm while still maintaining a relatively small number of parameters to be fitted.

To test the fit of the clustering model (Extended Data Fig. 6a), we shuffled 10,000 times the relative contribution values both across behavioral variables (Extended Data Fig. 6a, top) and across neurons (Extended Data Fig. 6a, bottom; the contributions for the cue period variables were re-normalized per neuron after shuffling). After each shuffling iteration, we repeated the clustering and recalculated the log-likelihood of the clustering model. The distribution of log- likelihood values for shuffled data was then compared to the log-likelihood of the clustering model on the real data.

The BIC score was used to select the number of clusters. It is a penalized likelihood term defined as $2(NlogL) + Mlog(n)$, where $NlogL$ is the negative log-likelihood of the data, M is the number of parameters of the GMM, and n is the number of observations. The first term rewards model with good fit, while the second term penalizes more complex models. The BIC score was calculated by the 'fitgmdist' function.

Alternative clustering analysis on the predicted traces

In Extended Data Fig. 7i,j, we used an alternative method to functionally cluster the neurons, in order to compare to the clusters described in Fig. 3. Behavioral predictors from one session were used to generate predicted activity traces based on the encoding model, for each neuron that had >5% variance explained by the behavioral model by multiplying the predictor matrix by the weights (n=233). A similarity matrix was constructed by taking the absolute correlation between the predicted traces for each neuronal pair. The similarity matrix was clustered via information-based clustering [20] using the published matlab code with parameters: T=0.1, Csize=5, InitNum=10. Neurons were assigned a cluster identity to the cluster for which they had the highest probability of belonging, provided that probability was higher than 0.75. The confusion matrix shown in extended Data Fig. 7j was constructed from neurons that had a cluster identity in both the relative contributions clustering approach (method used in the main paper) and the alternative method described here (clustering the similarity matrix obtained from the predicted neuronal traces; n=158). The value in bin i,j of the matrix was calculated by $\frac{\#\left(ID_{rel.\ contr.} = j \wedge ID_{pred.\ traces} = i\right)}{\#\left(ID_{rel.\ contr.} = j \vee ID_{pred.\ traces} = i\right)}$.

Quantification of reward prediction error signals with d'

In Fig. 5, the strength of modulation of reward responses by reward expectation was calculated using the d' measure as follows: 1- We divided rewarded trials into trials with either high reward expectation (HRE) or low reward expectation (LRE). For the pavlovian conditioning experiments, HRE trials were those where reward delivery was preceded by a tone, and LRE trials were those where reward delivery was not preceded by a tone. For the virtual reality experiments, trials were divided in two different ways: for the trial difficulty criterion, we ranked trials according to the strength of the evidence (absolute value of the difference between the total number of right- and left-sided cues). The top half of those trials (strong evidence) were considered HRE trials and the bottom half (weak evidence) were

considered LRE trials. For the previous outcome criterion, previously rewarded trials were HRE trials and previously unrewarded trials were LRE trials. 2- We calculated the average reward response in each trial by averaging activity in the first 2 s following reward delivery and subtracting from that the average activity in the 1 s preceding reward delivery. 3- The d' for the reward responses for HRE and LRE trials was calculated as follows:

$$d' = \frac{\mu_{LRE} - \mu_{HRE}}{\sqrt{0.5\left(\sigma^2_{LRE} + \sigma^2_{LRE}\right)}}$$

where μ and σ^2 are the mean and variance of the distribution of reward responses for the denoted trial group. Thus, positive d' values indicate activity consistent with a reward prediction error signal (stronger reward response for low reward expectation trials). To evaluate if RPE was significantly represented across the population (Fig. 5d,h,j) we tested if the d' distribution was significantly different from 0 using a 2-sided Wilcoxon signed rank test. For the d' distributions of the different neuronal clusters (Fig. 5j, right), p-values are shown after a Holm-Bonferroni correction for the 10 distributions. The number of neurons assigned to clusters 1 through 5 (which also had a significant reward response) are 62, 26, 18, 25, and 22 respectively.

Histology

After completion of behavioral experiments, mice were perfused with 4% PFA in PBS, and then brains were removed and postfixed in 4% PFA for 24 additional hours before transferring to 30% sucrose in PBS. After post-fixing, 40 micron sections were made with either a microtome (American Optical 860) or cryostat (Leica CM3050 S). Brain sections were washed with PBST (Phosphate buffered saline with 0.4% Triton x-100) for 30 min, and then placed in blocking buffer (10 ml PBST + 0.2 ml normal donkey serum + 0.1 g bovine serum albumin (sigma A7906–100G) for 1 hour. Sections were incubated overnight at 4° C in primary antibodies for TH (TH Ab; Aves labs, E.C. 1.14.16.2, chicken polyclonal anti-peptide antibody mixture, 1:1000 dilution) and GFP (Molecular probes G10362, rabbit monoclonal, 1:1000 dilution). Sections were then washed with PBST for 30 min, then incubated for 1 hour at room temperature in Alexa fluor 647 (Jackson ImmunoResearch Donkey-anti-chicken, 1:1000 dilution) and Donkey anti-rabbit Alexa fluor 488 (Jackson ImmunoResearch, 711–545-152, 1:1000 dilution). Following PBST washes, sections were mounted in 1:2500 DAPI in Fluoromount-G. Whole sections were imaged with a Nikon Ti2000E microscope.

Estimation of the neurons' location

In order to investigate the relationship between the activity of the neurons and their location in the VTA (Fig. 3c,d), we estimated each neuron's location by combining information about the position of the GRIN lens from histology with the location of the imaged neurons within the field of view. Histological slices stained for Tyrosine hydroxylase (TH) featuring the tract left by the GRIN lens (Extended Data Fig. 4a) were processed through the Wholebrain software [42] by applying registration points using the VTA, SNc and cerebral peduncle as primary markers. The center of the bottom of the lesion was used as a proxy for the center of

the lens, and its location was provided by the atlas coordinates output of the software. These coordinates are derived from the Allen mouse brain Common Coordinate Framework (CCF) mapped to stereotactic coordinates [42].

In order to directly estimate the optical properties of the GRIN lenses, we generated samples from a solution of agarose and fluorescent beads (10um, Molecular Probes). We first confirmed the size of the beads by imaging the samples directly with the 2-photon microscope which was calibrated by previous imaging of a 10um x10um grid (Thorlabs). We then proceeded to image the samples through the two types of GRIN lenses used. Given that GRIN lenses have different magnifications at different imaging depths, we calibrated the magnification factor at each depth by measuring the observed size of the beads in the x-y axes, and used that size to estimate the magnification factor. In order to relate the movement of the stage in the z-axis with the imaging depth of the imaged fields, we also measured the observed size of the beads across the z-axis. The z plane used to image each field of view was estimated by identifying the field of view from a z-stack that was previously obtained for each mouse.

For each neuron, the center of mass of its ROI was used as the marker for the neuron location within the field of view. The absolute location of the neuron was the vector sum of its distance from the lens center in the field of view to the measured location of the lens center in atlas coordinates. These estimates were used in Figs. 3 & 4 and Extended Data Figs. 4 & 6.

The relative concentration across the A/P or M/L axis of neurons belonging to a given cluster (Fig. 3d) was calculated as follows. First, the concentration of neurons belonging to a cluster was estimated using Gaussian kernel smoothing via the ‘ksdensity’ function in Matlab with a bandwidth of 50 um applied only on these neurons. Second, the relative concentration for each cluster was calculated as the concentration per cluster divided by the sum of concentrations calculated for all clusters. To calculate the 95% confidence intervals of the relative concentrations (Fig. 3d, dashed lines), we ran 10000 iterations where in each we randomized the cluster identities of the neurons and then proceeded to calculate the relative concentrations of each cluster as above. For each point in the A/P or M/L axis, the edges of confidence interval were the 2.5 and 97.5 percentiles of the distribution of concentrations calculated from the shuffled data. Significant spatial structure for each cluster along each axis was assessed by comparing the standard deviation of the relative concentrations of the data with that obtained from shuffled distributions, where shuffling was performed 10,000 times by randomizing the locations of the neurons relative to their cluster identity. The obtained p-values (Fig. 3d) were then Holm-Bonferroni corrected for the 10 conditions (5 clusters x 2 axes).

Signal and noise correlations

To investigate how the correlations between pairs of neurons were spatially organized in the VTA, we calculated signal and noise correlations for all pairs of neurons that were simultaneously recorded (Fig. 4c). The signal correlation between a pair of neurons was calculated by correlating the predictions of the encoding model for both neurons in the cue period or outcome period. The noise correlation was the correlation between the residuals

for each neuron pair. We also used an alternative method for estimating the noise correlations [43,25] (Extended Data Fig. 6c). The alternative noise correlation estimate between a pair of neurons (*i*,*j*) was calculated as follows: we first fit an augmented encoding model for neuron *i* which had as an additional predictor the activity of neuron *j*; we then calculated the normalized improvement in the fit using $\Delta V_n(i|j) = \frac{V(i|j) - V(i)}{V(i|j)}$, where *V(i|j)*, *V(i)* are the variances explained by the augmented and original (behavioral-only) encoding models respectively for neuron *i*. We repeated this procedure for neuron *j* and obtained $\Delta V_n(j|i)$. The noise correlation estimate was the mean of the two ΔV_n values. To investigate the relationship between pairwise signal and noise correlations and interneuronal distance we calculated Pearson's linear correlation coefficient and its associated p-value between the pairwise correlations and the pairwise distances for each condition (shown in each panel of Fig. 4c and Extended Data Fig. 6c).

Ex vivo recordings to compare GCaMP6f fluorescence with activity in DA neurons

In order to compare GCaMP6f fluorescence with spike times in DA neurons (Extended Data Fig. 2), we performed *ex vivo* slice imaging and electrophysiolgical recordings in Ai148×DAT::Cre mice. Mice were anesthetized with an i.p. injection of Euthasol (0.06ml/30g) and decapitated. After extraction, the brain was immersed in ice-cold carbogenated NMDG ACSF (92 mM NMDG, 2.5 mM KCl, 1.25 mM NaH2PO4, 30 mM NaHCO3, 20 mM HEPES, 25 mM glucose, 2 mM thiourea, 5 mM Na-ascorbate, 3 mM Na-pyruvate, 0.5 mM CaCl2·4H2O, 10 mM MgSO4·7H2O, and 12 mM N-Acetyl-L-cysteine) for 2 minutes. The pH was adjusted to 7.3–7.4. Afterwards coronal slices (300um) were sectioned using a vibratome (VT1200s, Leica) and then incubated in NMDG ACSF at 34°C for 15 minutes. Slices were then transferred into a holding solution of HEPES ACSF (92 mM NaCl, 2.5 mM KCl, 1.25 mM NaH2PO4, 30 mM NaHCO3, 20 mM HEPES, 25 mM glucose, 2 mM thiourea, 5 mM Na-ascorbate, 3 mM Na-pyruvate, 2 mM CaCl2·4H2O, 2 mM MgSO4·7H2O and 12 mM N-Acetyl-l-cysteine, bubbled at room temperature with 95% 02/5% CO2) for at least 45 mins until recordings were performed.

During cell-attached recordings, slices were perfused with a recording ACSF solution (120 mM NaCl, 3.5 mM KCl, 1.25 mM NaH2PO4, 26 mM NaHCO3, 1.3 mM MgCl2, 2 mM CaCl2 and 11 mM D-(+)-glucose, continuously bubbled with 95% O2/5% CO2) held at 30°C. Picrotoxin (100 μM) was added to the recording solution to block tonic inhibition and promote spontaneous activity. Cell-attached recordings were performed using a Multiclamp 700B (Molecular Devices, Sunnyvale, CA) using pipettes with a resistance of 4–6 MOhm filled with a solution identical to the recording ACSF. Infrared differential interference contrast–enhanced visual guidance was used to select neurons that were 3–4 cell layers below the surface of the slices, which were held at room temperature while the recording solution was delivered to slices via superfusion driven by peristaltic pump. Cell-attached recordings were collected once a seal (200 MOhm to >5 GOhm) between the recording pipette and the cell membrane was obtained. To generate bursts in cells that did not exhibit spontaneous bursting activity, a second glass pipette filled with recording ACSF containing 20 μM NMDA was placed above the recorded cell. Slight positive pressure (~12 psi) was briefly applied (100–250 ms) to generate bursting activity in the recorded cell. During bursts, spikes typically exhibited a gradual reduction in amplitude as observed previously [44].

Action potential currents were recorded in voltage-clamp mode with voltage clamped at 0 mV, which maintained an average holding current of 0 pA. Cell-attached currents were low-pass filtered at 1 kHz and digitized and stored at 10 kHz (Clampex 9; MDS Analytical Technologies). All experiments were completed within 4 hours after slicing the brain. Fluorescence was imaged using a CMOS camera (ORCA-Flash 2.8, Hamamatsu) at 30 Hz using a GFP filter cube set (exciter ET470/40x, dichroic T495LP, emitter ET525/50m).

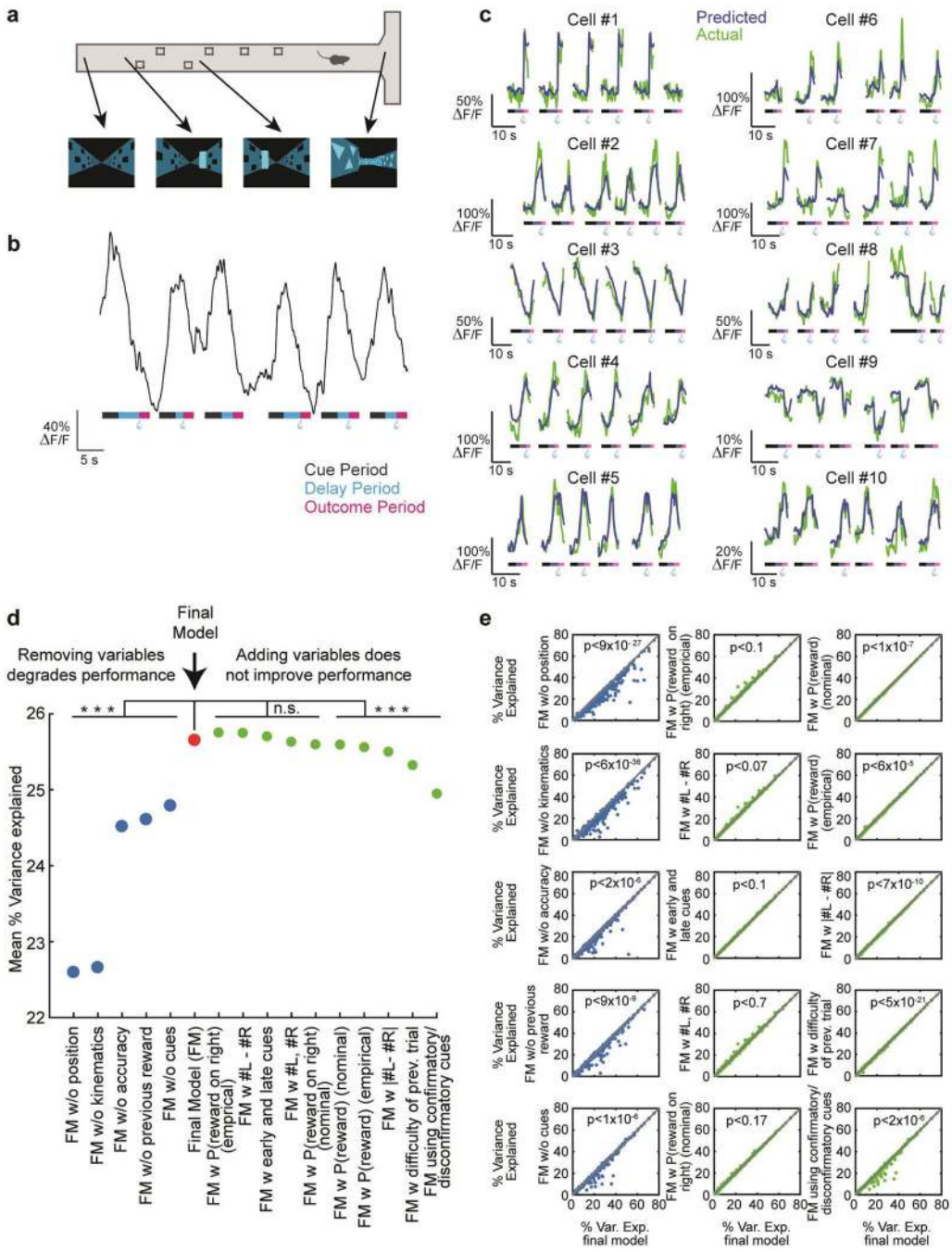
GCaMP6f kernel estimation

To generate fluorescence traces from simulated spike trains (Extended Data Fig. 7k) we estimated a GCaMP6f kernel from [39] by the following equation: $y = e^{-\frac{t}{500}} - e^{-\frac{t}{50}}$ where $t = [0, 1000]$ (t in ms).

Statistical procedures notes

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

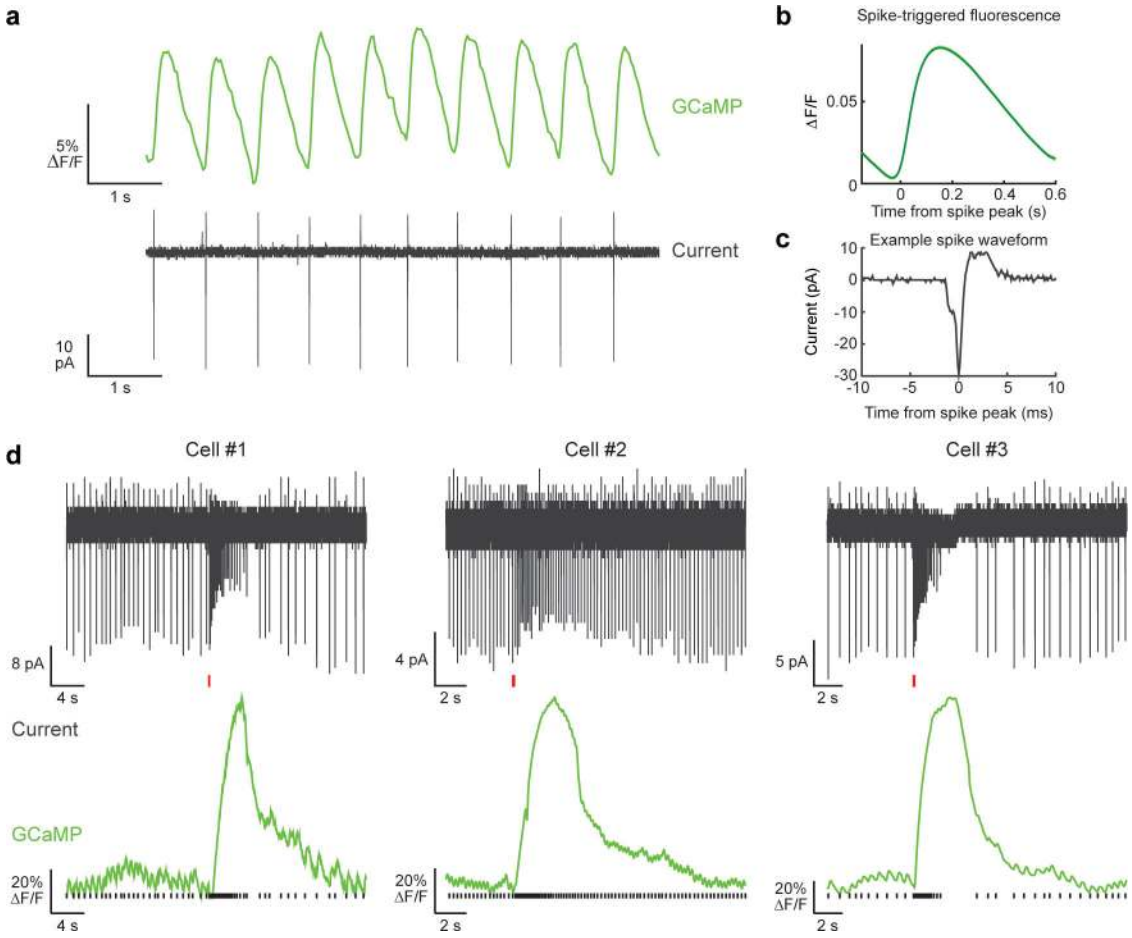
Extended Data



Extended Data Figure 1. Features of the VR task, encoding model predictions, and selection of the encoding model.

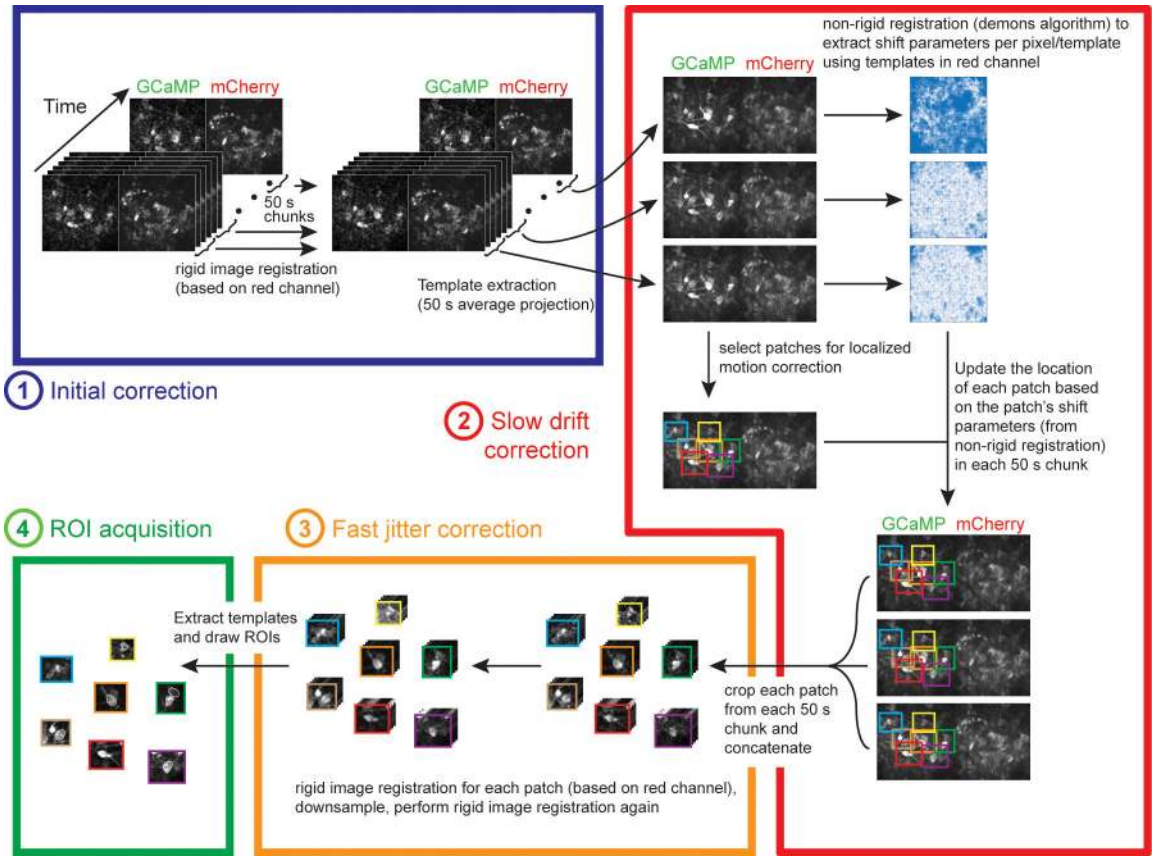
a, Example screenshots of the virtual world presented to the mouse in different positions along the maze. **b,** Activity trace during 6 consecutive trials of an example neuron that was significantly modulated by position in the central stem. The colored strip below the trace describes the trial epochs: cue period (gray), delay period (blue), outcome period (pink). Reward delivery is denoted by a water droplet. **c,** ΔF/F traces for 10 example neurons during 6 consecutive trials (green). Overlaid are the predictions of the behavioral model for these

trials (blue). The colored strip below each trace denotes the trial epochs: cue period (gray), delay period (blue), outcome period (pink). Reward delivery is denoted by a water droplet. **d,** Mean (across neurons) of percent variance explained (tested on held-out data with 5-fold crossvalidation) by the final model (red) and other models where a variables was either removed (blue) or added (green). See Methods for descriptions of all variables that were tested. All models for which a variable was removed from the final model performed significantly worse, based on comparing R^2 for all neurons ($p<2\times10^{-6}$, 2-sided paired t-test, n=303, Holm-Bonferroni correction for all model comparisons). For models where variables were added to those in the final model, the performance either did not exhibit a significant difference, or was degraded. See Methods for complete description of all models. e, Comparison of performance for all neurons of the final model (x-axis) and all the other models. Each panel shows the comparison with one model; significance of the 2-sided paired t-test (After Holm-Bonferroni correction) is shown in each panel. n=303 in all cases.



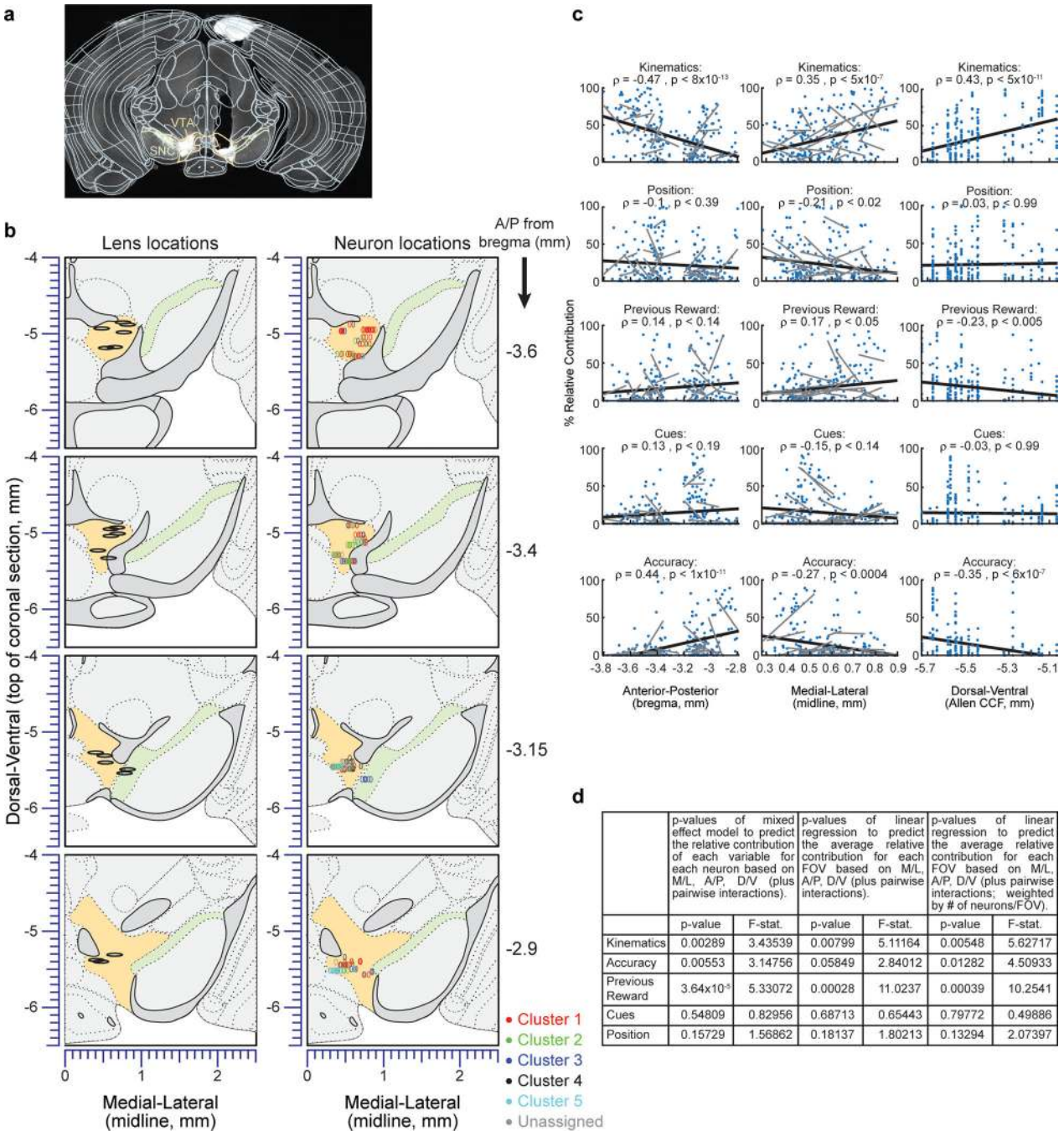
Extended Data Figure 2. Simultaneous calcium imaging and cell-attached recording in DA neurons in the VTA of Ai148×DAT::Cre mice.

a, Relative change in fluorescence (top) and cell-attached current (bottom) recorded simultaneously. **b,** Average spike-triggered fluorescence (average over n=126 spikes). **c,** Zoomed in spike waveform for the same cell as in (a). **d,** Examples of bursts from 3 different DA cellsdifferent DA cells, showing cell-attached current (top) and change in fluorescence (bottom). The spike times are shown with black bars under the fluorescence trace. The red horizontal bars under the current traces show the timing of NMDA puffs (see Methods).



Extended Data Figure 3. Motion correction procedure.

We developed a custom motion correction procedure to compensate for both non-rigid slow drift of the field of view (timescale: 10s of min) as well as non-rigid fast motion (timescale: 10s of ms). Importantly, the procedure avoids any use of interpolation, which can produce artifacts. The procedure consists of the following main steps: **1** (blue box) the entire movie is divided in non-overlapping 50 s chunks; in each chunk we perform rigid motion correction using standard cross-correlation methods (on the red channel). The template for each chunk is calculated by dividing the chunk into non-overlapping sections of 100 frames, calculating the mean image of each section, and obtaining the median of the mean images. **2** (red box) we use a non-rigid algorithm for image registration to align all the templates. The algorithm outputs shift parameters for every pixel and template. Separately, we manually draw patches that include neurons of interest in the first template. For each template, we use the shift parameters of all the pixels in each patch to estimate the average motion of the patch. We use that information to crop the patch from each 50 s chunk of the movie. **3** (orange box) we perform rigid motion correction (as above) on the concatenated patch movies, down-sample by a factor of 2 (to increase the signal strength) and then perform rigid motion correction again. **4** (green box) we extract the patch templates by using the mean projection, and hand draw ROIs of the objects of interest. See Methods for a detailed explanation of motion correction algorithm, and see Supplementary Video 2 for an example video before and after correction. Code available in: https://github.com/benengx/Deep-Brain-Motion-Corr.



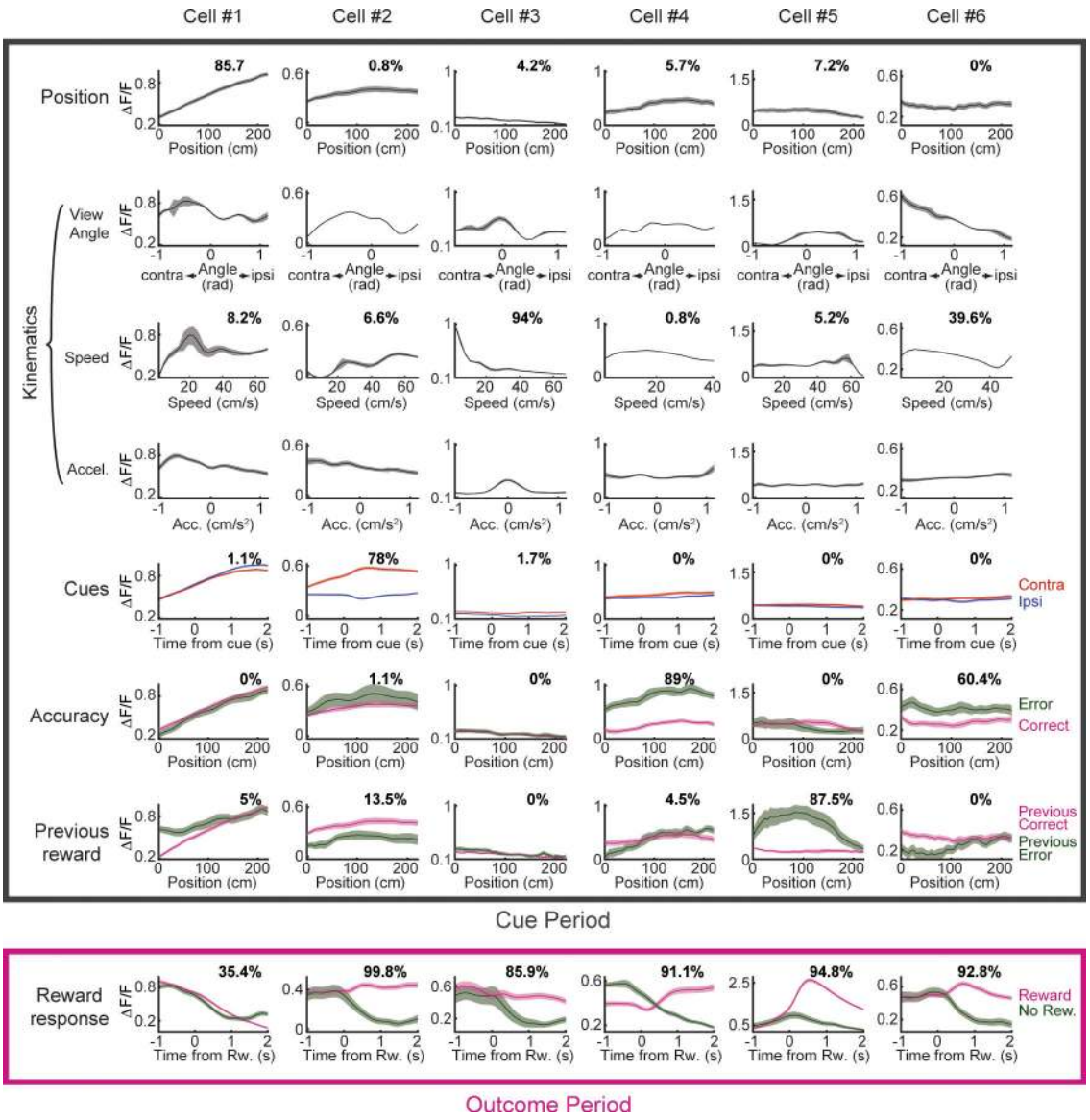
d

	p-values of mixed effect model to predict the relative contribution of each variable for each neuron based on M/L, A/P, D/V (plus pairwise interactions).		p-values of linear regression to predict the average relative contribution for each FOV based on M/L, A/P, D/V (plus pairwise interactions).		p-values of linear regression to predict the average relative contribution for each FOV based on M/L, A/P, D/V (plus pairwise interactions; weighted by # of neurons/FOV).	
	p-value	F-stat.	p-value	F-stat.	p-value	F-stat.
Kinematics	0.00289	3.43539	0.00799	5.11164	0.00548	5.62717
Accuracy	0.00553	3.14756	0.05849	2.84012	0.01282	4.50933
Previous Reward	$3.64x10^{-5}$	5.33072	0.00028	11.0237	0.00039	10.2541
Cues	0.54809	0.82956	0.68713	0.65443	0.79772	0.49886
Position	0.15729	1.56862	0.18137	1.80213	0.13294	2.07397

Extended Data Figure 4. Recovered neuron locations and validation of the spatial organization of neural responses.

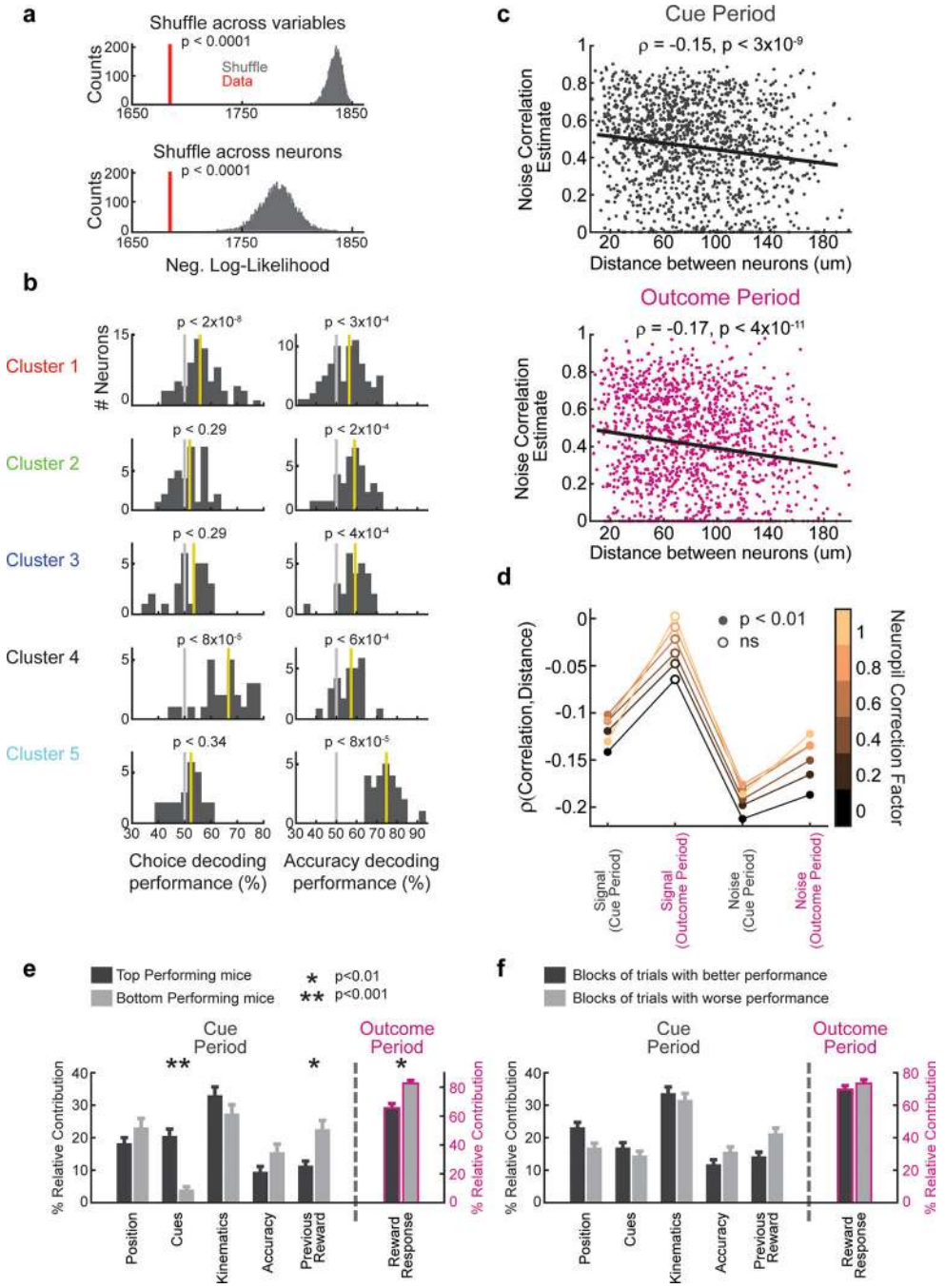
a, Example of lens location recovery. Coronal histological slices stained for Tyrosine hydroxylase were aligned to the Allen brain Atlas [45] using the Wholebrain software package [42]. The center of the lens was marked and its position in common coordinates was recovered by using the software. **b,** Left: recovered centers of GRIN lenses from all mice (black ellipses) are shown on top of the atlas images. Right: recovered locations of all neurons that entered the clustering analysis based on an encoding model R^2 during the cue period >5%

(n=233; see Methods for details on location recovery). Neurons are color-coded according to their cluster identity. **c,** Relative contributions of each behavioral variable as a function of neuron location along the A/P,M/L and D/V axes. In each row, the relative contribution of a behavioral variable is correlated with the A/P (left), M/L (middle) of D/V (right) locations. The correlation value and significance (after Holm-Bonferroni correction for all tests) is shown in the panel (n=233 in all cases). The linear fits of the entire population is shown by a black line, and linear fits of neurons belonging to individual mice (which had more than 5 neurons) are shown by gray lines. **d,** statistical tests of the spatial organization of responses to different behavioral variables that account for individual differences across mice. The table lists the p-values and F-statistics obtained for 3 statistical tests for the spatial organization of the cue-period variables. The first test was a mixed effect model which included all neurons that had good fit to the behavioral model during the cue period (R^2>5%, n=233). In this model, the relative contribution for a given variable to each neuron was the dependent variable, the A/P, M/L, D/V locations and their pairwise interactions were independent fixed effects, and the mouse identity was a random effect for the intercepts (MATLAB code: *model = fitglme(Data,'variable~ml*ap*dv-ml:ap:dv+(1|mouseID)')*). For this test, the degrees of freedom for the numerator and denominator respectively were 6 and 226. In the Field of View (FOV) tests, for every variable we averaged the relative contributions of all neurons in a given FOV. (For mice that had two FOVs we combined neurons from the two FOVs). A regression was run with the average relative contributions as the dependent variable, and the A/P, M/L, D/V of the lens locations and their pairwise interactions were independent fixed effects (n=19). In the weighted version of the FOV test, we additionally weighed each FOV observation by the number of neurons in that FOV. For these two tests, the degrees of freedom for the numerator and denominator respectively were 6 and 12. In all cases the listed p-values correspond to the F-test for the fixed effects.



Extended Data Figure 5. Average activity and relative contributions of different behavioral variables for several example cells.

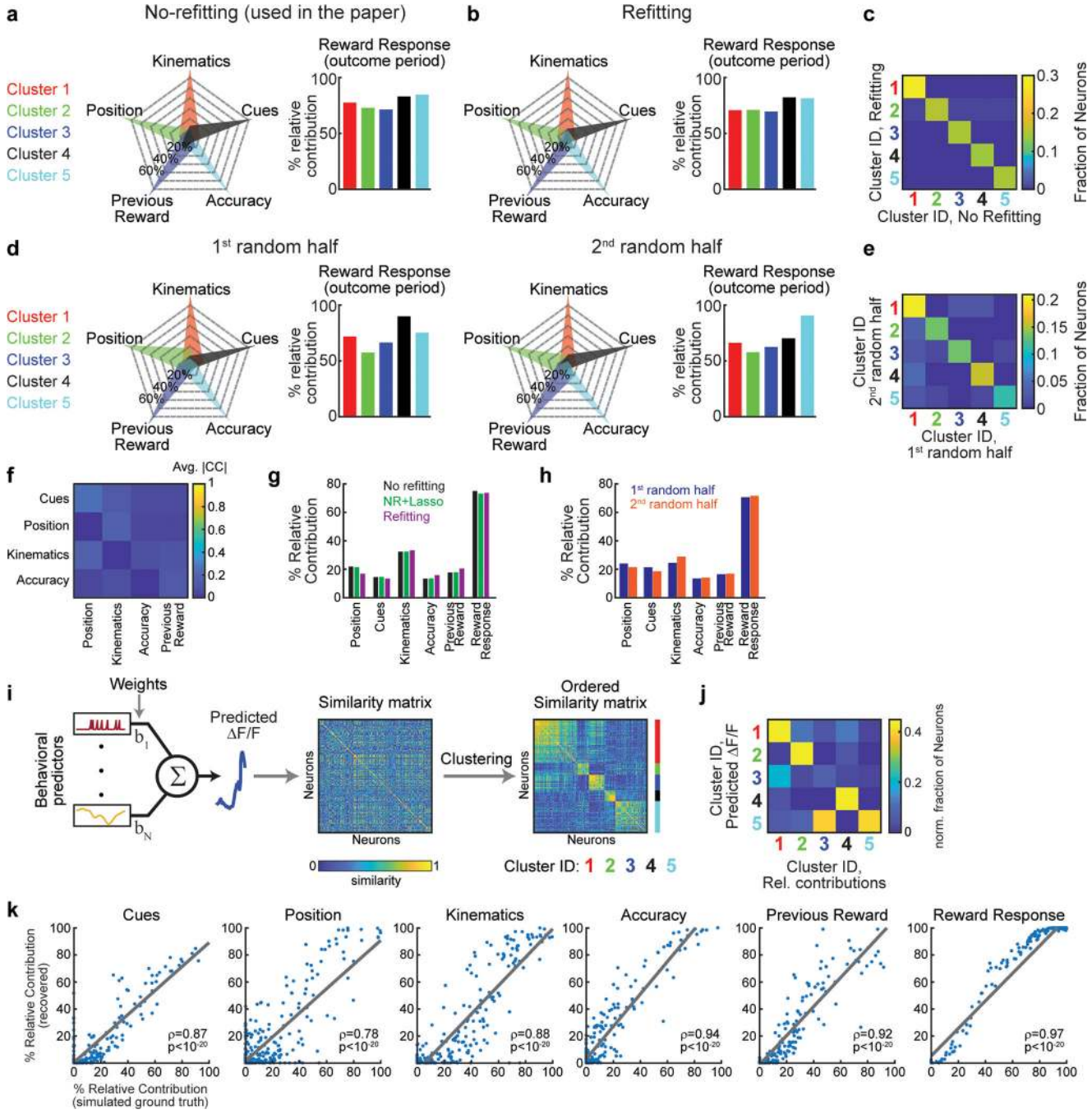
The panels show activity averages time-locked to different behavioral variables for 6 example cells. The percentage of relative contribution of the corresponding behavioral variable to the activity of each cell is displayed in each panel.



Extended Data Figure 6. Additional analyses of neural encoding.

a, Distributions of the negative log-likelihood of the clustering model (Fig. 3) for shuffled (gray) versus real data (red) indicates a significant fit of the clustering model. Top: Shuffling of relative contributions across variables. Bottom: Shuffling across neurons. **b,** Prediction of choice and accuracy from neurons in each cluster. For each neuron, decoding was performed by logistic regression using the average cue period activity (on a trial-by-trial basis) to predict choice or accuracy. Regression was performed using 10-fold crossvalidation (over trials). Separate decoders were trained to predict either choice or accuracy. Weighted

decoding was used to control for the different number of trials of each type (left/right choices or correct/incorrect trials; see Methods). Each panel shows a histogram of the decoding performance for a given variable (left column: choice, right column: accuracy) and a given cluster (rows). Gray vertical lines indicate 50% performance (chance level). Vertical yellow lines indicate the median of the distribution. Significance was assessed by a 2-sided Wilcoxon signed rank test and is presented after a Holm-Bonferroni correction for the 10 tests. For clusters 1 through 5, n= 74, 36, 27, 27, and 26 respectively. The predictive power of the different clusters is broadly consistent with their association with the different behavioral variables: choice was significantly predicted by neurons belonging to clusters 1 (associated primarily with kinematics, which contains the view angle component that is strongly related to choice) and 3 (associated primarily with cues, which determine choice for successful trials). The strongest predictive power for the mice's accuracy is exhibited by cluster 5, which is primarily associated with accuracy.**c,** Noise correlations estimated by an alternative method. Here, noise correlations were estimated by calculating the increase in variance explained by the behavioral-only encoding model when the second neuron activity was added to it as a predictor [25,43]. The noise correlation estimate is shown for all neuronal pairs (n=1492) during the cue period (left) and outcome period (right). **d,** To investigate the possible effect of neuropil contamination on the observed relationship between pairwise correlations and distance (Fig. 4), we systematically varied the neuropil correction factor from 0 to 1 and recalculated the relationship between correlations and interneuronal distance for the different conditions. In all cases, we find a similar pattern to the one presented in the main text: 1- A significant negative slope between distance and signal and noise correlations in the cue period. 2- A significant negative slope between distance and as noise correlations in the outcome period. 3- No relationship between distance and signal correlations in the outcome period. **e,** To investigate the relationship between task performance and neural encoding, mice were divided into 2 groups based on their task performance. The relative contributions of the behavioral variables were averaged separately for neurons belonging to the mice in each group. Consistent with modulation by reward expectation, we found that cue-related activity was stronger and reward responses were weaker in the top performing mice. Interestingly, previous reward (which does not provide useful information for task performance) was more strongly represented in the bottom performing mice (2-sided Wilcoxon signed rank test, n_1=129 neurons in the top performing mice, n_2=104 in the bottom performing mice, with Holm-Bonferroni correction for the 6 tests). **f,** To investigate the relationship between instantaneous performance and neural encoding, for each session, all trials were grouped into blocks of 10 consecutive trials with no overlap; these blocks were split into two groups based on whether the average performance in the block was greater or less than the median performance across all blocks in that session. The panel shows the relative contributions of all behavioral variables calculated separately for the better- or worse- performance blocks. The results did not show a significant difference for any of the variables (2-sided Wilcoxon signed rank test, n_1=n_2=233 neurons, with Holm-Bonferroni correction for the 6 tests), suggesting that the instantaneous performance of each mouse does not have a large effect on the strength of representation of the different variables.

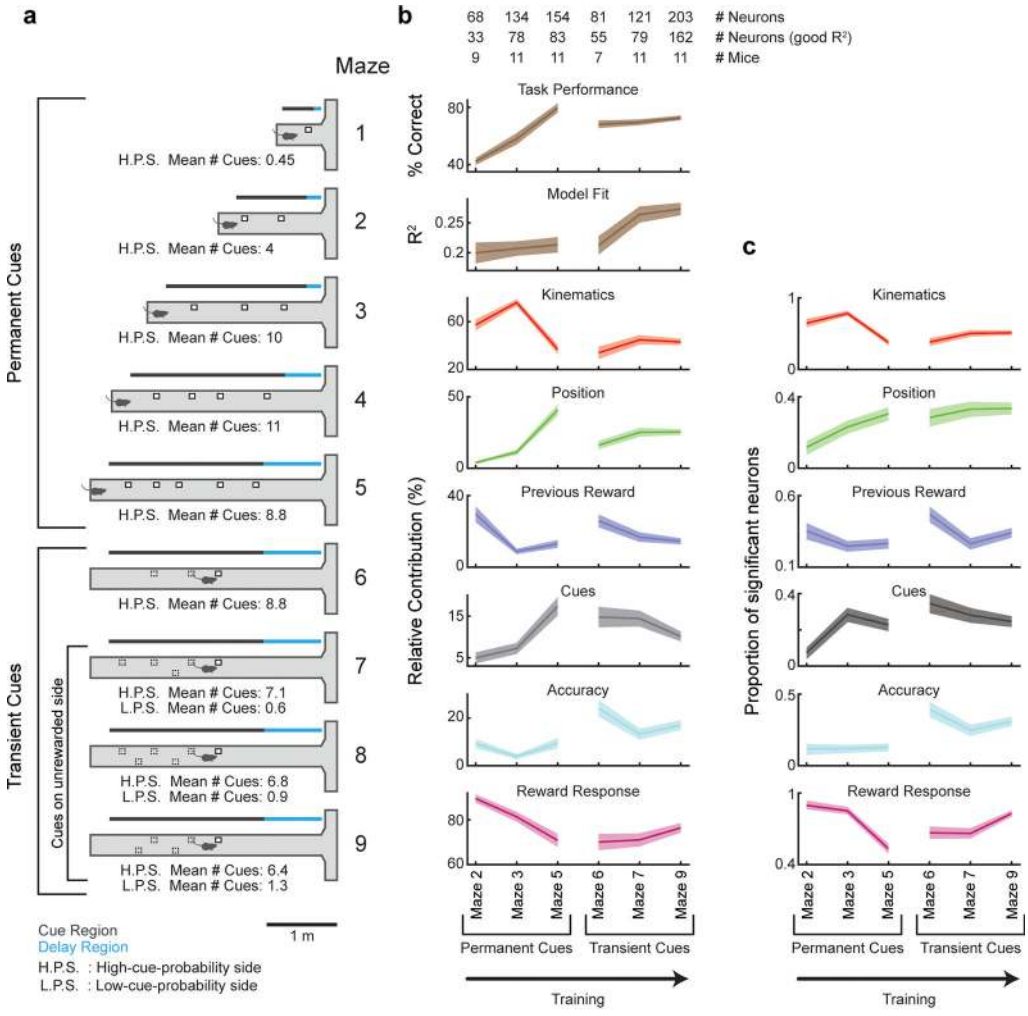


Extended Data Figure 7. Validation of the clustering procedure and encoding model.

a, Summary of average relative contributions of the different behavioral variables for neurons belonging to each cluster as calculated via the approach used in the paper (no-refitting; see Methods). Left: Average relative contributions of cue period behavioral variables to neural activity for each cluster. Right: average relative contribution of reward for each cluster. **b,** Same as **a**, but for the clustering analysis performed on the contributions calculated using the refitting approach (see Methods). **c,** Normalized confusion matrix for the cluster identities of each neuron, obtained by comparing the clustering of the relative

contributions based on either the no-refitting or the refitting approach (see Methods for description of 2 approaches). The main diagonal represents neurons for which the cluster identities matched (97.8%). **d,** Average relative contributions of clusters obtained by separately analyzing two random halves of the trials for each neuron. Correlations between the average relative contributions in each cluster across the two sets are as follows (n=5 in all cases): Position: $\rho = .99$, $p < 8\times10^{-5}$. Cues: $\rho = .99$, $p < 4\times10^{-4}$, Kinematics: $\rho = .99$, $p < 2\times10^{-4}$. Accuracy: $\rho = .99$, $p < 3\times10^{-4}$. Previous Reward: $\rho = .99$, $p < 0.001$. Reward Response: $\rho = .48$, $p < 0.42$. **e,** Normalized confusion matrix for the cluster identities of each neuron, obtained by clustering the two random halves of the data. The main diagonal represents neurons for which the cluster identities matched (79.1%). Note that chance level of matching is 20%. The matrix was calculated for neurons for which a cluster was assigned in the procedures for both halves of the data (>75% probability to belong to a cluster, n=91). **f,** Average absolute value of the correlations for all pairs of predictors across all behavioral variables during the cue period (average across all predictor pairs and mice). **g,** Average relative contributions assessed separately using 3 different approaches: 1- No refitting (NR; used in the paper). 2- No refitting + LASSO regularization (NR+L). 3- Refitting (R). Correlations between the results of the different approaches are as follows: ρ(NR,NR+L) = 1, $p < 7\times10^{-9}$. ρ(NR,R) = .99, $p < 1\times10^{-4}$. ρ(NR+L,R) = .99, $p < 8\times10^{-5}$ (n=6 in all cases). When omitting the reward response contributions: ρ(NR,NR+L) = 1, $p < 2\times10^{-5}$. ρ(NR,R) = .91, $p < 0.04$. ρ(NR+L,R) = .92, $p < 0.03$ (n=5 in all cases). Lasso regularization was applied using the 'lasso' function in Matlab; the mean square error (MSE) of the model was estimated using 5-fold crossvalidation, and we chose the lambda value that minimized the MSE. The results with lasso regularization were almost identical to the result without regularization, suggesting that there was not significant overfitting in our model. **h,** Average relative contributions assessed separately using two random halves of the data. For each neuron we randomly divided all the trials where the neuron was recorded into 2 separate subsets while matching the number of rewarded and previously rewarded trials between the subsets. Each subset of trials was then used to calculate the relative contributions of the behavioral variables. ($\rho = .99$, $p < 3\times10^{-4}$ for all behavioral variables (n=6), $\rho = .8$, $p < 0.11$ when omitting the reward response contributions (n=5)). **i,** We tested the robustness of the clustering results by performing an alternative clustering procedure based on the predicted neuronal traces. The panel depicts the analysis pipeline for this clustering approach: after learning the regression weights for all neurons, behavioral predictors from one session were used to generate predicted activity traces for all neurons. A similarity matrix was constructed by taking the absolute correlation between the predicted traces for each neuronal pair. The similarity matrix was clustered using information-based clustering [20] (see Methods) and ordered by the obtained clusters (right panel; cluster identity for each neuron depicted by a colored stripe to the right of the panel). **j,** Normalized confusion matrix for the cluster identities of each neuron, comparing the cluster identity obtained by clustering the relative contributions (method used in the main text; Fig. 3) and the alternative method described here (clustering the similarity matrix obtained from the predicted neuronal traces). The two clustering methods involve conceptual differences which may result in different clustering organizations. For example, the method used in Fig. 3, which clusters the relative contributions of the behavioral variables, is independent of a particular tuning for these variables, while the method presented here should be affected by such tuning (e.g. upward vs

downward position ramps). Nevertheless, we find a similar overall clustering structure between the two methods, with the following main differences: 1- original clusters 3 and 5 (associated with previous reward and accuracy) are joined in a single cluster (new cluster 5). 2- Original cluster 1 (associated with kinematics) is now split into 2 clusters (new clusters 1 and 3). Further investigation of the split of the kinematics cluster showed that the neurons that split from the main kinematics cluster have stronger modulation for the view angle component of kinematics (based on the regression coefficient values). Such a split could not occur in the formulation used in the main text which combined all the kinematics components (speed, acceleration and view angle). **k,** Further validation of the encoding model by simulating data with known relative contributions of the different behavioral variables. We replaced the activity of each neuron by a simulated trace that was computed using known relative contributions of the different behavioral variables as follows: first, the predictors corresponding to each behavioral variable were summed, resulting in one predictor per variable. Each of these predictors was z-scored and multiplied by a different relative contribution (taken from the values obtained for the real data). The scaled predictors were then summed, resulting in a single vector which forms the basis of the firing rate of the simulated neuron. To this vector we added a constant in order to obtain an average firing rate close to 5 Hz (which was observed in in-vivo electrophysiological recordings [22]). After zeroing negative values of this firing rate vector we used it to generate a spike train using a Poisson process. Finally, the spike train was convolved with an approximate GCaMP kernel (see Methods). We proceeded to estimate the relative contributions for the simulated trace using the encoding model procedure. Each panel shows the relative contributions used to simulate the traces (x-axis) and the recovered contributions (y-axis) for a given behavioral variable; the correlation between the original and recovered relative contributions and its associated p-value are denoted in each panel (n=233 in all cases).



d

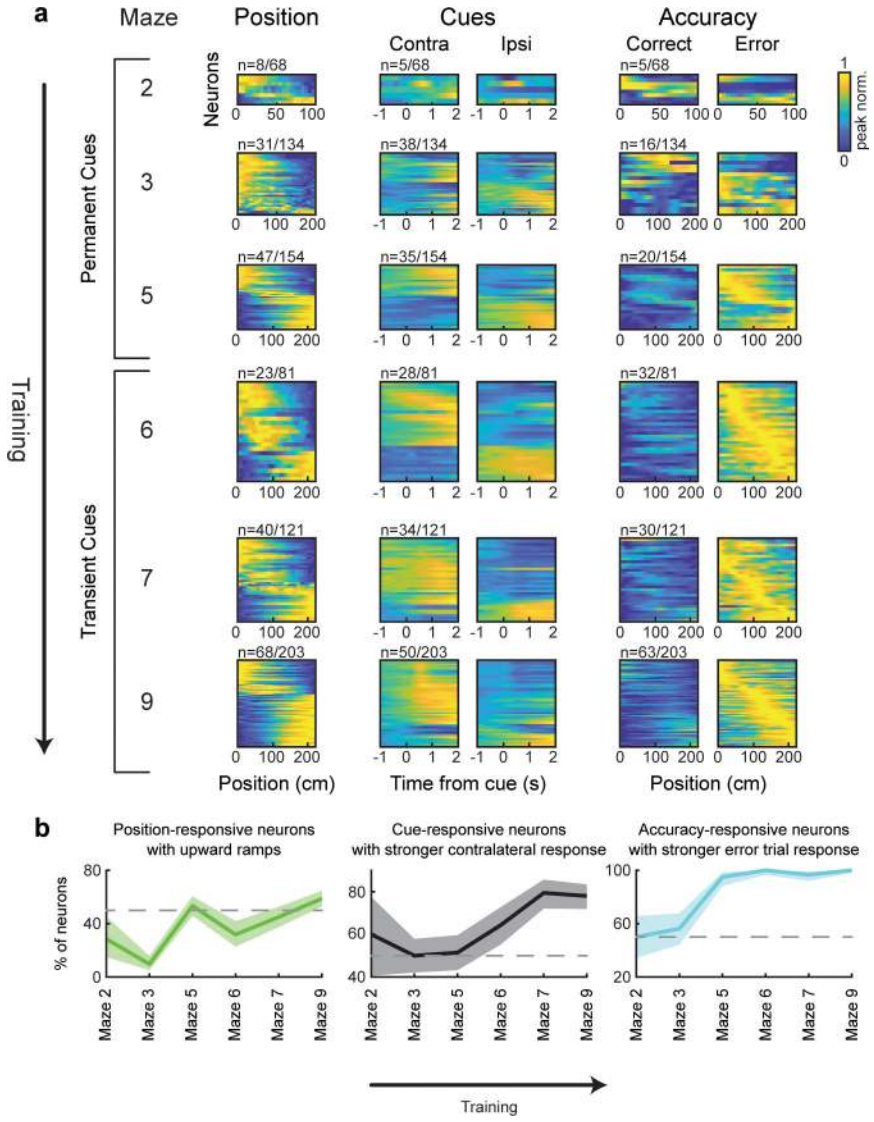
Maze #	1	2	3	4	5	6	7	8	9
Start region length (cm)	5	30	30	30	30	30	30	30	30
Cue region length (cm)	45	100	200	220	220	220	220	220	220
Delay region length (cm)	10	20	20	50	80	80	80	80	80
Permanent cues (y/n)	y	y	y	y	y	n	n	n	n
High-cue-probability side Poisson mean	0.45	4	10	11	8.8	8.8	7.1	6.8	6.4
Low-cue-probability side Poisson mean	none	none	none	none	none	none	0.6	0.9	1.3
Minimum performance for advancing (%)	0	60	80	80	80	80	75	75	-
Number of trials to calculate performance (for advancing)	20	80	100	100	100	100	100	100	-
Minimum number of sessions for advancing	0	0	1	1	2	2	2	1	-

Extended Data Figure 8. Evolution of neural responses throughout learning.

a, Schematic of the shaping protocol. Training consisted of 9 mazes with increasing task difficulty. In the first 5 mazes, cues were permanent and were visible from the beginning of the trial (but still became progressively bigger as the mouse approached them). From maze 6 onward, cues only appeared when the mouse approached within 10 cm of their location. From maze 7 onward, cues could also appear on the unrewarded side. Cues were randomly distributed along the cue region. The number of cues on each side was sampled from a Poisson distribution with the mean indicated for each maze. **b**, Task performance, model fit,

and relative contributions of the behavioral variables throughout learning. The total number of neurons, the number of neurons with good model fit during the cue period (R^2>5%; these were used to calculate the relative contributions of the behavioral variables during the cue period), and the number of mice analyzed in each training stage are indicated at the top. Shaded colors are s.e.m. The results showed that task performance increased steadily across the permanent cue mazes, and then dropped in the first transient cue maze, most likely due to the working memory component that is added in the transient cue mazes. The overall R^2 of the behavioral model increased across learning, indicating that over training, neural activity could be better explained by the measured behavioral variables. Interestingly, the relative contribution of position increased monotonically during the permanent cue mazes, but then dropped during the transient cue mazes, similar to the animals' performance across the mazes. This is consistent with the interpretation of positional ramps as reflecting a value signal [3,18] since the expected value at each position is closely related to reward expectation for that session, and reward expectation is determined by average task performance. The relative contributions for cues also increased during early learning, consistent with being a reflection of the strength of the cue-reward association. Note that this value is somewhat decreased in the last maze, in which (because of the increased task difficulty) each cue has a lower predictive power with respect to reward. The relative contribution of previous reward decreased across the permanent cue mazes, then transiently increased during the first transient cue session. Since relying on previous reward is the wrong strategy in this task, this decrease in the relative contribution of previous reward may relate to animals weighting previous reward more heavily during the major steps in training when they have not yet learned the correct strategy for solving the task. The relative contribution of kinematics declined over the training procedure. This may be due to the kinematic aspect of the behavior becoming less variable over training, as the animal's motor skills improved for VR navigation. Interestingly, the relative contribution of trial accuracy was significantly higher during the transient cue mazes than the permanent cue mazes. This result potentially suggests that DA activity is correlated with task performance preferentially when there is a working memory component. The reward response declined during the permanent cue mazes, and remained relatively consistent during the transient cue mazes; this is consistent with an RPE signal, since RPE implies negative modulation of reward responses by reward expectation (and reward expectation is related to task performance). **c**, Proportion of neurons that were significantly modulated by the different behavioral variables throughout learning (see Methods). Shaded colors show the 1 STD confidence intervals for a binomial distribution calculated using Jeffreys method. **d**, Details of the shaping procedure. The table lists the parameters of the mazes progressively used during the shaping of the behavior. The "permanent cues" field indicates if the cues were presented at the beginning of the trial; otherwise, each cue was presented when the mouse was 10 cm away from its location. "High- (and low) -cue-probability side mean" indicates the means of the Poisson distribution from which the number of cues presented on each side were drawn (at least 1 cue was always drawn); "none" indicates that no cues were presented for the low-probability side on any trial in that maze. The mice were automatically advanced to the next maze if the following criteria were met: 1- their performance was above a predetermined threshold ("minimum performance for advancing" field) for a given number of trials ("number of trials

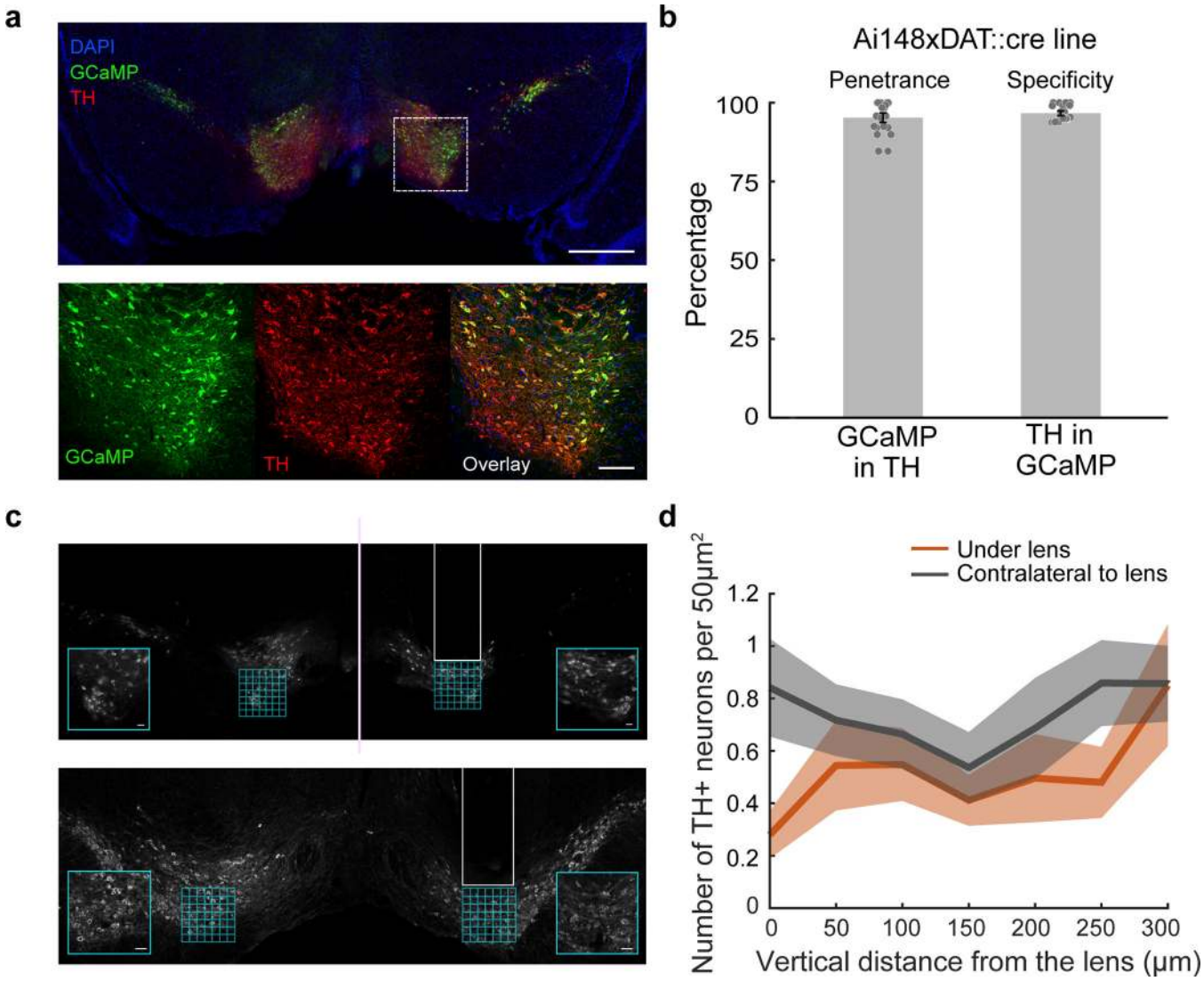
to calculate performance" field). 2- They completed at least n sessions in the current maze, where n is given by the "minimum number of sessions for advancing" field.



Extended Data Figure 9. Neural responses related to position, cues, and accuracy throughout learning.

a, For each behavioral variable (position, cues and accuracy), each heatmap contains all significant neurons for that maze, with each row representing the average response of one neuron (each neuron's activity is normalized by its peak). Statistical significance is assessed by comparing the F-statistic obtained from a nested model comparison with or without each behavioral variable to a distribution of the same F-statistic obtained from shuffled data (see Methods). In the case of position and accuracy, the averaging is over trials. In the case of cues, the averaging is across cue occurrences, and the average baseline activity was subtracted (in the second preceding the cue occurrence). The number of significant and total neurons for that variable and maze are indicated at the top of each heatmap. The height of the heatmaps for each maze is proportional to the average fraction of significant neurons (across variables) for that maze. **b**, Changes in tuning across learning. Left: percentage of neurons with significant responses to position that exhibited a positive slope in their average response. Middle: percentage of neurons with significant responses to cues that exhibited

higher response to contralateral cues (compared to ipsilateral cues). Right: percentage of neurons with significant responses to accuracy that exhibited higher response in error trials (compared to correct trials). Shaded colors show the 1 std. dev. confidence intervals for a binomial distribution calculated using Jeffreys method. The horizontal dotted lines indicate 50% in each panel. Position-selective neurons exhibited early in training more downward ramps than upward ramps (left panel, mazes 2 & 3). Since upward and not downward ramps are consistent with a value signal [3,18], this result suggests an evolution in the specific tuning -and not only the strength of representation- of this variable that is consistent with a value signal. Throughout training, cue-selective neurons are mostly selective for either contralateral or ipsilateral cues, and the preferential representation of contralateral cues develops late in training. This is interesting, because selectivity for contralateral vs ipsilateral cues is not a prediction of the RPE framework. Accuracy-selective neurons exhibit a strong bias towards elevated activity for error trials versus correct trials which was evident by the last permanent cue maze.



Extended Data Figure 10. Specific expression of GCaMP6f in midbrain dopamine neurons in the Ai148×DAT::cre mouse line.

a, Example GCaMP6f expression (green) and TH antibody staining (red). Square indicates location of high-magnification view of GCaMP expression in TH+ neurons. Upper scale bar: 500 μm. Lower scale bar: 100 μm. **b,** Quantification of penetrance and specificity of Ai148×DAT::cre line. Penetrance is the number of TH+ neurons also expressing GCaMP (mean: 95.2%; s.e.m.: 1.52%; *n*=11 sections (1082 cells, 2 mice)). Specificity is the number of GCaMP+ neurons that are also TH+ (mean: 96.7%; s.e.m.: 0.74%; *n*= 11 sections (1075 cells, 2 mice)). **c,** Examples of lesions caused by GRIN lens implants (left). Insets are higher magnification images of the regions where TH+ neurons were counted underneath the lens and compared to counts contralateral to the lens. Scale bar: 50μm. White overlay indicates location of the lesion. Cells were counted in 50 μm by 50 μm squares from 0 to 300 μm below the lens. **d,** Average number of TH+ neurons per 50 μm^2 by distance from the bottom of the lens. Orange: average count under the lens. Gray: average count from the contralateral hemisphere. Shaded colors are s.e.m. n = 11 mice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank J.Y. Choi, S.S.H. Wang, J. Pillow, D. Witten, L. Pinto, S. Bolkan, D. Lee, N. Engelhard, B. Deverett, A. Song, B. Briones, C. Brody, as well as the BRAINCOGS team and the Witten and Tank labs for advice on this work. We also thank E. Engel for reagents. Funding from from ELSC and EMBO (B.E.); NYSCF, Pew, McKnight, NARSAD, and Sloan Foundation (I.B.W.); ARO grants: W911NF-16-1-0474 (N.D), W911NF-17-1-0554 (I.B.W), and NIH grants: U19 NS104648-01, DP2 DA035149-01, 1R01DAA047869-01 and 5R01MH106689-02 (I.B.W.). I.B.W. is a New York Stem Cell Foundation—Robertson Investigator.

REFERENCES

1. Cohen JY, Haesler S, Vong L, Lowell BB & Uchida N Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 482, 85–88 (2012). [PubMed: 22258508]
2. Schultz W, Dayan P & Montague PR A Neural Substrate of Prediction and Reward. Science 275, 1593–1599 (1997). [PubMed: 9054347]
3. Howe MW, Tierney PL, Sandberg SG, Phillips PEM & Graybiel AM Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature 500, 575–579 (2013). [PubMed: 23913271]
4. Howe MW & Dombeck DA Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature 535, 505–510 (2016). [PubMed: 27398617]
5. Barter JW et al. Beyond reward prediction errors: the role of dopamine in movement kinematics. Front. Integr. Neurosci. 9, (2015).
6. Dodson PD et al. Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. Proc. Natl. Acad. Sci. U. S. A 113, E2180–8 (2016). [PubMed: 27001837]
7. da Silva JA, Tecuapetla F, Paixão V & Costa RM Dopamine neuron activity before action initiation gates and invigorates future movements. Nature 554, 244–248 (2018). [PubMed: 29420469]
8. Coddington LT & Dudman JT The timing of action determines reward prediction signals in identified midbrain dopamine neurons. Nat. Neurosci 21, 1563–1573 (2018). [PubMed: 30323275]
9. Kremer Y, Flakowski J, Rohner C & Lüscher C VTA dopamine neurons multiplex external with internal representations of goal-directed action. (2018). doi:10.1101/408062
10. Howard CD, Li H, Geddes CE & Jin X Dynamic Nigrostriatal Dopamine Biases Action Selection. Neuron 93, 1436–1450.e8 (2017). [PubMed: 28285820]
11. Parker NF et al. Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat. Neurosci 19, 845–854 (2016). [PubMed: 27110917]
12. Steinberg EE et al. A causal link between prediction errors, dopamine neurons and learning. Nat. Neurosci 16, 966–973 (2013). [PubMed: 23708143]
13. Bayer HM & Glimcher PW Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47, 129–141 (2005). [PubMed: 15996553]
14. Lak A, Nomoto K, Keramati M, Sakagami M & Kepecs A Midbrain Dopamine Neurons Signal Belief in Choice Accuracy during a Perceptual Decision. Curr. Biol 27, 821–832 (2017). [PubMed: 28285994]
15. Pinto L et al. An Accumulation-of-Evidence Task Using Visual Pulses for Mice Navigating in Virtual Reality. Front. Behav. Neurosci 12, 36 (2018). [PubMed: 29559900]
16. Barretto RPJ, Messerschmidt B & Schnitzer MJ In vivo fluorescence imaging with high-resolution microlenses. Nat. Methods 6, 511–512 (2009). [PubMed: 19525959]
17. Carelli RM Nucleus accumbens cell firing and rapid dopamine signaling during goal-directed behaviors in rats. Neuropharmacology 47, 180–189 (2004). [PubMed: 15464136]
18. Hamid AA et al. Mesolimbic dopamine signals the value of work. Nat. Neurosci 19, 117–126 (2016). [PubMed: 26595651]

19. Kim HF, Ghazizadeh A & Hikosaka O Dopamine Neurons Encoding Long-Term Memory of Object Value for Habitual Behavior. Cell 163, 1165–1175 (2015). [PubMed: 26590420]
20. Slonim N, Atwal GS, Tkacik G & Bialek W Information-based clustering. Proc. Natl. Acad. Sci. U. S. A 102, 18297–18302 (2005). [PubMed: 16352721]
21. Cox J, Pinto L & Dan Y Calcium imaging of sleep-wake related neuronal activity in the dorsal pons. Nat. Commun 7, 10763 (2016). [PubMed: 26911837]
22. Eshel N, Tian J, Bukwich M & Uchida N Dopamine neurons share common response function for reward prediction error. Nat. Neurosci 19, 479–486 (2016). [PubMed: 26854803]
23. Joshua M et al. Synchronization of Midbrain Dopaminergic Neurons Is Enhanced by Rewarding Events. Neuron 62, 695–704 (2009). [PubMed: 19524528]
24. Kim Y, Wood J & Moghaddam B Coordinated activity of ventral tegmental neurons adapts to appetitive and aversive learning. PLoS One 7, e29766 (2012). [PubMed: 22238652]
25. Pillow JW et al. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454, 995–999 (2008). [PubMed: 18650810]
26. Beier KT et al. Circuit Architecture of VTA Dopamine Neurons Revealed by Systematic Input-Output Mapping. Cell 162, 622–634 (2015). [PubMed: 26232228]
27. Lammel S et al. Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. Neuron 57, 760–773 (2008). [PubMed: 18341995]
28. Tsai H-C et al. Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science 324, 1080–1084 (2009). [PubMed: 19389999]
29. Surmeier DJ, Ding J, Day M, Wang Z & Shen W D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. Trends Neurosci. 30, 228–235 (2007). [PubMed: 17408758]
30. Panigrahi B et al. Dopamine Is Required for the Neural Representation and Control of Movement Vigor. Cell 162, 1418–1430 (2015). [PubMed: 26359992]
31. Lammel S et al. Diversity of transgenic mouse models for selective targeting of midbrain dopamine neurons. Neuron 85, 429–438 (2015). [PubMed: 25611513]
32. Daigle TL et al. A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. Cell 174, 465–480.e22 (2018). [PubMed: 30007418]
33. Dombeck DA, Khabbaz AN, Collman F, Adelman TL & Tank DW Imaging large-scale neural activity with cellular resolution in awake, mobile mice. Neuron 56, 43–57 (2007). [PubMed: 17920014]
34. Harvey CD, Coen P & Tank DW Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature 484, 62–68 (2012). [PubMed: 22419153]
35. Low RJ, Gu Y & Tank DW Cellular resolution optical access to brain regions in fissures: Imaging medial prefrontal cortex and grid cells in entorhinal cortex. Proceedings of the National Academy of Sciences 111, 18739–18744 (2014).
36. Aronov D & Tank DW Engagement of neural circuits underlying 2D spatial navigation in a rodent virtual reality system. Neuron 84, 442–456 (2014). [PubMed: 25374363]
37. Pologruto TA, Sabatini BL & Svoboda K ScanImage: flexible software for operating laser scanning microscopes. Biomed. Eng. Online 2, 13 (2003). [PubMed: 12801419]
38. Sage D & Unser M Teaching image-processing programming in Java. IEEE Signal Process. Mag. 20, 43–52 (2003).
39. Chen T-W et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499, 295–300 (2013). [PubMed: 23868258]
40. Kerlin AM, Andermann ML, Berezovskii VK & Reid RC Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. Neuron 67, 858–871 (2010). [PubMed: 20826316]
41. Pinto L & Dan Y Cell-Type-Specific Activity in Prefrontal Cortex during Goal-Directed Behavior. Neuron 87, 437–450 (2015). [PubMed: 26143660]
42. Fürth D et al. An interactive framework for whole-brain maps at cellular resolution. Nat. Neurosci 21, 139–149 (2018). [PubMed: 29203898]

43. Runyan CA, Piasini E, Panzeri S & Harvey CD Distinct timescales of population coding across cortex. Nature 548, 92–96 (2017). [PubMed: 28723889]

44. Mereu G et al. Spontaneous bursting activity of dopaminergic neurons in midbrain slices from immature rats: role of N-methyl-D-aspartate receptors. Neuroscience 77, 1029–1036 (1997). [PubMed: 9130784]

45. Lein ES et al. Genome-wide atlas of gene expression in the adult mouse brain. Nature 445, 168–176 (2007). [PubMed: 17151600]

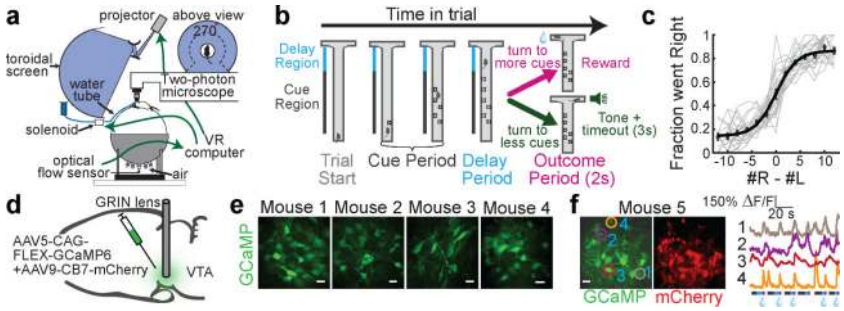


Figure 1. 2-photon imaging of VTA DA neuron during navigation and decision-making in virtual reality.

a, Schematic of the experimental setup. **b,** Schematic of an example trial. In the central stem of the maze, the mouse is presented with transient visual cues to either side ("cue period"). Turning to the arm with more cues results in reward delivery, while turning to the other arm results in a tone and a 3s timeout. **c,** Fraction right choices based on the difference in right vs left cues in each trial. Gray are all individual sessions used in this paper; black are mean, s.e.m., and logistic fit to the mean across sessions. **d,** Schematic of the surgical strategy. **e,** Fields of view for 4 example mice. Scale bar: 20um. **f,** Left: Simultaneous imaging of GCaMP and mCherry in another example animal, with 4 neurons demarcated. Right: traces from those 4 neurons during 6 consecutive trials. Bars below the traces indicate within-trial epochs: cue period (grey), delay period (blue), outcome period (pink). Water drop: reward delivery.

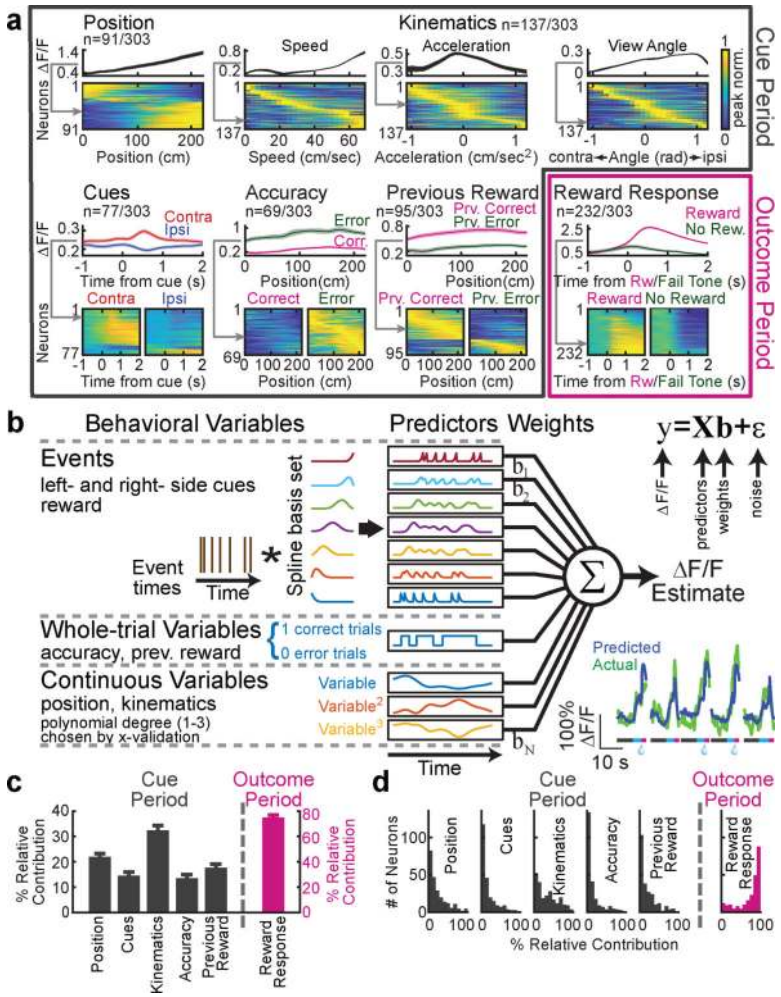


Figure 2. Quantifying VTA DA neuron responses to specific behavioral variables in the task.

a, Neural activity in relation to the following behavioral variables: position along the central stem of the maze, kinematics (speed, acceleration, view angle), cues (contralateral or ipsilateral to the recording side), accuracy (if the mouse made the correct choice at the end of the maze), previous trial reward (if the previous trial was rewarded), and reward (versus not). For each variable, the upper panel is the average **Δ**F/F of an example neuron while the lower panel contains all neurons significantly modulated by that variable, with each row representing the peak-normalized average response of each neuron (grey arrow indicates example neuron within heatmap). See Methods for statistics and averaging. **b,** Schematic of encoding model used to quantify the relationship between behavioral variables and activity of each neuron (see Methods). Inset: predicted and actual **Δ**F/F across 5 trials for one neuron; additional examples in Extended Data Fig. 1c. **c,** Relative contribution of each behavioral variable to explained variance of the neural activity, averaged across neurons. **d,** Same as **c**, but full distribution. All error bars are s.e.m.

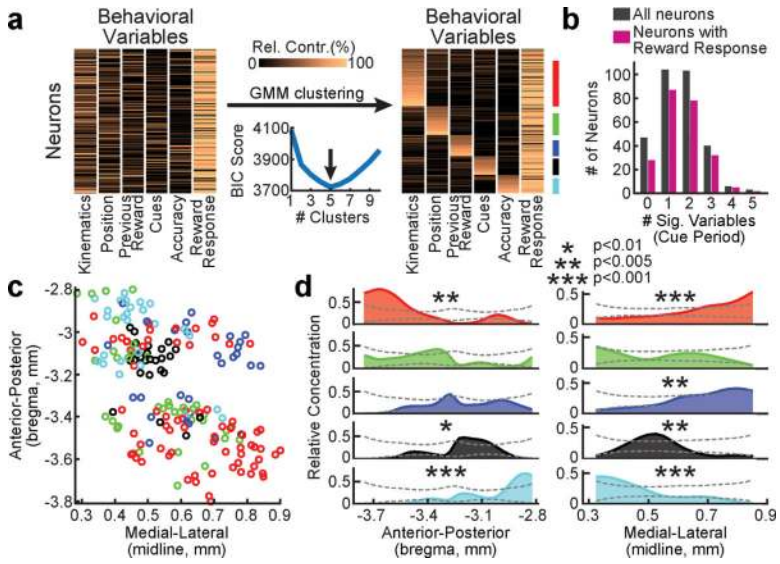


Figure 3. Functional and spatial organization of VTA DA neurons.

a, The clustering procedure. Left: Relative contribution of each behavioral variable to explained variance of neural activity for each neuron, before clustering (all neurons and variables are shown). Right: Same data grouped based on GMM clustering (ordered within each cluster by each neuron's probability to belong to the cluster). Colored vertical lines on the right denote cluster identity. Neurons with <75% probability to belong to any cluster not assigned to a cluster (<18% of neurons unassigned). Bottom middle: BIC scores used to select the optimal number of clusters. **b,** Histogram of the number of behavioral variables during the cue period for which neurons were significantly modulated by, for all neurons (grey) and for the subset of neurons with significant reward response (pink). **c,** Recovered locations within the VTA of each neuron along the A/P and M/L axes. Cluster identity denoted by color. **d,** Relative concentration of neurons belonging to each cluster across the A/P (left) and M/L (right) axes. Dashed lines indicate 95% confidence interval (see Methods).

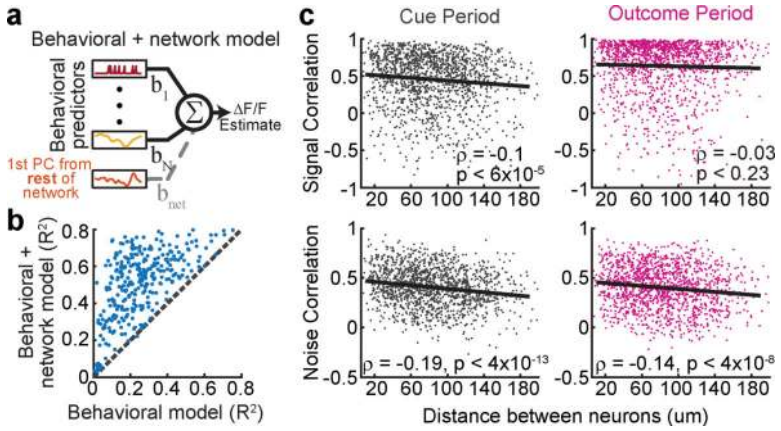


Figure 4. Spatial organization of signal and noise correlations in VTA DA neuron pairs.

a, Schematic of the expanded encoding model (behavioral + network model) which includes one additional predictor compared to that in Fig. 2b: the 1st principal component of the activity of all simultaneously recorded neurons other than the neuron being modeled. **b,** Comparison of the performance of the behavioral-only and the behavioral + network encoding models indicates high noise correlations. **c,** Signal and noise correlations for all simultaneously recorded pairs during the cue period (left) and the outcome period (right) as a function of the distance between the neurons (n=1492).

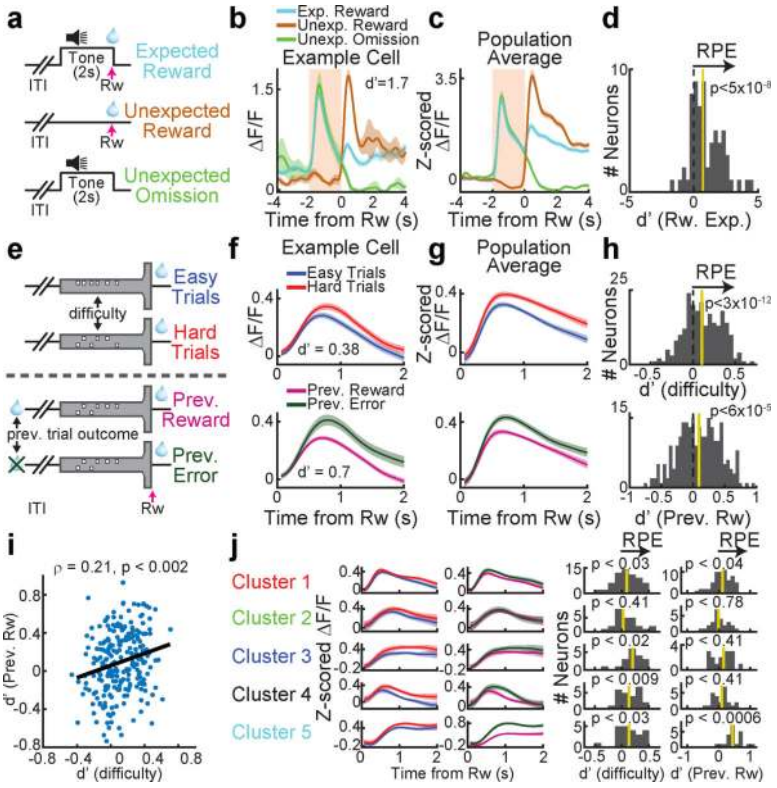


Figure 5. Two separable dimensions of reward expectation modulate reward responses in DA neurons during decision-making.

a, Schematic of the Pavlovian conditioning paradigm for data in panels b-d. **b,** An example cell where reward responses are modulated by expectation, consistent with RPE. d' compares the unexpected and expected reward response (see Methods). **c,** Same as b, but average population response. **d,** histogram of d' comparisons of unexpected and expected reward for all neurons. n=8 mice and n=65 neurons. **e,** In the VR T-maze, two dimensions of reward expectation were quantified: trial difficulty, and previous trial outcome. **f,** An example DA neuron modulated by both RPE dimensions. **g,** Same as f, but average population response. **h,** d' histograms for both RPE dimensions for all reward-responsive neurons (n=232). **i,** Across the population, a significant (but noisy) correlation between the 2 dimensions of RPE. **j,** Reward responses in most functionally defined clusters are significantly modulated by RPE across at least 1 dimension, as shown by the average responses (left) and the d' histograms (right; see Methods for details on significance). In all cases, shaded colors are s.e.m.