Complete representation of action space and value in all striatal pathways.

Moritz Weglage[1], Emil Wärnberg[1], Iakovos Lazaridis[1], Ourania Tzortzi, Konstantinos Meletis*
Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden.

*Correspondence: dinos.meletis@ki.se
[1]Contributed equally

ABSTRACT
The dorsal striatum plays a central role in motor and decision programs, such as the selection and execution of particular actions and the evaluation of their outcomes. A standard circuit model has emerged based on the striatal organization where projection neurons with specific molecular and long-range connectivity identities encode discrete and possibly opposing action signals. We used large-scale cell-type specific imaging of calcium signals during motor and decision behaviors to map the activity of individual striatal projection neurons (SPNs) that form the three major output pathways of the striatum – SPNs in the D1+ direct, the A2A+ indirect, and the Oprm1+ patch pathway. We found that during exploration or choice behaviors SPNs showed a pathway-independent representation of the discrete phases and action variables. The tuning of individual SPNs was action and context-dependent, together covering the entire task space, and included pathway-independent representation of decision-variables such as action value in a dynamic choice task. We propose that the three major SPN pathways broadcast in parallel the complete representation of the task space to downstream targets, including task- and phase-specific signals of action value and choice.

INTRODUCTION

The selection of specific actions is based on computations that produce prediction and evaluation of the action outcome, and correct action selection is essential for the survival of all species. Action selection computations have been associated with neuron activity in basal ganglia circuits, where the striatum plays a central role in integrating information from cortical and subcortical circuits, which is then propagated to downstream targets for action execution[1–3]. The striatum has been neuroanatomically divided into two major output pathways: the direct pathway targeting the globus pallidus interna (GPi) and the substantia nigra (SN), and the indirect pathway targeting the globus pallidus externa (GPe)[4,5]. A circuit model has emerged based on the dichotomous organization of the striatum, where the direct and indirect pathway differentially control motor programs and explain the pathophysiology of movement disorders[6–8]. In this model, the striatal pathways regulate motor behaviors through antagonistic signals. Neurochemical definitions can further divide the striatum into compartments [9,10], for example into patch (also known as striosome) and matrix compartments where the striatal patches exhibit high levels of mu opioid receptor (MOR) expression[11,12] and form a distinct pathway that projects to the GPi and SN[13,14]. Distinct gene expression patterns can be used to genetically target and visualize the direct, indirect, and patch pathway[15–17].

Altogether, evidence from neuroanatomy and physiology have supported that the direct and indirect pathways have opposing effects on behavior. In support of this circuit model, optogenetic manipulation of the direct and indirect striatal pathways has shown their differential role in reinforcement as well as action[18–20]. In contrast, the concomitant activation of both direct and indirect pathways during movements suggests a mixed motor representation[21], and the direct and indirect pathways show complementary roles in action sequences[22], which supports a more complex representation of motion and action in the striatum. Imaging of neuron activity in the direct and indirect pathway[23], in striatal patches[24,25], and in mouse models of movement disorders[26], has together provided support for the opposing role of striatal pathways in the standard model for simple motor and evaluation programs, but also challenged the unitary representation of behavioral events.

The striatum has been shown to encode key variables in decision-making such as action value[27–29], and the striatal patches have been associated with evaluation of actions[30]. To address the role of the three major outputs of the dorsomedial striatum in the representation of motion, actions, and decision-making variables, we recorded the activity of single SPNs belonging to the direct (D1+), indirect (A2A+), or the patch (Oprm1+) pathway using transgenic mice in different behavioral contexts.

RESULTS

The activity of dorsomedial D1+, A2A+, and Oprm1+ SPNs during locomotion.

To define the role of the three major output pathways of the dorsal striatum during action selection and evaluation, we imaged the genetically-encoded calcium sensor GCaMP6s in SPNs through implanted GRIN lenses targeting the anterior part of the dorsomedial striatum in the right hemisphere (Fig. 1A, Suppl. Fig. 1). We imaged the three major output pathways in dorsomedial striatum, defined genetically by the expression of D1 (i.e. D1-Cre for direct pathway), A2A (i.e. A2A-Cre for indirect pathway), or Oprm1 (i.e. Oprm1-Cre for patches) (Fig. 1B). We applied the CNMF-E algorithm to extract regions-of-interest (ROIs) and deconvolved the calcium traces for single neurons[31] to better capture the physiological range dynamics of the neuron activity.

We exposed mice to an open field arena and recorded the activity of SPNs during normal exploration and locomotion. We found on the population level that the D1+, A2a+, but also Oprm1+ pathway showed on average an increase in activity upon movement initiation and a decrease after stopping (Fig. 1D-E). To map in more detail the SPN activity during specific behavioral events, we extracted a set of discrete behaviors in the open field arena, and plotted the activity of selected Oprm1+ neurons to map their tuned responses (Fig. 1F-H). We found many examples of Oprm1+ neurons that showed representation of simple movement categories, such as left versus right turning or running. To better

extract the activity features from SPNs in the three pathways, we computed tuning scores for the main behavioral categories for all recorded SPNs. The distribution of tunings was unimodal for all behavioral categories, but wider than expected by chance (Suppl. Fig. 2). Based on the distributions, we defined the significantly tuned neurons (Fig. 1I-J). Overall, we found significantly tuned neurons that represented for example left turns (contralateral to the recording side), or running speed, and the proportion of tuned neurons for different behavioral categories was similar in the D1+, A2A+, as well as Oprm1+ pathway (Fig. 1K-L). We found that neurons belonging to a specific tuning type were marginally closer in space than expected by chance for all three pathways but did not form spatial clusters (Fig. 1M-N). To assess whether the neuron activity contained enough information to represent the different behavioral categories, we trained support-vector machines (SVMs) to decode the behaviors from the neuron activity data. Supporting a complete representation of the different behavioral categories in all striatal pathways, we found that the behavioral decoding accuracy was high for activity data from SPNs belonging to the D1+, A2A+, as well as Oprm1+ pathway (Fig. 1O-P).

The D1+, A2A+, and Oprm1+ SPNs represent the action space of a choice task.

To image SPN activity during a more complex behavioral task that requires evaluation of actions and updating of predictions regarding action outcome, we modified a two-alternative choice task[23]. In this task, mice were trained to freely initiate a trial by nose-poking into a central port, which allowed them to choose to nose-poke on the right or left side port to receive a liquid reward (Fig. 2A). The probability of reward delivery in the one of the side ports (correct port) was set to 75%. After a reward was delivered, there was a 5% chance of the rewarded side switching. Importantly, mice were not given any cue to indicate the rewarded port, forcing them to keep track and update their choices based on recent trial outcome history. We trained mice in the task for at least 3 weeks, resulting in correct port choice in 67% of the trials, indicating that the mice learned to dynamically adapt their choices to trial outcome (Fig. 2B-C). Mice showed appropriate switching behavior, by updating their choice based on the reward history, showing a win-stay (Fig. 2D), or lose-switch (Fig. 2E) strategy.

To define whether neuron activity reflected strategically different trial types, we focused our analysis on win-stay trials (e.g. reward in the right port followed by choice of right port) versus lose-switch trials (e.g. no reward in the right port followed by choice of left port) (Fig. 2F). We plotted the single neuron activity from D1+, A2A+, and Oprm1+ pathways in the five discrete phases of the win-stay trial, which clearly showed that the peak activities tiled the entire task structure (Fig. 2G). Interestingly, the peak activities from all recorded pathways were similarly distributed across all trial phases of the behavior.

To further explore the relation between neuron activity, trial structure and behavioral strategy, we pooled single-trial recording data for trials following a right choice from many recording sessions to construct one pseudosession, which comprised around 20-30 pseudotrials of each trial type. We then analyzed the single-pseudotrial activity data by performing a dimensionality reduction to extract task-relevant principal components (principal component analysis, PCA), and we found that neuron activity contained structured information to describe the entire task trajectory, but also to discriminate the trial type (win-stay, lose-stay, lose-switch; Fig. 2H). The neuron activity therefore encodes two critical aspects of performing the task: the progression through the trial as well as the decision variables differentiating trial types.

The D1+, A2A+, and Oprm1+ SPNs share a similar representation of the trial phases.

To address the relation between neuron activity in the three pathways and how they represent the progression through the trial, we focused on how individual neurons in the D1+, A2A+, and Oprm1+ pathways were modulated by discrete trial phases.

We found clear examples of Oprm1+ neurons highly tuned to discrete phases in the task, representing specific movements (e.g. center to right turn) as well as outcome and choice signals (Fig. 3A-B).

Furthermore, we found examples of similar sharp tuning for all trial phases in the Oprm1+ pathway (Suppl. Fig. 3). To investigate the tuning profile of all striatal pathways, we calculated tuning scores for twelve discrete task phases (Fig. 3C-D). We found that there was a similar distribution of phase tuned neurons. The tuning in the D1+, A2A+, and Oprm1+ pathways covered all trial phases, and all pathways showed a tuning bias for trial phases with contralateral movements (Fig. 3E). We selected the significantly tuned neurons for the four different movements in the task and visualized their average activity during the phase. We found that the average amplitude was similar and phase-specific for all pathways (Fig. 3F). Importantly, the tuning of single neurons was most often not limited to a single trial phase, and the tuning profiles did not form well-separated clusters, suggesting a rich representation of the trial (Fig. 3G, Suppl. Fig. 3). When we calculated the spatial organization of neurons with a particular tuning profile, we found that these were closer than expected by chance but not distinctly clustered, further supporting the absence of clear functional organization of SPNs in space (Fig. 3H-I).
The rich tuning profile found in the D1+, A2A+, and Oprm1+ pathways suggested that neuron activity should include enough information to categorize the trial phase on a single trial basis. To test this, we trained SVMs to predict trial phases only from neuron activity. We found that activity from only a relatively small fraction of the neurons was sufficient to decode any of the twelve task phases with high accuracy, and that the decoding accuracy was similar for neuron activity from the D1+, A2A+, and Oprm1+ pathways (Fig. 1J-K). The high accuracy of these predictions demonstrates that all three pathways form trial space representations which are reliable on a trial-by-trial basis. Furthermore, the tiling of peak neuron activity across a trial phase, exemplified by Oprm1+ neurons, suggested that activity contained phase substructure (Fig. 3L). We therefore trained SVMs on activity of D1+, A2A+, and Oprm1+ pathways and confirmed that we could accurately predict the progression of the animal through the trial phase (Fig. 3M-N).

The phase tuning of D1+, A2A+, and Oprm1+ SPNs is task-specific and highly conserved across days.
The tuning of single neurons to particular behaviors in the open field and trial phases in the choice task could represent similar actions (e.g. turning, stopping). We therefore aimed to define whether SPNs evidenced comparable tuning to similar motions across different contexts and environments. We recorded the activity of single neurons from the D1+, A2A+, and Oprm1+ pathways first in the open field and immediately after that in the two-choice task, without removing the miniscope. This ensured the complete preservation of neuron identity in the field of view. We found that neurons with significant tuning to left turns in the open field did not maintain that tuning profile during left turns in the choice task (Fig. 4A). Comparing the activity of all neurons in the D1+, A2A+, and Oprm1+ pathway between the open field and the choice task showed that the tuning identity of SPNs was not conserved across contexts (Fig. 4B).
To evaluate whether the tuning profile within a single context is stable across imaging sessions (i.e. between days), we tracked the activity of single neurons in the choice task over days to weeks (Fig. 4C). We found that significantly phase-tuned neurons maintained their tuning profile across several recording sessions, even for very distinctive phase tunings (Fig. 4D). Importantly, we verified SPNs' conservation of tunings across days by using SVMs trained on neuron activity from one session to predict trial phase in subsequent sessions (Fig. 4E). In summary, we found that SPNs are not purely motion-tuned, but instead tuned in a phase- and context -specific fashion, forming complex action representations.

D1+, A2A+, and Oprm1+ pathways represent all aspects of the choice task structure.
The tuning of neurons in the D1+, A2A+, and Oprm1+ pathway to specific trial phases revealed a structured activity pattern that represented the entire trial structure. In addition to the representation

of the trial phases, mice in the choice task also have to keep track of more abstract task variables, such as the value and outcome of actions. To map the neuron activity that differentially represents the variables that match value and decision aspects of the task, we first identified neurons with activity patterns that showed modulation by the trial outcome and subsequent choice, which define the win-stay versus lose-switch strategy. We found single neurons, for example in the Oprm1+ pathway, that showed modulation by trial type, either increasing their activity during lose-switch decisions or increasing activity in win-stay strategies (Fig. 5A). Since the value modulation was phase-specific in the example neurons, we calculated the win-stay versus lose-switch selectivity score for each trial phase separately. We found that the D1+, A2A+, and Oprm1+ pathway selectivity scores were similarly distributed (Fig. 5B, Supp Fig. 5). In addition, all three pathways showed a similar proportion of neurons significantly tuned to the win-stay and to lose-switch trial type, further supporting a pathway-independent representation of the task (Fig. 5C). The value modulation was often phase-specific and not persistent throughout the trial, and it was common for SPNs to show some value modulation in more than one phase. The proportion of phases with value modulation was similar across SPNs from all three pathways (Fig. 5D-F).

To extract the decision variables that accounted for win-stay versus lose-switch, and specifically the impact of reward history on the behavior, we estimated the trial by trial action value using logistic regression (Fig. 5G). The action value estimate accurately reflected the choices and decisions made by the mice (Fig. 5H). To visualize the neuron activity showing trial type selectivity, we focused on a single decision-relevant phase following the outcome in the right port (the right port to center port turn) and the representation of action value in that phase (Fig. 5I-N). We first selected neurons with a significant lose-switch selectivity score in this phase, and plotted the average activity of each neuron during the three main trial types (win-stay, lose-stay, lose-switch).

The lose-switch selective neurons were as expected more active in the lose-switch trial type compared to the win-stay trial type (Fig. 5I). During the lose-stay trial types, these neurons instead showed an intermediary activity profile, suggesting that the population is encoding value rather than trial type identity. We found that the activity of the lose-switch selective neurons was negatively correlated with action value (Fig. 5J). Importantly, since vigor of behavioral response as a metric of motivation could confound the activity signals, we verified that the activity and action value negative correlation could not be explained by a correlation with the action duration (Fig. 5K). We repeated the same analysis for neurons with a significant win-stay selectivity score. The win-stay selective neurons showed increased activity in win-stay trials compared to lose-stay trial in the center to right turn phase, and showed an intermediate activity level in lose-stay trials (Fig. 5L). Correspondingly, the activity of win-stay selective neurons showed positive correlation with the action value, which was not explained by the action duration (Fig. 5M-N).

The action value signal in the right to center turn phase prompted us to determine whether the value and decision variables could be found in other task phases as well. We therefore trained SVMs to predict win-stay versus loses-switch trial types for each of the eight discrete behavioral phases in the task. The single neuron activity in the D1+, A2A+, or Oprm1+ pathway contained enough information to accurately decode the trial type in every trial phase (Fig. 5O). In contrast, SVMs trained on only the action duration could not predict the trial type, again demonstrating the distinction was not due to vigor.

Since the task structure could be divided into three discrete trials types (i.e. win-stay, lose-switch, lose-stay), we used the SVM model - which was only trained using neuron activity in win-stay and lose-switch trials - to estimate the probability of a trial being a win-stay rather than a lose-switch trial type. The SVM prediction confidence in assigning trials as win-stay types correlated with the action value, whereas high prediction confidence for lose-switch trial type correlated with low action values (Fig. 5P). The SVM decoder was not trained on the lose-stay trial type, and it was therefore interesting that the SVM predicted that this trial type was similar to lose-switch for trials with action value approaching

zero, whereas lose-stay trials with higher action value were more similar to low-confidence win-stay trials. Supporting the evidence for pathway-independent representation of action value and trial type, the dependence of the SVM decoding of the trial type on action value was found in the D1+, A2A+, and Oprm1+ pathway (Fig. 5P-Q). In summary, the action value representation by SPNs is phase-specific during a trial and independent of action duration or motivation.

DISCUSSION

We describe here that the activity of SPNs defines the action space including key variables related to trial type and decision-making (e.g. value of the action). The activity did not simply reflect the individual actions as unitary and discrete motions (e.g. left turns) but instead captured the complexity of the task structure, by integrating value and phase information for the execution of a specific action. Individual SPN activity tiled the entire task space, forming a continuous representation of all the actions required to perform the task. Supporting the role of these representations for the proper task execution is the conserved and highly stereotypical tuning of single SPNs to specific trial phases over many separate behavioral sessions and days. We were surprised to find similar tuning in the three molecularly and neuroanatomically distinct output pathways of the striatum, considering the proposed specialization of each pathway. The high-dimensional representation of the entire task and the context-dependent tuning of individual SPNs must be considered when applying simple circuit models of the basal ganglia to understand motor and action behavior.

The standard basal ganglia model is supported by the optogenetic manipulation of direct and indirect pathways, demonstrating their opposing or differential effects on behavior[18–20]. Similarly, recording of direct versus indirect pathway SPNs in dorsomedial striatum has shown a pathway bias in value encoding[32]. In comparison, recording of the striatal pathway activities during motor behavior does not reveal a clear distinction between the pathway activities[21,22]. Interestingly, the encoding of action value has also been recorded in dorsolateral striatum[33], supporting a more complex and rich representation of motor and decision-making variables across striatal regions.

The sequential activation of SPNs has been proposed to be a key mechanism in representing the motor program either as start-stop signals[34] and action chunking[35] or as continuous sequence representations[20,36,37], while other studies have emphasized role of SPNs in categorical coding of value[27,29]. The activity in dorsolateral striatum during a simple locomotion task could not be clustered[37], suggesting that the motor-related activity of SPNs is generally not low-dimensional and instead contains mixed representations of several variables. Our findings support a model where the encoding of action sequences and value are simultaneously represented during action selection, by merging signals for motor and decision signals in a continuous representation that is relevant for proper task performance. The striatal patches have been primarily linked to action evaluation, but we find that similar to the direct/indirect pathways the patch SPNs also show a continuous representation of several aspects of the motor and action program, in a discrete and phase-tuned fashion instead of a sustained value or trial type signal.

It will be important to determine the input pathways that contribute to the rich striatal representation of SPNs. Studies have shown that the dorsal striatum contains neurons that encode spatial information[38,39 41,42], visual and tactile information[40], delay periods[36] and time[41], signals that together can shape the representation of behaviorally relevant information. Even if the corticostriatal functional connectivity has not been comprehensively defined at the detail of single SPNs, the corticostriatal organization has been proposed to be generally pathway-specific[42–44], although evidence also points to converging organization[45]. The dorsomedial striatum receives inputs from the frontal cortex, which also contains signals on the value of tasks[46,47]. The multi-tuning and multiplexing of task-relevant signals has been observed in a number of cortical circuits[48–50], and some of these circuit calculations are likely to be transmitted and represented in the striatal circuitry[51].

We found that SPNs integrate detailed motion signals, with signals about the context, value, and trial phase to produce a rich representation of the progression through the task as well as action value and choice variables. We therefore propose that the dorsal striatum broadcasts an ergocentric (greek [έργον]: work, task) representation of the entire task to all downstream targets: the neuron activity is a merged representation of the phase-specific actions together with abstract representation of key decision-making variables over the entire task space including upcoming choices and the value of specific actions.

REFERENCES

1. Hikosaka, O., Kim, H. F., Yasuda, M. & Yamamoto, S. Basal Ganglia Circuits for Reward Value–Guided Behavior. *Annu. Rev. Neurosci.* **37**, 289–306 (2014).
2. Cox, J. & Witten, I. B. Striatal circuits for reward learning and decision-making. *Nat. Rev. Neurosci.* **20**, 482–494 (2019).
3. Gurney, K., Prescott, T. J. & Redgrave, P. A computational model of action selection in the basal ganglia. I. A new functional anatomy. *Biol. Cybern.* **84**, 401–410 (2001).
4. Tepper, J. M., Abercrombie, E. D. & Bolam, J. P. Basal ganglia macrocircuits. in *Progress in Brain Research* (eds. Tepper, J. M., Abercrombie, E. D. & Bolam, J. P.) vol. 160 3–7 (Elsevier, 2007).
5. Kreitzer, A. C. Physiology and pharmacology of striatal neurons. *Annu. Rev. Neurosci.* **32**, 127–147 (2009).
6. Albin, R. L., Young, A. B. & Penney, J. B. The functional anatomy of basal ganglia disorders. *Trends Neurosci.* **12**, 366–375 (1989).
7. Alexander, G. E. & Crutcher, M. D. Functional architecture of basal ganglia circuits: neural substrates of parallel processing. *Trends Neurosci.* **13**, 266–271 (1990).
8. Nelson, A. B. & Kreitzer, A. C. Reassessing Models of Basal Ganglia Function and Dysfunction. *Annu. Rev. Neurosci.* **37**, 117–135 (2014).
9. Olson, L., Seiger, Å. & Fuxe, K. Heterogeneity of striatal and limbic dopamine innervation: Highly fluorescent islands in developing and adult rats. *Brain Res.* **44**, 283–288 (1972).
10. Graybiel, A. M. & Ragsdale, C. W. Histochemically distinct compartments in the striatum of human, monkeys, and cat demonstrated by acetylthiocholinesterase staining. *Proc. Natl. Acad. Sci.* **75**, 5723–5726 (1978).
11. Pert, C. B., Kuhar, M. J. & Snyder, S. H. Opiate receptor: autoradiographic localization in rat brain. *Proc. Natl. Acad. Sci. U. S. A.* **73**, 3729–3733 (1976).
12. Märtin, A. *et al.* A Spatiomolecular Map of the Striatum. *Cell Rep.* **29**, 4320-4333.e5 (2019).
13. Jiménez-Castellanos, J. & Graybiel, A. M. Compartmental origins of striatal efferent projections in the cat. *Neuroscience* **32**, 297–321 (1989).
14. Fujiyama, F. *et al.* Exclusive and common targets of neostriatofugal projections of rat striosome neurons: a single neuron-tracing study using a viral vector. *Eur. J. Neurosci.* **33**, 668–677 (2011).
15. Gerfen, C. R. *et al.* D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. *Science* **250**, 1429–1432 (1990).
16. Gerfen, C. R. & Surmeier, D. J. Modulation of Striatal Projection Systems by Dopamine. *Annu. Rev. Neurosci.* **34**, 441–466 (2011).
17. Gong, S. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).
18. Kravitz, A. V., Tye, L. D. & Kreitzer, A. C. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nat. Neurosci.* **15**, 816–818 (2012).

19. Tai, L.-H., Lee, A. M., Benavidez, N., Bonci, A. & Wilbrecht, L. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat. Neurosci.* **15**, 1281–1289 (2012).
20. Geddes, C. E., Li, H. & Jin, X. Optogenetic Editing Reveals the Hierarchical Organization of Learned Action Sequences. *Cell* **174**, 32-43.e15 (2018).
21. Cui, G. *et al.* Concurrent activation of striatal direct and indirect pathways during action initiation. *Nature* **494**, 238–242 (2013).
22. Tecuapetla, F., Jin, X., Lima, S. Q. & Costa, R. M. Complementary Contributions of Striatal Projection Pathways to Action Initiation and Execution. *Cell* **166**, 703–715 (2016).
23. Klaus, A. *et al.* The Spatiotemporal Organization of the Striatum Encodes Action Space. *Neuron* **95**, 1171-1180.e7 (2017).
24. Bloem, B., Huda, R., Sur, M. & Graybiel, A. M. Two-photon imaging in mice shows striosomes and matrix have overlapping but differential reinforcement-related responses. *eLife* **6**, e32353 (2017).
25. Yoshizawa, T., Ito, M. & Doya, K. Reward-Predictive Neural Activities in Striatal Striosome Compartments. *eNeuro* **5**, (2018).
26. Parker, J. G. *et al.* Diametric neural ensemble dynamics in parkinsonian and dyskinetic states. *Nature* **557**, 177–182 (2018).
27. Samejima, K., Ueda, Y., Doya, K. & Kimura, M. Representation of Action-Specific Reward Values in the Striatum. *Science* **310**, 1337–1340 (2005).
28. Lau, B. & Glimcher, P. W. Value representations in the primate striatum during matching behavior. *Neuron* **58**, 451–463 (2008).
29. Wang, A. Y., Miura, K. & Uchida, N. The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nat. Neurosci.* **16**, 639–647 (2013).
30. Friedman, A. *et al.* A Corticostriatal Path Targeting Striosomes Controls Decision-Making under Conflict. *Cell* **161**, 1320–1333 (2015).
31. Giovannucci, A. *et al.* CaImAn an open source tool for scalable calcium imaging data analysis. *eLife* **8**, e38173 (2019).
32. Shin, J. H., Kim, D. & Jung, M. W. Differential coding of reward and movement information in the dorsomedial striatal direct and indirect pathways. *Nat. Commun.* **9**, 1–14 (2018).
33. Stalnaker, T. A., Calhoon, G. G., Ogawa, M., Roesch, M. R. & Schoenbaum, G. Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. *Front. Integr. Neurosci.* **4**, (2010).
34. Jin, X. & Costa, R. M. Start/Stop Signals Emerge in Nigrostriatal Circuits during Sequence Learning. *Nature* **466**, 457–462 (2010).
35. Graybiel, A. M. The basal ganglia and chunking of action repertoires. *Neurobiol. Learn. Mem.* **70**, 119–136 (1998).
36. Akhlaghpour, H. *et al.* Dissociated sequential activity and stimulus encoding in the dorsomedial striatum during spatial working memory. *eLife* **5**, e19507 (2016).
37. Sales-Carbonell, C. *et al.* No Discrete Start/Stop Signals in the Dorsal Striatum of Mice Performing a Learned Action. *Curr. Biol.* **28**, 3044-3055.e5 (2018).
38. van der Meer, M. A. A., Johnson, A., Schmitzer-Torbert, N. C. & Redish, A. D. Triple Dissociation of Information Processing in Dorsal Striatum, Ventral Striatum, and Hippocampus on a Learned Spatial Decision Task. *Neuron* **67**, 25–32 (2010).
39. Hinman, J. R., Chapman, G. W. & Hasselmo, M. E. Neuronal representation of environmental boundaries in egocentric coordinates. *Nat. Commun.* **10**, (2019).
40. Reig, R. & Silberberg, G. Multisensory Integration in the Mouse Striatum. *Neuron* **83**, 1200–1212 (2014).
41. Mello, G. B. M., Soares, S. & Paton, J. J. A Scalable Population Code for Time in the Striatum.

Curr. Biol. **25**, 1113–1122 (2015).
42. Gerfen, C. R. The neostriatal mosaic: striatal patch-matrix organization is related to cortical lamination. *Science* **246**, 385–388 (1989).
43. Lei, W., Jiao, Y., Mar, N. D. & Reiner, A. Evidence for Differential Cortical Input to Direct Pathway versus Indirect Pathway Striatal Projection Neurons in Rats. *J. Neurosci.* **24**, 8289–8299 (2004).
44. Wall, N. R., De La Parra, M., Callaway, E. M. & Kreitzer, A. C. Differential innervation of direct- and indirect-pathway striatal projection neurons. *Neuron* **79**, 347–360 (2013).
45. Smith, J. B. *et al.* Genetic-Based Dissection Unveils the Inputs and Outputs of Striatal Patch and Matrix Compartments. *Neuron* **91**, 1069–1084 (2016).
46. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).
47. Hwang, E. J. *et al.* Corticostriatal Flow of Action Selection Bias. *Neuron* **104**, 1126-1140.e6 (2019).
48. Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. & Churchland, A. K. Single-trial neural dynamics are dominated by richly varied movements. *Nat. Neurosci.* **22**, 1677–1686 (2019).
49. Stringer, C. *et al.* Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, (2019).
50. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).
51. Peters, A. J., Steinmetz, N. A., Harris, K. D. & Carandini, M. Striatal activity reflects cortical activity patterns. *bioRxiv* 703710 (2019) doi:10.1101/703710.

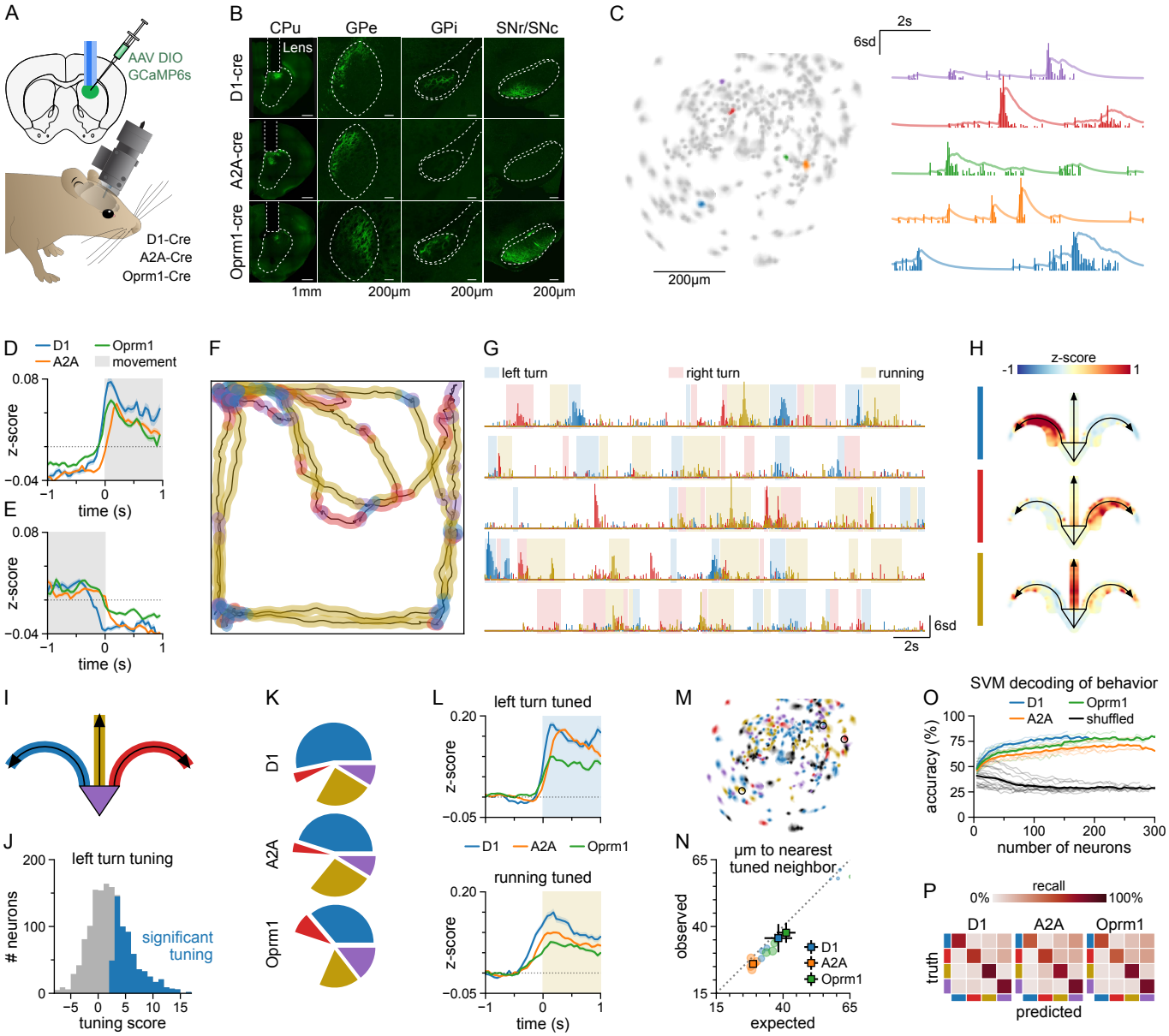
A
AAV DIO
GCaMP6s
D1-Cre
A2A-Cre
Oprm1-Cre
B
CPu
GPe
GPi
SNr/SNc
Lens
D1-cre
A2A-cre
Oprm1-cre
1mm
200μm
200μm
200μm
C
2s
6sd
200μm
D
D1
A2A
Oprm1
movement
z-score
time (s)
E
z-score
time (s)
F
G
left turn
right turn
running
6sd
2s
H
z-score
-1
1
I
J
left turn tuning
neurons
significant tuning
tuning score
K
D1
A2A
Oprm1
L
left turn tuned
z-score
D1
A2A
Oprm1
running tuned
z-score
time (s)
M
N
μm to nearest tuned neighbor
observed
expected
D1
A2A
Oprm1
O
SVM decoding of behavior
D1
A2A
Oprm1
shuffled
accuracy (%)
number of neurons
P
recall
0%
100%
D1
A2A
Oprm1
truth
predicted

Figure 1. The activity of dorsomedial D1+, A2A+, and Oprm1+ SPNs during locomotion.
A) Illustration of experimental approach to image neuron activity in SPNs of the dorsomedial striatum (DMS). A Cre-dependent AAV vector was injected into D1-Cre (direct pathway), A2A-Cre (indirect pathway), or Oprm1-Cre (patch pathway) mice (top). Mice were implanted with a GRIN lens. Calcium signals were recorded with head-mounted miniscopes (bottom).

B) Images show Cre-dependent GCaMP6s expression (green) in caudate putamen (CPu), globus pallidus externa (GPe), globus pallidus interna (GPi), and substantia nigra (pars reticulata and compacta, SNr/SNc).

C) Example of neurons detected in the field-of-view of a single recording session by the CNMF-E algorithm (left map). Calcium signals from the five neurons color-coded in the map. Transparent lines show denoised fluorescence and opaque bars the deconvolved signal.

D) Population average activity aligned to movement onset. Lines show mean, shaded areas ±SEM; n=632 D1-neurons from 7 sessions, 1604 A2A-neurons from 6 sessions and 1465 Oprm1-neurons from 10 sessions.

E) Population average activity aligned to movement offset. Lines show mean, shaded areas ±SEM; n=632 D1-neurons from 7 sessions, 1604 A2A-neurons from 6 sessions and 1465 Oprm1-neurons from 10 sessions.

F) Example showing the behaviors detected in the open field arena (2 minutes, Oprm1+ mouse). The behavior was segmented into left turns, right turns, running bouts and quiescence. Color-code same as in panel G.

G) Deconvolved signals for three selected neurons in the patch pathway (Oprm1+ mouse, behavior shown in F).

H) Heatmaps of the average responses of the three neurons in G during the four segmented behaviors. Arrows represent the three movements, the triangle quiescence.

I) Open field task schematic indicating the color coding of the segmented behaviors ; code used in panels K, M and P.

J) Histogram of the left turn tuning scores of all Oprm1+ neurons (n=1465 from 10 sessions).

K) The proportion of neurons of each primary tuning for all three Cre lines. Color code follows I. n=632 D1-neurons from 7 sessions, 1604 A2A-neurons from 6 sessions and 1465 Oprm1-neurons from 10 sessions

L) Average activity of left turn and running-tuned populations around the onset of the respective behavior. n(left turn) = 514 tuned D1-neurons from 7 sessions, 1324 tuned A2A-neurons from 6 sessions, 1210 tuned Oprm1-neurons from 10 sessions. n(running) = 330 tuned D1-neurons from 7 sessions, 1032 tuned A2A-neurons from 6 sessions, 737 tuned Oprm1-neurons from 10 sessions.

M) Example field of view illustrating the spatial distribution of individual neurons color-coded according to their primary tuning. Oprm1+ mouse, color code follows I.

N) The observed average distance of each neuron significantly tuned to a behavior to its closest neighbor tuned to the same in comparison to the average minimum distance expected by chance (computed based on repeated shuffles of neuron identity). Circles: single session averages, radii proportional to the number of recorded neurons. Squares: average value for each Cre-line, sessions weighted by the number of neurons. Error bars: ±SEM (bootstrapped). N=7 sessions from D1-Cre mice, 6 sessions from A2A-Cre mice, 10 sessions from Oprm1-Cre mice.

O) Decoding accuracy of support vector machines (SVMs) trained to predict the current behavior from neuron activity. Thin lines: single sessions; thick lines: average for each Cre-line. Black lines: behavior labels segment-wise shuffled. N=7 sessions from D1-Cre mice, 6 sessions from A2A-Cre mice, 10 sessions from Oprm1-Cre mice.

P) Average confusion matrix of SVM predictions. All of a session's neurons were used to train the SVM; confusion matrices were averaged with the population size as weights. Behaviors are color coded following I. N=7 sessions from D1-Cre mice, 6 sessions from A2A-Cre mice, 10 sessions from Oprm1-Cre mice.

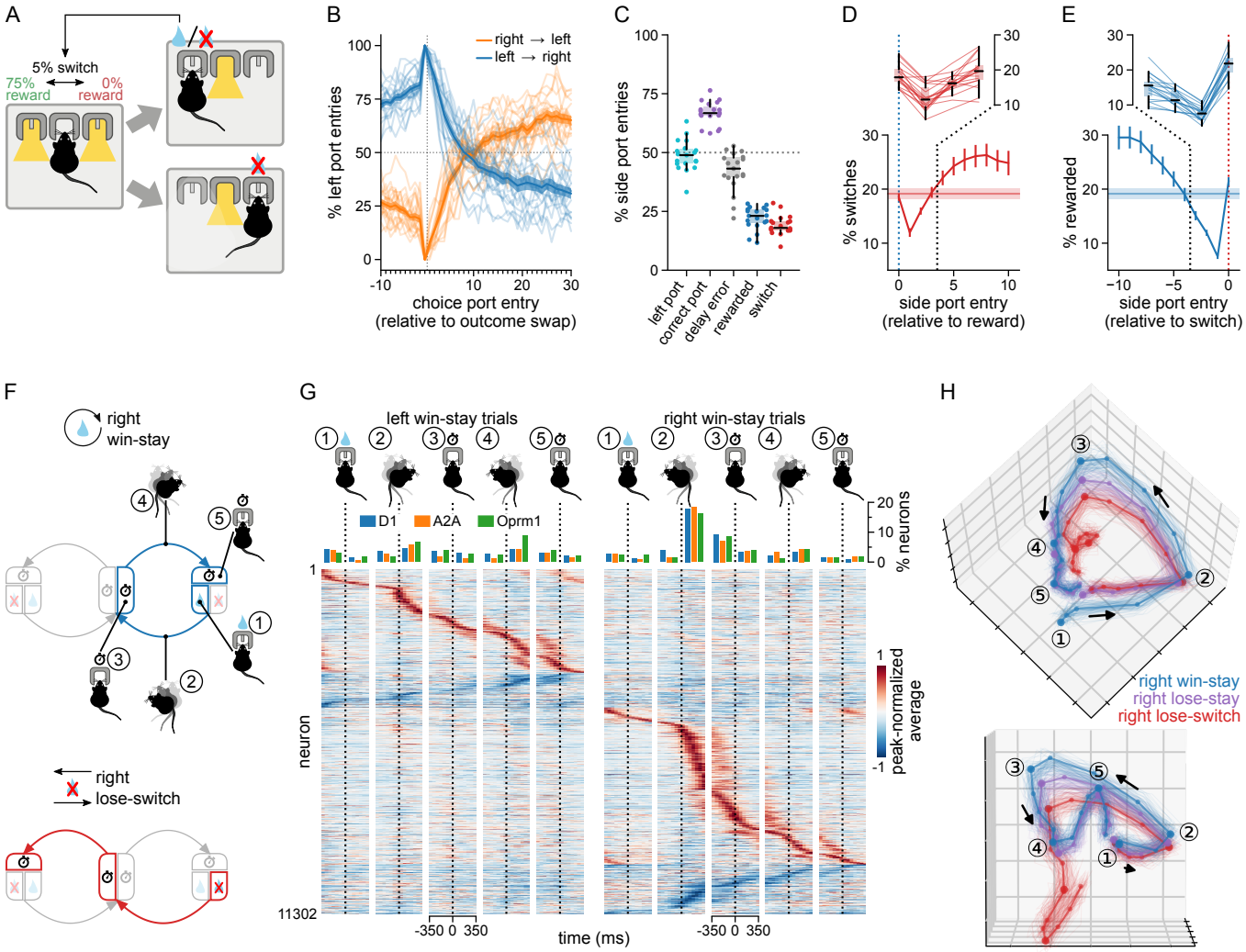
A
5% switch
75% reward
0% reward
B
right → left
left → right
% left port entries
choice port entry
(relative to outcome swap)
C
% side port entries
left port
correct port
delay error
rewarded
switch
D
% switches
side port entry
(relative to reward)
E
% rewarded
side port entry
(relative to switch)
F
right win-stay
right lose-switch
G
left win-stay trials
right win-stay trials
D1
A2A
Oprm1
% neurons
neuron
peak-normalized average
time (ms)
H
right win-stay
right lose-stay
right lose-switch

Figure 2. The D1+, A2A+, and Oprm1+ SPNs represent the action space of a choice task.
A) Illustration of the 2-choice switching task. Mice initiated a trial by poking the central port, followed by a choice for one of the side ports. Only one side port delivered sucrose rewards (reward probability of 75%). 5% of reward deliveries were followed by a covert switch of reward delivery side. To initiate trials or receive rewards, mice had to stay a minimum of 350 ms in the respective ports.

B) Mice successfully shifted their port preference after covert switches of the rewarded port. A switch of the reward from the right to the left port was followed by a gradual increase in the fraction of left port entries on subsequent trials (red line), a left to right switch vice versa (blue line). Reward port switches only occured after rewarded choices. n = 3167 switches. Thin transparent lines: individual mice.

C) Key statistics of the behavior. *Correct port*: entry into the port currently assigned the reward. *Delay error*: failure to trigger a choice port due to premature withdrawal (<350ms) from same or the initiation port. *Switch*: choice of the port not selected last. n=299083 trials from 271 sessions from 19 mice. Transparent dots: individual mice.

D) Win-stay behavior: The probability of a choice switch was lowest immediately after a rewarded choice (side port entry 0). Horizontal line: overall average fraction of switch choice port entries. Thin transparent lines: individual mice.

E) Lose-switch behavior: Mice were least likely to obtain rewards (i.e. were most likely to lose) on choice port entries preceding a switch choice (side port entry 0). Horizontal line: overall fraction of rewarded choice port entries. Thin transparent lines: individual mice.

F) Illustration of "win-stay" (top) and "lose-switch" (bottom) trials after right side port outcomes. The outcome phase was treated as the start of the trial to reflect the dependence of a stay or switch choice on the outcome obtained last; a win-stay trial starts with a win (reward), a lose-switch trial with a loss (no reward). The task space was defined by twelve trial phases. Rectangles correspond to time spent in the center or outcome ports (left, right). Arrows represent movements between the ports. The center port is split into two phases according to the direction of the upcoming choice. Side ports are subdivided into the delay period (clock in upper rectangle), and reward (drop, inner part) and omission (x, outer part) phases.

G) The average response of neurons to the phases of left and right port win-stay trials, centered on the phase onset times (time 0, window: ±350 ms). Neurons were peak-sorted. The bar charts indicate the percentage of neurons of each pathway (color-coded) with a positive peak falling within the 350 ms time window in the raster below. Pooled data from all recordings.

H) Pseudo-trial activity trajectories based on single trial data pooled over all neurons recorded. The trajectories separate by trial type (color-coded), appearing ordered according to value implicit in the strategic choice made (win-stay > lose-stay > lose-switch). Trials were binned in scaled time (4 bins per phase). Trials were drawn in random order for every neuron. D1+ n=376; A2A+ n=1663; Oprm1+ n=1864 neurons; from 5, 5, 16 sessions, respectively. Mean trajectories of the pseudo-trial data used for dimensionality reduction (PCA; thick lines) and resampled pseudo-trials (thin lines).

Error bars and shading: ±SEM. Boxplot whisker range: upper (lower) quartile to highest (lowest) value within 1.5 IQR

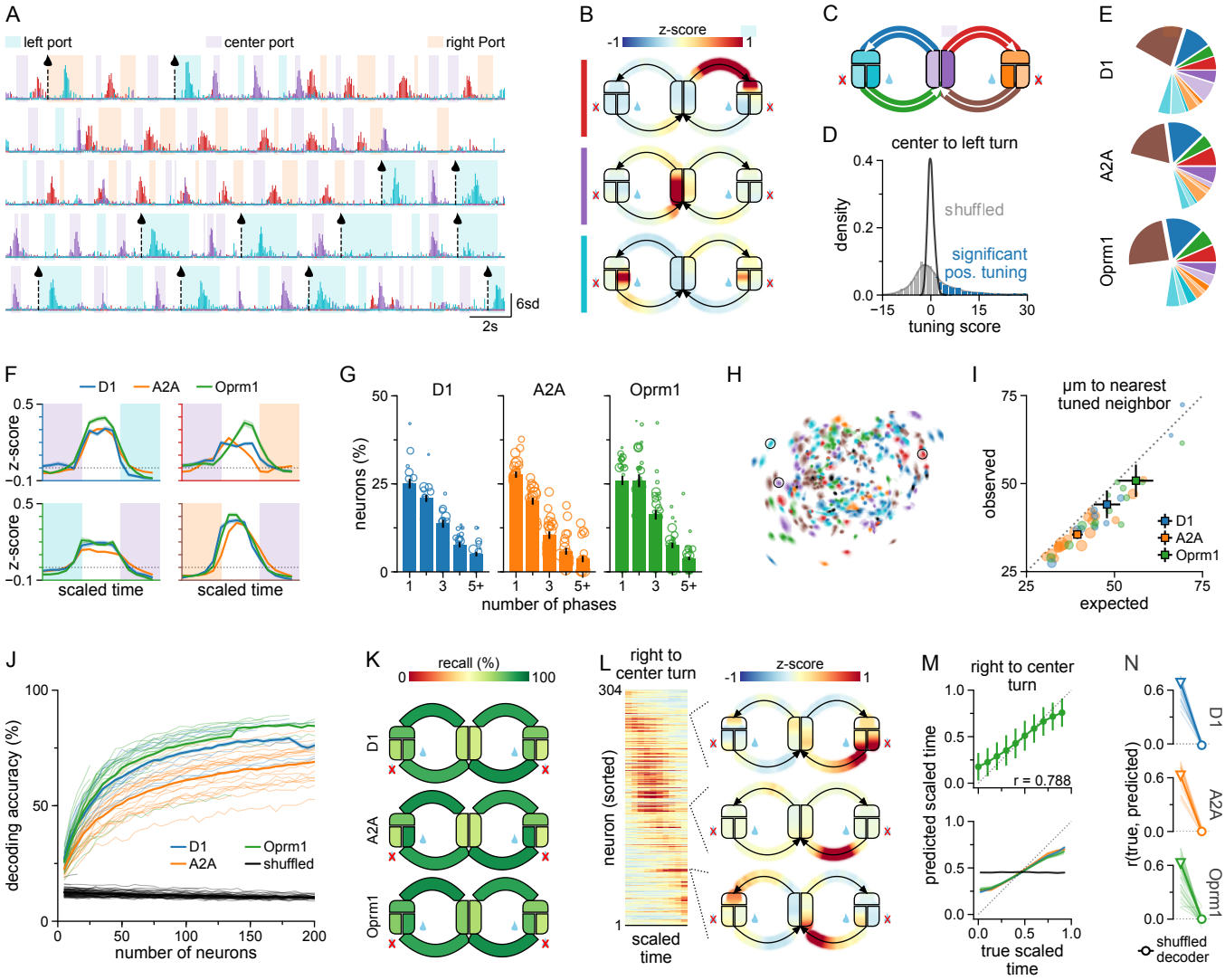
A
left port
center port
right Port
6sd
2s
B
z-score
-1
1
C
D
center to left turn
density
shuffled
significant
pos. tuning
tuning score
E
D1
A2A
Oprm1
F
D1
A2A
Oprm1
z-score
scaled time
G
D1
A2A
Oprm1
neurons (%)
number of phases
H
I
µm to nearest
tuned neighbor
observed
expected
D1
A2A
Oprm1
J
decoding accuracy (%)
number of neurons
D1
A2A
Oprm1
shuffled
K
recall (%)
D1
A2A
Oprm1
L
right to
center turn
neuron (sorted)
scaled
time
z-score
M
right to center
turn
predicted scaled time
r = 0.788
true scaled
time
N
r(true, predicted)
D1
A2A
Oprm1
shuffled
decoder

Figure 3. The D1+, A2A+, and Oprm1+ SPNs share a similar representation of the trial phases.
A) Deconvolved traces from three example neurons, respectively tuned to the center-to-left turn (red), the center port delay phase which precedes a left turn (purple), and left port reward delivery. Shading indicates port occupancy. Black dashed lines and drop pictograms mark reward delivery.

B) The average activity of three example neurons in every trial phase over the entire recording session plotted as heatmaps structured according to the 2-choice task schematic. The neurons are the same as in panel A; colorbars indicate correspondence. Trial-level phase activity scaled to uniform length.

C) Illustration of the 2-choice task schematic divided into twelve phases. Reference for phase color-coding in panels E, F and G.

D) Histogram of the center-to-left movement tuning scores of all Oprm1+ neurons (n=2793). The tuning scores were computed by z-scoring neurons' mean activity with the mean and standard deviation of a sampling distribution obtained using block-wise shuffled behavior data. Significance: mean activity above the 99.5th percentile.

E) Pie charts proportioned according to neurons' primary phase-tuning, per Cre-line. Colors follow C. n=1943 D1-neurons, 6566 A2A-neurons, 2793 Oprm1-neurons.

F) Average activity of 4 of the 12 populations defined by phase-tuning (the movements) during the phase they are tuned to. Phase-tuning indicated by the color of the axes; colors as in C. Background as in A. n(center-to-left neurons) = 413/1517/751; n(center-to-right neurons) = 279/928/426; n(left-to-center neurons)=228/883/424; n(right-to-center neurons)=581/1785/996 (D1/A2A/Oprm1).

G) Bar plots of the weighted mean percentage of neurons significantly tuned to several counts of phases, by Cre-line. Sessions were weighted by the number of neurons. Circles: single sessions, radii proportional to number of neurons. Error bars: ±SEM (bootstrapped).

H) Oprm1+ single-cell ROIs for an example session, color-coded by primary tuning. Black: no significant tuning. Circles mark ROIs whose activity is shown in A and B.

I) Similarly tuned neurons are marginally closer in space than expected by chance (N=16 D1-sessions, 19 A2A-sessions, 30 Oprm1-sessions).

J) Decoding accuracy (fraction correctly predicted phases) for randomly sampled SPN sub-populations of increasing size. Thin lines: single sessions. Thick lines: average per Cre-line. N=16 D1-sessions, 19 A2A-sessions, 30 Oprm1-sessions.

K) The SVM decoding recall by phase when trained on the full population of each session. The average recall was weighted by the number of neurons in each session.

L) Sequential neuron activity through a task phase. Raster plot of the average activity for all neurons recorded in an example session (Oprm1+, right port to center port movement). Neurons sorted by the peak of the activity (left panel). Heatmaps show three example neurons across the entire task space in the same session (right panel).

M) The trial phase substructure can be decoded from neuron activity. Top panel: the actual versus the decoded time progression through the phase (right port to center port turn) for the session shown in L. Bottom panel: the actual versus decoded progress through the phase averaged over all sessions. Individual sessions weighted by the number of recorded neurons; Cre lines color coded as in N. Black line: decoder trained on shuffled data. Shaded area: ±SEM (bootstrapped).

N) The correlation between true and predicted progress through right-to-center turns for all sessions (thin lines), compared to the correlation from shuffled data. Thick lines indicate averages, weighted by the number of recorded neurons in each session. Error bars: ±SEM (bootstrapped).

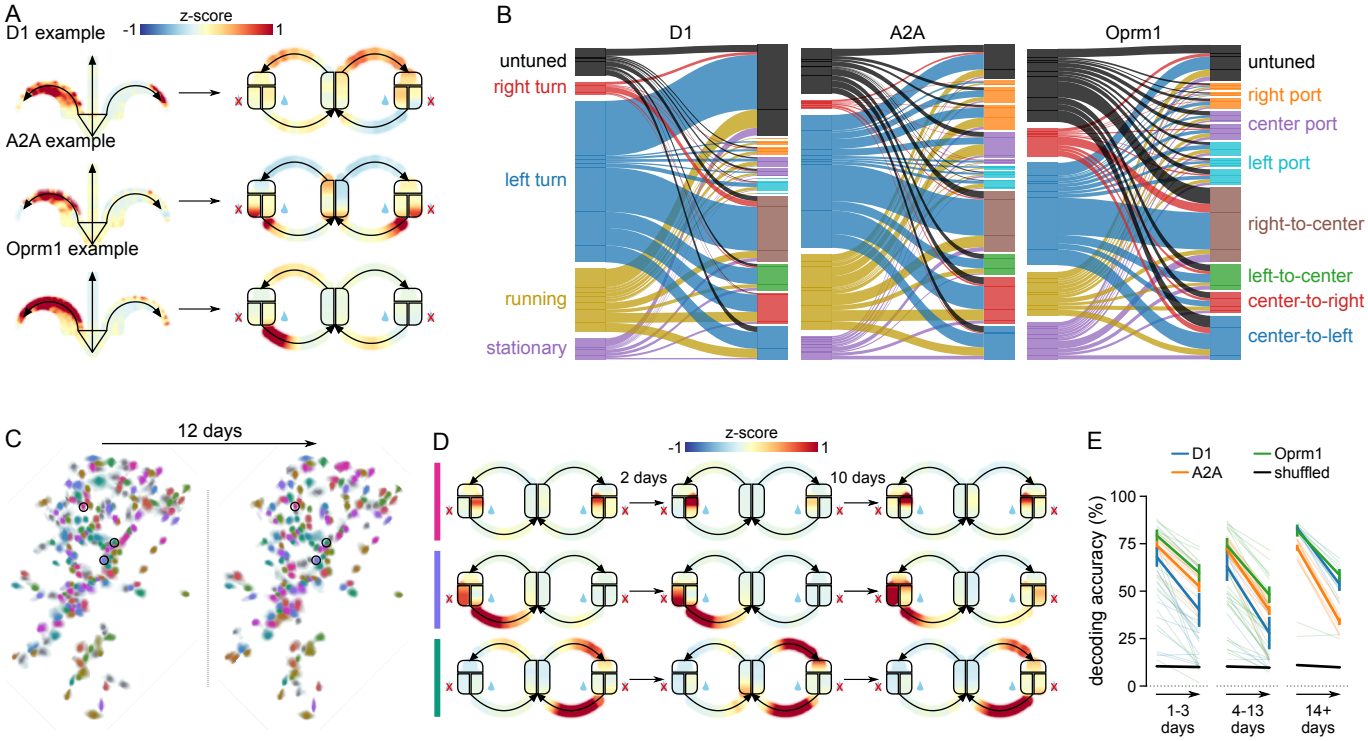
A
D1 example
z-score
-1
1
A2A example
Oprm1 example
B
D1
A2A
Oprm1
untuned
right turn
left turn
running
stationary
untuned
right port
center port
left port
right-to-center
left-to-center
center-to-right
center-to-left
C
12 days
D
z-score
-1
1
2 days
10 days
E
D1
A2A
Oprm1
shuffled
decoding accuracy (%)
100
75
50
25
0
1-3 days
4-13 days
14+ days

Figure 4. The phase tuning of D1+, A2A+, and Oprm1+ SPNs is task-specific and highly conserved across days.

A) Examples of how individual left-turn tuned neurons (left) change their tuning in the 2-choice task (right).

B) Alluvial plot showing how the primary tunings in the open field (left side) change in the 2-choice task (right side). Neurons in A and B were recorded without detaching the miniscope; the comparisons did not necessitate the matching of spatial filters between sessions.
n=406 D1-neurons, 1111 A2A-neurons and 985 Oprm1-neurons were followed.

C) Spatial filter map of an example recording session (Oprm1-Cre) and aligned filters from a session recorded 12 days later. The depicted filters were transformed by the registration process. Colors are chosen arbitrarily, but consistent for aligned neurons. Gray filters are without a match.

D) Trial heatmap of 3 neurons matched across 3 sessions, including those shown in C. The colored bars match the color of the respective, circled filters in C.

E) SVMs decode the trial phase across sessions (color code shows pathway). Each pair of sessions from a single mouse is shown with a transparent line (thickness proportional to the number of aligned neurons for that pair). Opaque lines correspond to the average decoding accuracy across pairs of sessions, weighted by the number of aligned neurons. SVM decoding accuracy on shuffled data (black). N(1-3 days) = 65/55/85 session pairs; N(4-13 days) = 25/60/85 session pairs; N(14+ days) = 25/30/25 session pairs (D1/A2A/Oprm1). Error bars: ±SEM (bootstrapped).

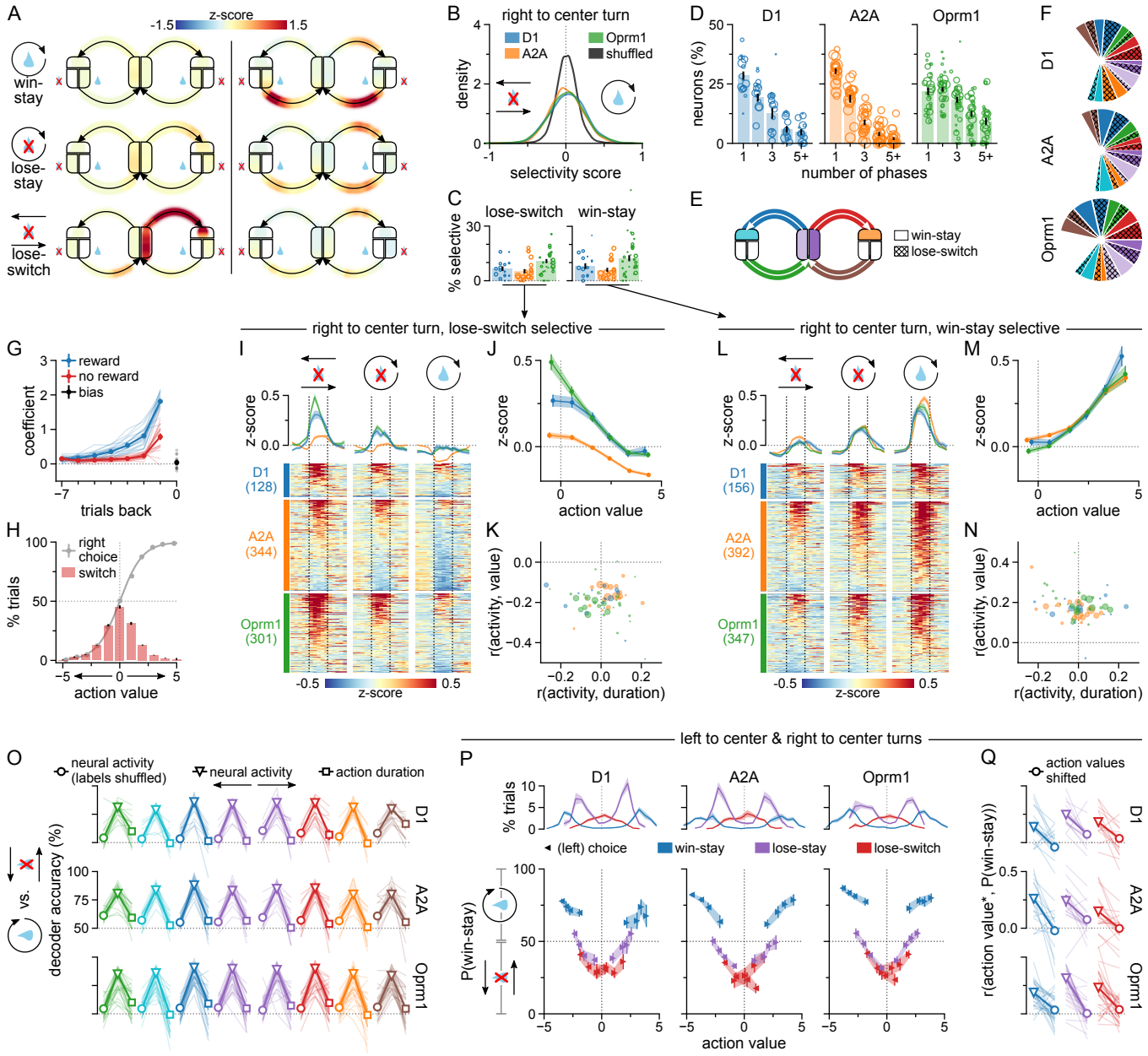
A
z-score
-1.5
1.5
win-stay
lose-stay
lose-switch
B
right to center turn
D1
A2A
Oprm1
shuffled
density
selectivity score
C
lose-switch
win-stay
% selective
D
D1
A2A
Oprm1
neurons (%)
number of phases
E
win-stay
lose-switch
F
D1
A2A
Oprm1
right to center turn, lose-switch selective
right to center turn, win-stay selective
G
reward
no reward
bias
coefficient
trials back
H
right choice
switch
% trials
action value
I
z-score
D1 (128)
A2A (344)
Oprm1 (301)
J
action value
K
r(activity, value)
r(activity, duration)
L
D1 (156)
A2A (392)
Oprm1 (347)
M
N
left to center & right to center turns
O
neural activity (labels shuffled)
neural activity
action duration
vs.
decoder accuracy (%)
P
% trials
(left) choice
win-stay
lose-stay
lose-switch
P(win-stay)
action value
Q
action values shifted
r(action value*, P(win-stay))

Figure 5. The D1+, A2A+, and Oprm1+ pathways represent all aspects of the choice task structure.
A) Example of SPNs with phase-specific evaluation for stay or switch choice. Example Oprm1+ neuron (left side) shows elevated activity in the center port preceding right port choices, as well as during the choice action, only in the context of left-to-right lose-switch trials when the value of the left choice is low. Example Oprm1+ neuron (right side) shows increased activity when returning from the choice port to the center in win-stay trials, when the value of the previous choice is high. Note that it also responds, to a lesser degree, during return movements in lose-stay trials, when the last choice is of intermediate value due to rewards received on trials prior to the recent loss (see G).

B) Distributions of the phase-specific lose-switch versus win-stay selectivity scores for neurons (color code shows pathway). Neuron activity for win-stay trials (positive scores) versus lose-switch trials (negative scores). Selectivity scores: area under the ROC curve, scaled from -1 to 1. Shuffled distribution (black). (n=1943 D1+ neurons, 6566 A2A+ neurons, 2793 Oprm1+ neurons).

C) The weighted mean percentage of neurons significantly selective for lose-switch (left) and win-stay (right) trials during the right port to center port phase (circles: single sessions, radii proportional to the number of neurons, error bars: ±SEM bootstrapped).

D) Bar plots of the weighted mean percentage of neurons significantly win-stay or lose-switch selective to several counts of phases. Sessions were weighted by the number of neurons (circles: single sessions, radii proportional to the number of neurons, error bars: ±SEM bootstrapped).

E) Illustration of the choice task with phase color-coding used in panels F and O.

F) The primary phase-specific win-stay or lose-switch selectivity of SPNs. (n=1943 D1-neurons, 6566 A2A-neurons, 2793 Oprm1-neurons).

G) Logistic regression models the impact of the outcomes of the previous seven trials on the current port choice. Mice repeated recently rewarded choices (blue), and were likely to repeat the previous choice even after one reward omission (red). Bias coefficient (black): positive values indicate right port bias. Logistic regression coefficients for individual mice (thin lines) and the mean coefficients (thick lines).

H) Average fraction of right port choices (grey dots) and choice switches (red bar plots) as a function of relative action value (binned). Grey curve: logistic regression prediction (data pooled over all mice). Positive (negative) action values predict right (left) port choices. Action values are the sum of coefficients shown in G. Outcome history dictates the coefficient (reward or no reward) used for each of the previous seven trials, side choice its sign (right: positive, left: negative). Dots and bars: mouse average (N=19 mice). Error bars:±SEM.

I) Scaled time raster plots of the average activity of the lose-switch selective neurons in lose-switch, lose-stay and win-stay trials. The dotted lines mark the start and the end of the right port to center port motion. The line plots above the rasters show the population mean activity.

J) The mean trial activity of lose-switch selective neurons at various action values (average neuron activity, binned).

K) The session-averaged Pearson correlations for lose-switch selective neurons to movement duration versus to action value. Radii reflect the number of selective neurons in the session.

L) Scaled time raster plots of the average activity of the lose-switch selective neurons in lose-switch, lose-stay and win-stay trials. The dotted lines mark the start and the end of the right port to center port motion. The line plots above the rasters show the population mean activity.

M) The mean trial activity of win-stay selective neurons at various action values (average neuron activity, binned).

N) The session-averaged Pearson correlations for win-stay selective neurons.

O) Phase-specific decoding accuracy of predicting win-stay versus lose-switch trial type from neuron activity using cross-validated linear SVMs. For each trial phase SVMs were trained on data with the behavior labels shuffled (circles), on the neuron activity data (triangles), or only on phase duration (squares). Accuracy for individual sessions (thin lines) and mean accuracy (thick lines). Sessions weighted by the number of neurons (N=16 D1-sessions, 19 A2A-sessions, 30 Oprm1-sessions).

P) The SVMs probability estimates for outcome port (left or right side) to center port turns being win-stay rather than lose-switch as a function of action value. Action value correlates with the SVM prediction confidence. Win-stay trial turns are more confidently predicted in high action value trials. (blue). Lose-switch trial turns are more confidently predicted in low action value trials (red). Lose-stay trial turns with low action value (magenta) appear increasingly like lose-switch turns. Correlation scores are quantified in Q. Binning by quartiles of the pooled action value distributions. Triangles: mean probability estimate of sessions weighted by number of neurons. Shading and error bars: ±SEM (bootstrapped). Top panel: density of each trial type by action value; average of sessions ±SEM.

Q) Correlation between action value and the estimated probability of choice port to center turns being win-stay rather than lose-switch (action value sign reversed for left choice trials). Correlation of individual sessions (thin lines) and average correlation of sessions weighted by the number of neurons (thick lines). Color-coding as in P.

METHODS

Animals

The 19 adult female and male transgenic mice (25–35 g) used in experiments were kept on a 12-hour light/dark cycle. We used D1-Cre (Drd1-cre EY262Gsat), A2A-Cre (Adora2a-cre KG139Gsat), and Oprm1-Cre mice. Mice were single housed after the surgeries and placed on food restriction during behavioral training and recording in the 2-choice task (maintained at min. 85% of their free-feeding body weight). All procedures were approved by the Swedish local ethics committee for animal experiments (Stockholms djurförsöksetiska nämnd, approval N166/15).

Surgeries

For stereotactic surgery, the mice were anesthetized with isoflurane (2% in air) and administered buprenorphine (0.03 mg/kg, subcutaneous) for analgesia. Buprenorphine was administered for post-surgery pain-relief. A feedback-controlled heating pad maintained body temperatures at 36°C throughout the surgical procedures. All animals were unilaterally microinjected with 400 nl of AAV5-CAG-Flex-GCaMP6s into the dorsomedial striatum in the right hemisphere (AP: 1.0, ML: 1.25, DV: -2.3; rate of 100 nl/min). The pipette was retracted 5 min after the injection finished. Two weeks after the viral injection, gradient-index (GRIN) endoscope lenses (Inscopix) of a 1 mm diameter were implanted 100-200 μm above the viral injection site. Prior to lowering the lens, portions of the overlaying cortex were aspirated using a 0.5 mm-diameter, blunt-point needle with sharpened edges which was attached to a vacuum pump. The GRIN lens was fixed in place using dental cement. 4-6 weeks after GRIN lens implantation, the baseplate was anchored to the skull with dental cement to support the detachable miniscope (Inscopix) during imaging. To determine the optimal placement of the baseplate, the field of view was monitored live throughout the procedure using the baseplate-attached miniscope. The miniscope was not refocused over the course of the experiments in order to improve tracking of individual neurons across days.

Open field behavior

We tracked the mice in the open field (49×49 cm) using DeepLabCut (DLC; Mathis, 2018). We manually labeled the base of the tail, the center of the body and the left and right ears in 1000 frames, sampled from all sessions. After running DLC on every open field video, we transformed the video coordinates (in pixels) to world coordinates (in cm) using a perspective transform matching the four corners of the box.

We based our movement analysis on two markers tracked by DLC: the base of the tail and the center of the body. We used a custom particle filter (2000 particles; diagonal gaussian observation noise with σ=1 cm; diagonal gaussian innovation noise with σ=1 cm/frame; gaussian likelihood penalty on distance between tail base and center of body with μ=3 cm and σ=1 cm; *systematic*

resampling as defined in Doucet & Johansen (2011, p. 13)) to estimate smoothed x- and y-position as well as running speed, allocentric body direction, angular speed and body elongation.

We then used a custom optimization algorithm to classify each video frame as either *left turn, right turn, running* or *stationary*. First, we chose a scoring function $P_b(s,e)$ that defined how well a segment starting at frame s and ending at frame e would fit behavior b. In particular, we chose

$$P_{\text{left turn}}(s,e) = \begin{cases} -10 & \text{if } e-s<5 \\ -10 & \text{if } a(s,e)<e-s \\ -10 & \text{if } d(s,e)>10(e-s) \\ a(s,e)-0.5(e-s)-10 & \text{otherwise} \end{cases}$$

$$P_{\text{right turn}}(s,e) = \begin{cases} -10 & \text{if } e-s<5 \\ -10 & \text{if } a(s,e)>e-s \\ -10 & \text{if } d(s,e)>10(e-s) \\ -a(s,e)-0.5(e-s)-10 & \text{otherwise} \end{cases}$$

$$P_{\text{running}}(s,e) = \begin{cases} -10 & \text{if } e-s<5 \\ -10 & \text{if } |a(s,e)|>e-s \\ 5d(s,e)-10 & \text{otherwise} \end{cases}$$

$$P_{\text{stationary}}(s,e) = -5$$

where $a(s, e)$ is the signed angle difference between the animal in frame s and frame e measured in degrees, and $d(s, e)$ is the distance in cm between the animal's position in frame s and frame e. Note that both $a(s, e)$ and $d(s, e)$ can be computed in constant time. The negative terms acted as a cost on the number of segments to encourage the segments to be as long as possible. Having this definition, we wanted to find segments of behaviors such that the total score was maximized, i.e.

$$\underset{a_i,b_i,N_{\text{seg}}}{\textit{maximize}} \quad \sum_{i=1}^{N_{\text{seg}}} P_{b_i}(a_i, a_{i+1}-1)$$

where N_{seg} is the number of segments, a_i is the start of segment i and b_i is the behavior of segment i ($a_1=1$, $a_i<a_{i+1}$, a_N=number of frames). The optimal $\{N_{seg}, a_i, b_i\}$ can be found very efficiently with dynamic programming on the recurrence relation:

$$S_t = \max_{k,b} S_{t-k-1} + P_b(t-k,t)$$

To improve running time further, we constrained $k<2000$ so that the algorithm ran in linear time.

2-choice task

We trained the food-restricted mice to perform a self-paced, probabilistic 2-choice task inside custom-built operant chambers which contained three nose poke ports. The mice initiated a choice in the center port, and were then free to choose one of the two side ports. One side port yielded sucrose solution with 75% probability upon entry (15% sucrose, 3.75 µl), while the other yielded nothing. After a reward, there was a 5% probability of a reward port switch. When no reward was delivered (due to failed trial initiation, incorrect port choice or reward omission) the rewarded side remained the same. There were no cues to indicate the correct port choice nor the occurrence of a reward port switch. The volume of the sucrose reward was doubled in a random 10% of the rewarded trials. LEDs located in the ports indicated whether a trial was in the initiation phase (center LED on) or in the choice phase (side LEDs on). A 350 ms infrared beam break was required to trigger any port. The task was controlled and task events recorded using a pyboard microcontroller.

The effect of trial outcome history on port choice, used to compute trial-by-trial relative action value, was estimated by logistic regressions, as previously reported (Tai et al., 2012). Animals failed to trigger the reward ports in almost half of their attempts, predominantly due to having failed to wait out the delay in the center port first, exiting it early (<350 ms; see Figure 2C, delay error). This suggests they missed or disregarded the LED cues indicating whether the initiation nose poke was successful and remained unaware of their error. Thus, we included every side port entry in the analysis, irrespective of whether or not the choice was properly initiated in the center port.

Imaging data acquisition and preprocessing

Raw calcium imaging videos (1440×1080, 20 fps) were acquired using miniscopes and the nVista Acquisition Software (1.2.0, Inscopix). To accurately align calcium transients and behavior, we recorded the sync pulses provided by the miniscope's data acquisition box in the output of the microcontroller controlling the operant task. The sync pulses also served to trigger a Blackfly USB3 video camera (FLIR) used to record both the operant and the open field experiments. 2-choice task imaging sessions lasted approximately 1 hour, open field sessions 20 min (see table S?). Calcium imaging videos were spatially downsampled (2×2 pixel bins), cropped, and motion corrected using the Inscopix Data Processing Software and exported as Neurodata Without Borders (NWB) files. We used the CaImAn (1.4.2; Giovannucci, 2019) implementation of the CNMF-E algorithm (Zhou, 2018) to extract spatial filters and deconvolved fluorescence traces of individual neurons. Imaging frames were downsampled (2×2 pixel bins) once more prior to running the algorithm. To adjust for bleaching, the rolling z-score (±5 min window) of the deconvolved activity traces was computed and used in all subsequent analysis. To match individual neurons across days, we used CaImAn's ROI-registration function.

Definition of trial phases

Each calcium imaging frame was assigned to a trial phase solely based on the infrared beam brakes detected around that frame. Frames recorded while a particular port's beam was broken were registered as occupations of that port (i.e. nose pokes). When no beam was broken, frames were classified as movements between the port occupied last and the port entered next. We excluded thus labeled movements bypassing the center port (side port to side port), movements lasting longer than 1.5s (animal unengaged), as well as movements starting and ending in the same port (reentries), from further analysis. Nose pokes preceding unengaged periods and reentries were also excluded. We divided center port nose pokes into two different phases, depending on which side port was entered next (upcoming left or right choice). Side port nose pokes were split into three phases: the *delay* phase preceding the outcome presentation (<350ms in port) and the two mutually exclusive outcome phases (>350ms in port), *omission* and *reward*. Note that this classification approach disregards whether ports were triggered or not, i.e. nose pokes shorter than 350 ms and nose pokes addressing inactive ports (due to a delay error in the previous nose poke) are included in the analysis (cf. Behavior).
For drawing trial heatmaps (e.g. Figure 3B) and for sub-phase decoding (Figure 3L-N), we assigned a *progress score* to each frame, defined as the count of consecutive previous frames spent in the current phase, divided by the total number of frames in the current phase. Note that this scaling of the progress was therefore based on *time* between beam breaks, not on physical space.

Pooled-population activity trajectories

The activity trajectories in figure 2H are based on single-trial activity pooled over neurons of all Cre lines and recording sessions (without matching neurons across sessions). Individual trials in which the animal did not follow the task structure perfectly, e.g. by reentering a port or disengaging from the task mid-way, were excluded from this analysis, as were trials in which any phase lasted fewer than 250 ms (5 frames). Whole sessions were excluded if there were fewer than 20 trials left of any of the three trial types visualized (right win-stay, right lose-stay and right lose-switch). To pool the single trial data from different sessions into pooled-population pseudo-trials, we contracted all trials to uniform length by binning the trial phases in scaled time (4 bins per phase, resulting in a total of 20 data points per trial). For every trial type, we then randomly sampled trials of that type from every neuron in every session, combining them into pseudo-trials of pooled-population activity. Trials were drawn without replacement; the total number of pseudo trials of a specific type was therefore determined by the minimum number of trials of that type performed in any session. We applied PCA to one pooled data set obtained in this way, treating the neurons as features to be reduced and the time points of the concatenated pseudo-trials as samples. We then plotted the average trial trajectory using the first three principal component scores (thick lines in 2H). To evaluate the reproducibility of this approach, we resampled the pooled-population data set several times, obtaining different subsets and combinations of single-trial data, and projected these pseudo-trials into the same PCA space (thin lines).

Note that the random drawing of trials forced the neurons in the pseudo-session to be conditionally independent given the trial phase and type, and therefore effectively canceled out any principal components that were not aligned to our parameterization of the task. This was done on purpose to enhance the task-relevant structure in the neural activity and we term the principal components calculated this way "task-related principal components".

Visualization of trial heatmap

To create the trial heatmaps we first assigned a trial phase and a phase progress (0 to 1) to each calcium imaging frame. Next, we used these to calculate a pixel coordinate in the schematic (full image was 501×251 pixels). For the center port, the y-coordinate increased with increased phase progress, such that 0 was at the bottom and 1 was at the top of the rectangle. Two separate x-coordinates were used depending on the upcoming choice. For the side ports, the y-axis was reversed compared to the center port so that increasing phase progress was going downwards. The scaling was such that the beginning of the delay phase always was at the top and the end of the delay phase was at the indicated line. The reward and omission phases similarly progress downwards. The omission phase is shorter than the reward phase, but scaled to have the same length in the visualization.

For leftwards movements, i.e. center-to-left and right-to-center, the x-coordinate was linearly decreasing with phase progress. For rightward movements, i.e. center-to-right and left-to-center, it was linearly increasing with phase progress. The y-coordinate of the four movements was the square of the progress, followed by a translation and scaling. The phase progress therefore does not follow the tangent of the parabola between the respective ports but rather the x-coordinate.

We then calculated the average deconvolved z-scored activity for each pixel and smoothed the averages with a gaussian kernel ($\sigma = 7$ pixels). Finally, we applied the colormap to the smoothed value at each pixel. We set the color according to the mean activity and the transparency according to how many times the (smoothed) pixel was visited.

Calculation of neuron tuning scores

We created surrogate data by first dividing the session into blocks of consecutive left and right trials and then shuffling these blocks. This procedure meant most of the behavioral statistics of the session were preserved (number of left and right trials, stay and switch probabilities, time between trials and between phases within a trial, etc), while the connection to the neural activity was broken. To calculate the tuning score, we calculated the average z-scored fluorescence during each of the 12 trial phases. As a comparison, we did the same for 1000 iterations of shuffled blocks as described above. We classified a neuron as *positively tuned* to a trial phase if the average activity using the real data was larger than in 99.5% of the shuffled iterations or, reversely, as *negatively tuned* if it was less than 99.5% of the shuffled iterations. In addition to these discrete classifications, we also defined a *tuning score* as the real average minus the grand average of the shuffled iterations and

divided by the standard deviation of the averages of the shuffled iterations. We call the 12 tuning scores for a given neuron the *tuning profile* of that neuron.
For the open field, we used a similar procedure to define positively and negatively tuned neurons, as well as tuning scores, to the four behaviors. The shuffle was taken over the behavioral segments (left turn, right turn, running, stationary) without consideration of transition probabilities.
To determine how well a neuron's activity discriminated win-stay from lose-switch trials in the choice task, we computed a *selectivity score* based on receiver operating characteristic (ROC) analysis. The selectivity score was defined as the area under the ROC curve (AUC), scaled from -1 to 1, i.e. equaled 2 x (AUC - 0.5). To establish whether a neuron's selectivity score was statistically significant, we opted for a similar approach as described for the phase tuning scores above. For every neuron, we computed 1000 selectivity scores based on shuffled behavior data for comparison with the actual selectivity score. If the actual score was higher than 99.5% of the shuffled data scores, we classified the neuron as significantly win-stay tuned, if instead it was lower than 99.5% of the latter, we designated it significantly lose-switch tuned.

Support vector machines to decode behavior from neuron activity

We z-scored the deconvolved fluorescence traces across the full session for each neuron. For every phase of each individual trial, we calculated the average z-scored deconvolved fluorescence for all the neurons. Using the trial phase as the label with the vector of average fluorescences as corresponding features, we trained support vector machines (SVMs) with linear kernels and slack C=1 (sklearn.svm.SVC). To handle multiclass labels (12 different trial phases), we employed the *one-versus-rest* strategy.

In Figure 3J, we randomly selected a subset of neurons and randomly divided 80% of the trial phases into a training set and 20% into a test set. We repeated this procedure 10 times for each session and number of selected neurons (steps of 5 neurons) and report the average accuracy of these 10 iterations. Accuracy was measured as the fraction of trial phases correctly predicted without accounting for the frequency of each label. Note that this is not traditional 5-fold cross-validation because the train and test split (as well as the selection of neurons) is redrawn in each iteration. In the open field (Figure 1O) we used a similar procedure, but with segments of behavior rather than trial phases.

In Figure 3K, we did five iterations of 80%-20% train-test splits for each session and used all the available neurons in each iteration. We calculated the fraction of each phase type correctly recalled for each session and averaged these values weighted by the number of neurons in the respective session. In Figure 1P we used the same principle to calculate confusion matrices.

When decoding across days (Figure 4E) we only included matched neurons. For all pairs of sessions from the same animal, we trained five iterations of support vector machines on 80% of the trial phases from the session with the earliest recording date. We tested the SVM from each

iteration on the remaining 20% of the same session (*same-day-decoding*) as well as on all the trial phases from the other session (*across-days-decoding*). We calculated the average accuracy for each pair of sessions. In Figure 4E, each line corresponds to one such *pair*. Note that this implies the same session is included multiple times. For each genotype and bin of days, the thick lines indicate the average accuracy over all pairs, weighted by the number of matched neurons.

For decoding lose-switch versus win-stay, we averaged the deconvolved and z-scored traces for each phase in each trial, as described above. We then discarded side-port phases and analyzed the remaining six trial phases separately. For each type of trial phase, we discarded all win-switch or lose-stay trials and trained separate SVMs to predict whether the individual phases belonged to a win-stay or a lose-switch trial. Accuracies are reported as the fraction of trial phases that were correctly predicted in a 5-fold cross-validation schema. For comparison, for each decoding we also repeated the respective procedure with one instance of shuffled data created with the same algorithm as when calculating the tuning scores. For Figure 5O, we also compared each SVM trained on neural data to an SVM trained on a single feature, namely the duration of the trial phase that is to be predicted. If the difference in neural activity was only due to different vigor of the movements, these SVMs would reach the same or better performance as the ones trained on neural data.

To see how the SVM decoding varied with the action value, we took the SVMs trained on lose-switch and win-stay trials applied them to all three types of trials. Note that these SVMs never saw any lose-stay trials in the training phase and therefore can never predict lose-stay. Instead of using the binary classification labels, we used the probability they assigned to win-stay over lose-switch for each trial. In Fig. 5P, we binned trials by trial type and action value (quartile bins) and showed the mean probability outputted by all SVMs for the respective bin. We also correlated this probability (without binning) with the action value of each trial separately for each trial type. For comparison, we calculated the same correlation on data where the action value was either shifted forward or backward 10-30 trials (with wrap-around). This preserved the trial-by-trial structure in the action value but dissociated it with the actual neural activity.

Analysis software

In addition to the explicitly mentioned tools, we used Python 3.7, with numpy 1.17.3 (van der Walt et al, 2011), scipy 1.3.0 (Virtanen et al, 2020), pandas 0.25.3 (McKinney, 2010) throughout. The support vector machines and other machine learning tools were from scikit-learn 0.21.3 (Pedregosa et al, 2011). Plots were rendered with the matplotlib 3.1.1 (Hunter, 2007), seaborn 0.9.0 and figurefirst 0.0.6 (Lindsay et al, 2017) plotting libraries. Some computationally heavy parts were written in Cython (Behnel, 2011).

References

Doucet, Arnaud and Johansen, Adam. *A Tutorial on Particle Filtering and Smoothing: Fifteen years later* in Handbook of Nonlinear Filtering (Oxford University Press, 2011).

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825-2830 (2011)

Giovannucci, Andrea; Friedrich, Johannes; Gunn, Pat; Kalfon, Jeremie; Brown, Brandon L; Koay, Sue Ann; Taxidis, Jiannis; Najafi, Farzaneh; Gauthier, Jeffrey L; Zhou, Pengcheng; Khakh, Baljit S; Tank, David W; Chklovskii, Dmitri B; Pnevmatikakis, Eftychios A (2019). CaImAn: An open source tool for scalable Calcium Imaging data Analysis. eLife, pp.e38173

Harris, C.P., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*.

Mathis, A., Mamidanna, P., Cury, K.M. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* 21, 1281–1289 (2018). https://doi.org/10.1038/s41593-018-0209-y

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero,
John D. Hunter. Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, 9, 90-95 (2007), DOI:10.1109/MCSE.2007.55 (publisher link)

Stéfan van der Walt, S. Chris Colbert and Gaël Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation, Computing in Science & Engineering, 13, 22-30 (2011), DOI:10.1109/MCSE.2011.37 (publisher link)

Stefan Behnel, Robert Bradshaw, Craig Citro, Lisandro Dalcin, Dag Sverre Seljebotn and Kurt Smith. Cython: The Best of Both Worlds, Computing in Science and Engineering, 13, 31-39 (2011), DOI:10.1109/MCSE.2010.118

Wes McKinney. Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference, 51-56 (2010)

Tai, L. H., Lee, A. M., Benavidez, N., Bonci, A., & Wilbrecht, L. (2012). Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. Nature Neuroscience, 15(9), 1281–1289. https://doi.org/10.1038/nn.3188

Theodore Lindsay, Peter T. Weir, Floris van Breugel. FigureFirst: A Layout-first Approach for Scientific Figures. Proceedings of the 16th Python in Science Conference 57-63 (2017). DOI: 10.25080/shinma-7f4c6e7-009

Zhou, P., Resendez, S.L., Rodriguez-Romaguera, J., Jimenez, J.C, Neufeld, S.Q., Giovannucci, A., Friedrich, J., Pnevmatikakis, E.A., Stuber, Garret D , Stuber, G.D., Hen, R., Kheirbek, M.A., Sabatini, B.L., Kass, R.E., Paninski, L. (2018). Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. eLife, pp.e28728.