VTA dopamine neurons multiplex external with internal representations of goal-directed action

Kremer Yves*[1], Flakowski Jérôme*[1], Rohner Clément[1], Lüscher Christian [1,2,3]

[1]Department of Basic Neurosciences, University of Geneva, CH-1211 Geneva, Switzerland

[2]Clinic of Neurology, Department of Clinical Neurosciences, University Hospital Geneva, CH-1205 Geneva, Switzerland

[3]Lead contact

*equally contributing authors

corresponding authors : yvkremer@gmail.com
christian.luscher@unige.ch

Summary

Dopamine (DA) released from VTA neurons in response to external cues or rewards may represent a learning signal. Here we introduce a spatial task where a reward-predicting cue instructs mice to navigate to a specific location. Simultaneous in vivo single-unit recordings revealed that DA neurons multiplexed an internal representation of the animal's goal-directed actions (locomotion, licking, distance) with phasic responses to cue and reward. Neuronal activity discriminated between rewarded and failed trials, generating an error signal even in the absence of external cues. Following a contingency change, mice readily learned to move to a different location, which became impossible, if the internal error signal was jammed by optogenetic stimulation. We conclude that a multiplexed internal representation of the task modulates VTA DA neuron activity, engaging a learning process that leads to the behavioral adaptation of goal directed action.

Keywords

Dopamine, ventral tegmental area, multiplexing, reward prediction error, locomotion, goal-directed action, negative learning signal, freely moving, single-unit recordings.

Introduction

Navigating through their environment, animals encounter salient stimuli that can shape actions based on previous rewards or punishments. When contingencies change, the behavior needs to be adapted rapidly by updating the relevant associations. Prediction of reward based on prior outcome contributes to the learning process to optimize behavioral performance. Midbrain dopamine (DA) neurons can support learning of cue-reward associations by encoding the reward prediction error (RPE) (Schultz et al., 1997). Unexpected rewards evoke a large phasic response in DA neurons. If a reward-predicting cue reliably precedes reward delivery, the phasic response shifts to the salient cue and the burst evoked by the now predicted reward decreases. These responses scale with value (Tobler et al., 2005), probability (Fiorillo et al., 2003), timing between events (Hollerman and Schultz, 1998) and if preceded by several stimuli, the RPE signal shifts to the earliest reliable reward-predicting cue (Schultz et al., 1993). The

hidden states inferred from previous training or subjective valuation impact on the RPE computation (Starkweather et al., 2017).

The activity of midbrain DA neurons also correlates with additional behavioral parameters. For example, DA neurons can code for motivation (Hamid et al., 2016), action initiation (Syed et al., 2015), distance to reward (Howe et al., 2013) or judgment of time between stimuli (Soares et al., 2016). There is evidence for specialization according to the anatomical location of the neuron. For example, VTA DA neurons support cue-associated Pavlovian conditioning (Saunders et al., 2018) while substantia nigra pars compacta (SNc) cells trigger action initiation and invigorate movement (Alves et al., 2018; Panigrahi et al., 2015). Based on single cell optical recordings, reward predictions and activity related to motor output may be encoded by distinct populations of DA neurons (Howe and Dombeck, 2016). Alternatively, it has been hypothesized that the phasic component of DA neuron firing encodes the RPE while more sustained changes in tonic firing in the same cell would encode movement vigor or motivation (Niv et al., 2005).

DA neuron functions are often derived from experimental readouts at the population level such as fast-scan cyclic voltammetry or bulk fluorescence imaging of calcium indicators (e.g. fiber photometry) in target regions or the entire midbrain population. For technical reasons, single-unit electrophysiological recordings have been carried out in restrained or head-fixed animals, precluding the exploration of the interplay between locomotion and RPE computation, the variable timing between relevant events and the perceived probability.

To overcome these limitations, we have designed a spatial learning task where a mouse had to search for a rewarded zone while freely moving around the arena. The only way to identify this zone was to stop inside for several seconds, which would trigger a CS. Once the CS presented, the animal could then collect the reward and engage in the next trial. Comparing failed and successful trials, we were able to observe the difference in neural responses and disentangle RPE from several parameters associated with locomotion. To this end, while mice executed the task, we performed single-unit recordings of the ventral tegmental area (VTA) and

simultaneously monitored RPE and locomotion. We found that VTA DA neurons exhibited phasic responses to salient events but also computed a representation of goal-directed actions, reliably modulating their firing rate to reflect motor output. This signal discriminated between rewarded and error trials even when animals failed to elicit cues due to incorrect behavior, thereby demonstrating that it was internally generated by the animal. Finally, we establish a link of causality by optogenetically jamming this error signal with tonic firing, which precluded the mouse to adapt to the changing contingency. Thus, VTA DA neurons are causally involved in learning not only from external cues but also from the outcome of goal-directed actions, both of which are multiplexed in the same cell.

Results

In the spatial task a reward-predicting light stimulus taught the mouse to go to a defined, but unmarked location of the operant chamber to obtain a reward (Fig.1A). During a pre-training phase, mice were conditioned to a standard CS-US paradigm, associating a randomly occurring 4s light stimulus (CS) to the availability of a liquid reward (US, fat solution; 5% of lipofundin). The reward was available for 4s after the light stimulus had been switched off and was only delivered if the mouse licked against a drinking spout during this time window. The first lick occurring during the reward window triggered an opening of the valve which delivered a single drop to the animal (Fig. 1B, top). Once this association had been learned, we switched to the cue-guided spatial task, where the CS would be given after the mouse had spent 2s in a small (4x4cm), unmarked zone of the operant chamber. The reward zone (RZ) was such that the animal had to remain still while waiting for the CS and then move to the spout to collect the reward (Fig. 1B, below). If the mouse left the RZ too early or failed to lick, no reward was delivered (error trial). Mice rapidly associated their spatial position to the CS and optimized their locomotor pattern over days (Fig. 1C-D). As a consequence the number of RZ entries and the reward rate increased with each session (Fig. 1E-F). Compared to pre-training sessions, where the number of rewards was limited by the random inter-trial interval imposed by the experimenter, in the spatial task the maximum reward rate was limited only by the duration of the executed sequence (approx. 10s) and the animal's performance. Once mice had located the RZ, they produced highly reproducible locomotor output and followed specific

spatial trajectories between the zone and the drinking spout (Suppl. Video S1). Across all training phases, mice were video-tracked at 20Hz and all relevant behavioral events time-stamped (Fig. 1D).

We carried out the spatial task in DAT-cre mice (n=10) injected with virus expressing cre-dependent channelrhodopsin (ChR2) and implanted with a 16-channel optrode mounted into a microdrive (Fig. 2A). We recorded n=161 extracellular single-units in the VTA (plus several neurons above and below the VTA, Fig. S1A & Star methods) and found a large variety of spiking patterns, regardless of the learning stage and even when comparing simultaneously recorded neurons (Fig. 2B). To quantify the firing rate changes with behavior, we used a receiver-operating characteristic (ROC) (Cohen et al., 2012). Because precise timing between events was not determined by the task but largely depended on the animal's behavior, recordings were aligned to 5 individual events of interest (CS, US, lick burst onset, speed peak and acceleration peak). After a dimensional reduction using independent component analysis (ICA) in combination with the mean firing rate of each cell, a hierarchical clustering was applied to group the neurons based on their spiking patterns (Fig. 2D-E). Nine clusters emerged, out of which one (#7) contained 80 neurons with phasic responses to CS and US suggestive of DA neurons. In fact, when an optogenetic identification procedure was applied (Fig. 2C), all positive neurons (n=17) fell into cluster #7, thus confirming their identity as DA neurons. By contrast, non-DA neurons showed sustained firing aligned to specific time points of the trial or were activated during specific motor output such as licking or locomotion. Compared to previous studies in head-fixed animals (Cohen et al., 2012; Sadacca et al., 2016), our analysis revealed clusters reflecting motor output in non-DA neurons, which were modulated by locomotor speed or acceleration (cluster 3 & 4; n=6 & 25 neurons) or licking (cluster 1 & 2; n=4 & 3 neurons). Continuous spike density functions (cSDF) of the remaining clusters could not easily be attributed to individual behaviors and exhibited more complex structures (Fig. 2E).

The most prominent features of the DA neuron population were the large, phasic CS- and US-evoked responses (Fig. 3A). We expected that as the mice learned to associate the cue to the reward the US would decrease (Schultz et al., 1993). Surprisingly, in 31% (25/80

neurons), the response to the predicted US was significantly larger than the CS response, whereas 46% (37/80 neurons) showed larger responses to the CS than to the US (Wilcoxon rank sum test, $p<0.05$). Furthermore, the distribution of response amplitudes measured in optogenetically identified DA neurons (5/17 with USresp>CSresp vs. 11/17 with CSresp>USresp; Wilcoxon rank sum test, $p<0.05$) was similar to responses across all recorded DA neurons (Fig. 3B). Remarkably, the two optogenetically identified neurons in Fig. 3A, one showing predominantly responses to the CS whereas the second cell showed much larger responses to the US, were recorded simultaneously i.e. during the same behavior and at the same learning stage.

In parallel, 51/80 of DA neurons also showed slow ramping or decreasing firing rates (Fig. 3C, top) when modulated by goal-directed actions (Fig. 3C-D; Fig. S2D for more examples) during task execution (Fig. 3C,D,F). This modulation was more pronounced between CS and US (58/80 DA neurons modulated between CS-US vs 27/80 DA neurons outside CS-US) and contrasted with non-DA neurons where modulation was often observed outside the CS-US window (75/81 non-DA neurons modulated between CS-US vs 70/81 non-DA neurons outside CS-US) (Fig. 3F). To further isolate motor responses from CS or US responses, we repeated the same analysis after excluding any motor events occurring closer than 1250ms to either of these two events. Putative DA neurons and identified neurons showed similar response profiles (Fig. S2E)

We next sought to estimate how many factors contributed to the firing pattern of a given neuron. We applied an encoding approach using a regularized Poisson generalized linear model (GLM) (Fig. 4A) to the entire, continuous session taking into account all monitored behavioral variables and external events: CS, US, distance to reward, in/out of RZ, licks, acceleration and speed (Fig 4B). We then evaluated the contribution of each variable by calculating the correlation between the actual neural signal and the encoder prediction.

DA neurons typically required 3-4 contributors to explain their firing pattern based on both external events and licking and locomotion parameters (Fig. 4C; Fig. S3 & Star methods).

We then ranked them according to the probability of their contribution (Fig. 4D) and confirmed the contribution of each variable by removing them one by one while monitoring the decrease in correlation between the recorded signal versus the prediction (Fig. 4E). This confirmed that DA neurons multiplexed external events and actions. When comparing the decoding performance of the GLM within the CS-US window and outside the CS-US, we found a stronger correlation for the former (Fig. 4F). This indicates that salient, reward-predicting events provide a time window during which goal-directed actions are closely monitored by DA neurons.

We next had a closer look at error trials, during which the mouse made timing mistakes and failed to trigger the CS but still approached and licked against the drinking spout to test for reward availability. This allowed us to compare successful trials to error trials with similar motor output (Fig. 5A, left). When monitoring the neural activity of DA neurons (Fig. 5A, right), we found striking differences in both, the averaged trial histogram as well as in individual trials during a one hour session. This contrast was most prominent when neural activity was aligned to transitions of goal directed actions, such as lick bursts or locomotion. Therefore we compared successful and error trials at the time of zone entry, when the mouse left the zone and at lick burst on- and offset.

This analysis revealed significant differences in the neural activity when comparing successful and failed trials. Even when no CS or US occurred during error trials, neurons exhibited an internally generated error signal aligned to the mouse's actions. In VTA DA neurons, it consisted not only of decreased phasic firing but was turned into a negative error signal further emphasized by pauses occurring in the tonic firing (Fig 5A). 31/80 (41%) of DA neurons showed a significant increase in spike pause probability in at least one of the events (10/80 for enter zone event ,18/80 for leave zone event, 15/80 for lick burst onset and 19/80 for lick burst offset) in failed trials compared to successful trials (Fig. 5B-C). Discriminating neurons were not necessarily specific to a single event (Fig. 5D).

To establish a link of causality, we designed experiments during which we would modulate the negative error signal and monitor the impact on the performance with a

contingency change (change of location of the RZ). We either applied an optogenetic stimulation during the activity pause in error trails or enhanced the pause through optogenetic inhibition. To this end, we injected DAT-cre mice (n=12) bilaterally with virus expressing either cre-dependent channelrhodopsin (ChR2) (n=5) or archaerhodopsin (eArch3.0) (n=3) (Fig.6A). We used DAT-cre animals injected with virus expressing green fluorescent protein as controls (n=4). In fully trained mice, the spatial task would initially run normally with RZ1 activating the CS. After 20 min, the RZ switched to a different location (RZ2). ChR2-mice that returned to RZ1 would receive an optogenetic stimulation simulating DA neuron tonic firing to jam the pause present during error trials. Animals injected with eArch3.0 received a 500 ms orange light stimulation, to enhance the negative error signal in order to facilitate the learning of the new contingency. eYFP-injected control mice received either of these stimulation patterns at random (Fig.6C).

ChR2-mice that received the optogenetic stimulation, persevered in RZ1 after RZ2 was introduced, while control mice started to search for RZ2 and successfully triggered the CS within the second 20min block (Fig.6B). The number of times that the CS was triggered was significantly lower in stimulated ChR2-mice along with a reduction of rewards (Fig.6D). There was no significant improvement in performance with eArch3.0-inhibition. It is possible that the pause already occurring in the internal negative error signal could not be enhanced. The rapid adaptation to the changing contingency in control mice further supports that the natural error signal is already optimized, leading to a ceiling effect.

Discussion

Our single-unit recordings in freely moving mice show that RPE and motor related activity are not segregated by anatomical regions or confined to specialized cells but instead encoded by different firing modes in a given VTA DA neuron. In other words, we provide experimental evidence for multiplexed signaling in DA neurons and show that the RPE signal and movement related firing take advantage of the different firing modes of DA neurons to be processed in parallel and that they are modulated in an outcome-dependent manner thus contributing to the learning of a new contingency.

Multiplexing as an efficient mean to transmit complex information by a single channel, for example by dividing the frequency or time domain. Frequency-division multiplexing has been proposed for DA neuromodulation on theoretical grounds (Hiroyuki, 2014; Niv et al., 2007; Oster et al., 2015; Schultz, 2007) but experimental evidence remains scarce and although conjunct encoding of reward- and movement-related signals exist (Barter et al., 2015; Puryear et al., 2010), formal identification of neuronal cell-types remain elusive.

Multiplexing may arise from the convergence of many inputs onto VTA DA neurons, which receive inputs from motor control (M1, M2) and general locomotor regions (diagonal band of Broca, medial septal nucleus, pedunculopontine nucleus, laterodorsal tegmentum) (Fuhrmann et al., 2015; Watabe-Uchida et al., 2012; Xiao et al., 2016) and inputs shaping the RPE computation. The latter are believed to be more widely distributed, neurons contributing to the positive RPE signal have been found in striatum, ventral pallidum, subthalamic nucleus, pedunculopontine nucleus and lateral hypothalamus (Tian et al., 2016). For the negative error signals, the lateral habenula has emerged as the main input (Matsumoto and Hikosaka, 2007; Tian and Uchida, 2015). Interestingly specific inputs can differentially affect individual firing modes in DA neurons thereby precisely addressing a single channel of the multiplexed signal (Floresco et al., 2003). Without surprise, optogenetic manipulations of VTA DA neurons have revealed strong behavioral effects, which may be explained by the many information streams treated in parallel (Beier et al., 2015; Lammel et al., 2008, 2011, 2012; Xiao et al., 2016).

Our data shows that the firing patterns of individual DA neurons contain much task-related information, which may serve as a learning signal. This may be reflected in the observation that the RPE component did not show an extinction of the US response. Even in expert animals, we found a large fraction of neurons with a response that was stronger to the predicted US than to the CS. This was not due to different stages of learning, as all mice had fully associated the CS with the availability of the US before they were introduced to the spatial task. Moreover, during recording sessions where several DA neurons were monitored, we found DA neurons showing large responses to the US next to DA neurons that did not respond to the US. Fig. 3A shows such an example of two simultaneously recorded and optogenetically identified DA neurons, at the same learning stage and during the same behavior (see also Fig. 2B for 3 simultaneously recorded DA neurons). We also found examples of both responses during the pre-training phase (Fig. S2A), ruling out the possibility that the RZ interfered with the RPE computation.

A possible explanation for this non-canonical phasic response to a predicted reward might be the variable time interval between CS and US. This would mean that compared to a setting where strict probabilities and timing between events are imposed by the experimenter, more natural behaviors maintain a phasic DA response to rewards.

The phasic DA signal has been proposed to contain two components, early responses (0-250ms) correspond to a saliency response when the event is detected, the second component (250-500ms) represents the value of the event that occurred (Nomoto et al., 2010). In our study, CS responses have both, early and late components, whereas US responses predominantly occurred during the early 250ms window (Fig S2C), which may reflect a saliency and a value component for the former and an exclusive saliency signal for the latter. This suggests that the detection of both, CS and US, are important in challenging tasks, whereas the CS carries the value prediction of the upcoming US (Fig S2B).

The movement-related information encoded by DA neurons may also contribute to learning. Using peri-event analysis (Alves et al., 2018) and a generalized linear model approach

(Allen et al., 2017), we found a mixture of external events and goal-directed actions predicting the neural signal of VTA DA neurons consistent with task-related modulation in DA firing rates described in monkeys while performing a forelimb reaching task (Schultz, 1986). This was in contrast to non-DA neurons whose firing patterns could largely be explained by motor output. The fact that motor output alone did not reliably modulate the spiking of DA neurons but rather tracked specific actions during relevant time windows such as the CS-US time-interval, further suggests that it represents a learning signal. This is in line with optogenetic manipulations where direct stimulation of VTA GABAergic neurons show disruption of motor output (Van Zessen et al., 2012), whereas phasic firing of VTA DA neurons reinforces action rather than driving motor output (Tsai et al., 2009). In SNc DA neurons, transient increases in firing have been observed before action initiation and optogenetic stimulation promoted movement initiation (Alves et al., 2018). We found brief as well as sustained increases or decreases in spiking upon motor output, following action initiation and the modulation was dependent on the outcome of the individual trial (successful versus failed trial), all indicating that VTA DA neurons track performance of goal-directed actions. Actions occurring during failed trials produced significant decreases in spiking and increased the probability for neurons to produce a pause. Pausing in VTA DA neurons mediate a negative RPE in a Pavlovian task (Chang et al., 2015), which we causally linked to learning from contingency changes. The negative prediction error signal may thus not be limited to event evaluation but also involves action evaluation.

Depression in DA neuron firing has been observed in monkeys at the time of expected reward delivery during learning of a visual discrimination task (Hollerman and Schultz, 1998), suggesting that such an error learning signal may have escaped studies where animals are overtrained or errors not taken into account, although they could contribute to learning. Moreover, when the experimenter strictly controls cues and rewards, the animal cannot influence the outcome of a trial and will not be able to assign a value to an action. In our spatial task, the animal can build a model of the task structure and continuously improve future outcomes by adjusting its actions. DA signaling has been proposed to underlie this type of action selection (Costa, 2011; Nakahara and Hikosaka, 2012). Another analogous situation may apply to songbirds, which during song learning, when presented with a distorted auditory

feedback, produced a similar negative performance signal in VTA DA neurons (Gadagkar et al., 2016).

But how might the signal then be de-multiplexed? Downstream of the VTA, cell-type specific changes in nucleus accumbens (NAc) can occur during learning (Atallah et al., 2014). In fact D1- and D2-receptors expressed on individual medium spiny neurons differentially regulate synaptic plasticity of afferent glutamate transmission (Shen et al., 2008; Yagishita et al., 2014), which has been proposed to underlie reinforcement learning (Schultz, 1998). As D1- and D2-receptors have different affinities, it makes them ideal candidates to decode a multiplexed DA-signal (Dreyer et al., 2010; Marcott et al., 2014), and differentiate between the phasic and tonic spiking. In line with this interpretation does closed-loop stimulation of either D1- or D2-striatal medium spiny neurons (MSNs) during movement execution bi-directionally shift specific parameters of the stimulated movement (Yttri and Dudman, 2016). Similar mechanisms could underlie the optimization of task-related parameters in ventral striatum during goal-directed actions. Similar DA-dependent mechanisms have been reported in other downstream areas of the VTA: stimulation of VTA DA projections to prefrontal cortex showed opposing effects between phasic and tonic stimulation (Ellwood et al., 2017) and functional imaging studies in humans have implicated the posterior medial frontal cortex in learning from errors and showed that it was dopamine and D2-receptor dependent (Klein et al., 2007).

Our results may also have translational implications. Presentation of salient cues to patients with Parkinson's disease can improve symptoms, for example by overcoming freezing of gait (Gilat et al., 2018; Ginis et al., 2017). If elevated DA levels constitute an underlying mechanism, then our data indicate that adding cues may recruit otherwise silent VTA DA neurons, which are less affected by neurodegeneration than SNc neurons (Surmeier et al., 2017) and could partially compensate in regions where SNc and VTA projections overlap. Another implication might be that in substance abuse, where pharmacological jamming of the error signal by the drug in addition to environmental cues could drive the persistent consumption despite negative consequences and lead to addiction (Keiflin and Janak, 2015; Pascoli et al., 2015).

Acknowledgments

We would like to thank Anthony Holtmaat and Alexandre Pouget for helpful comments on the manuscript and all members of the Lüscher lab for stimulating discussions. This work was supported by the Swiss National Science Foundation (SNSF; FNS310030B_170266) and by the European Research Council (ERC; UE7-MESSI-322541).

Author contributions

Y.K. & C.L. designed the experiments. Y.K. & C.R. trained the animals. Y.K. built the set-up, performed the surgeries and recordings. Y.K. and C.R. performed the histology and confocal imaging. Y.K., J.F. and C.R. analyzed the data. Y.K., J.F. and C.L. wrote the manuscript.

Declaration of interests

The authors declare no competing interests.

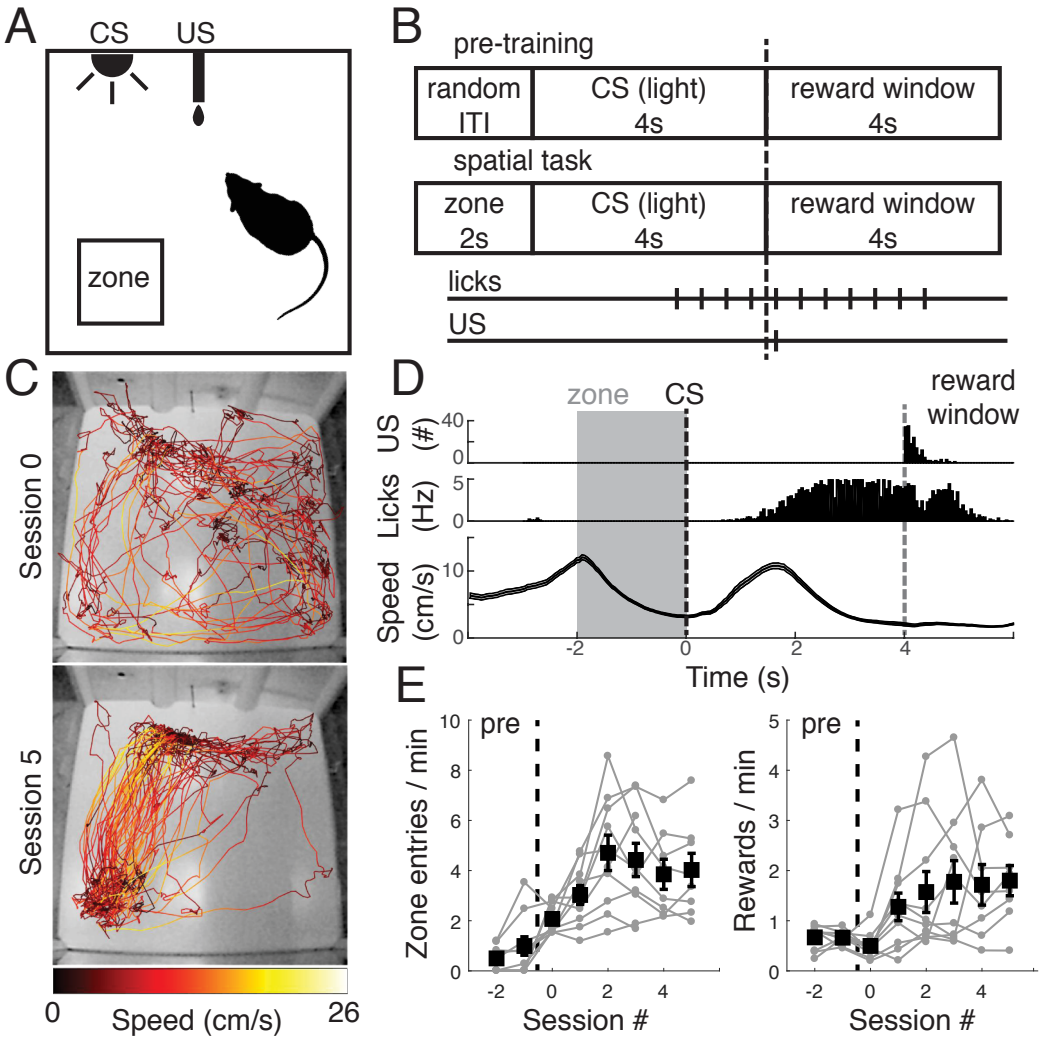
A
CS
US
zone
B
pre-training
random ITI
CS (light) 4s
reward window 4s
spatial task
zone 2s
CS (light) 4s
reward window 4s
licks
US
C
Session 0
Session 5
0
Speed (cm/s)
26
D
zone
CS
reward window
US (#)
Licks (Hz)
Speed (cm/s)
Time (s)
E
pre
Zone entries / min
Session #
pre
Rewards / min
Session #

Figure 1

Figure 1. The cue-guided spatial task.

(A) Schematics of the operant chamber for training and single-unit recordings. The reward zone (RZ) is highlighted for illustration but invisible for the mice that had to make the association between their spatial position and the CS.

(B) During pre-training, mice had to associate a randomly occurring 4s light cue (CS) to the availability of a liquid reward (US). The first lick performed during a 4s reward window received a single drop of fat solution (top). Once this association had been learned, we switched to the spatial task, during which mice had to wait for 2s in the RZ to trigger the CS which indicated that the reward could be collected (below).

(C) Track plots constructed from 10.000 video frames (~6-7 min) showing the locomotor pattern of the same mouse at the first and 6th session. Note that the animal optimized its trajectories maximizing the reward rate.

(D) Behavioral variables recorded during session 5 shown in panel (C). Individual licks, locomotion and US distribution during the reward window aligned to the CS. The greyed area indicates the 2s period spent inside the RZ.

(E) Behavioral performance of all mice (n=10). Sessions -2 & -1 correspond to the last two pre-training sessions. During the spatial task, all mice learned the association between the RZ and the CS, increasing the time spent inside the RZ as well as the number of rewards received (mean ± s.e.m.).

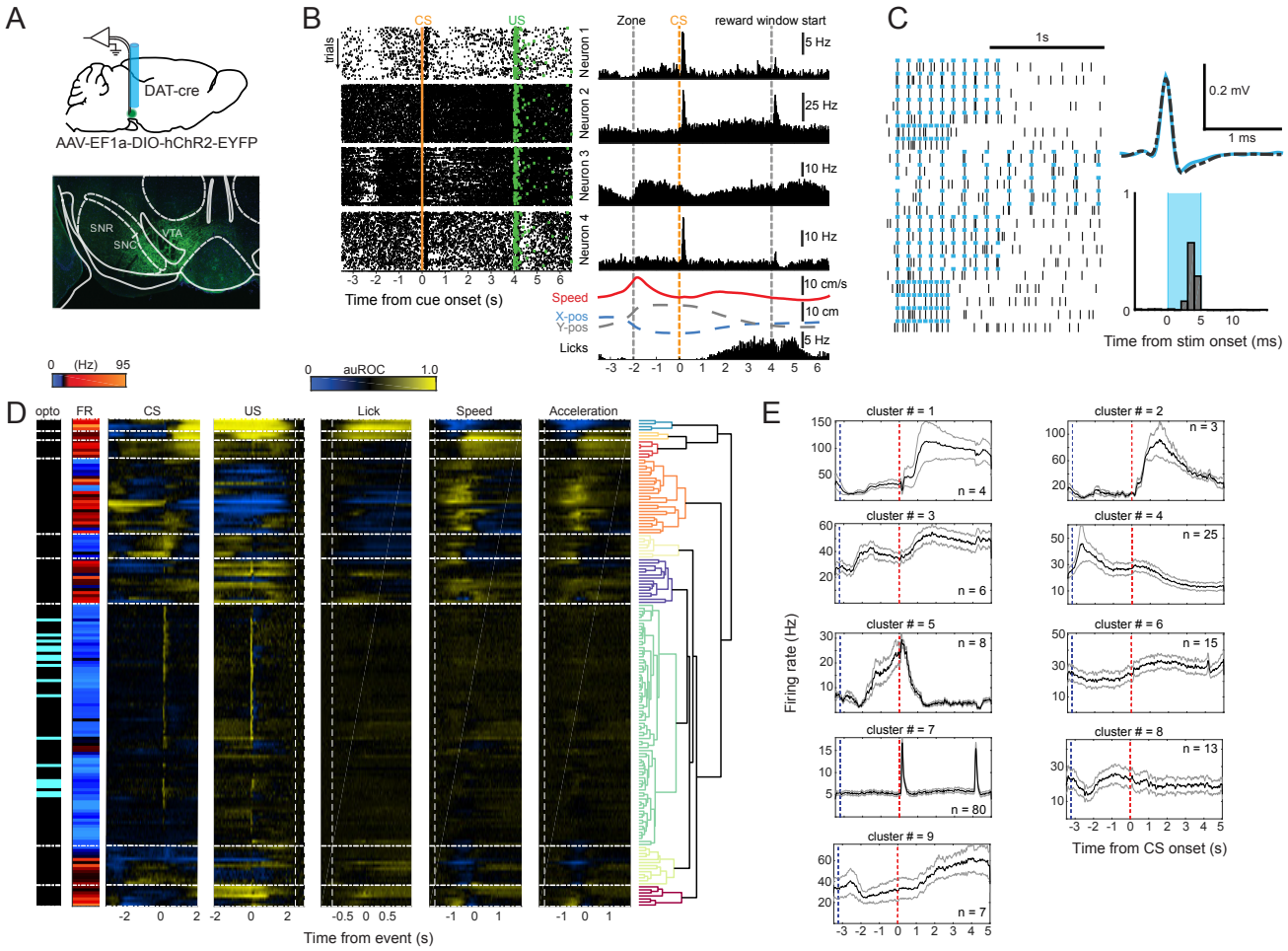
A
DAT-cre
AAV-EF1a-DIO-hChR2-EYFP
SNR
SNC
VTA
B
CS
US
trials
Neuron 1
Neuron 2
Neuron 3
Neuron 4
Time from cue onset (s)
Zone
reward window start
5 Hz
25 Hz
10 Hz
10 Hz
Speed
10 cm/s
X-pos
Y-pos
10 cm
Licks
5 Hz
C
1s
0.2 mV
1 ms
Time from stim onset (ms)
0 (Hz) 95
0 auROC 1.0
D
opto
FR
CS
US
Lick
Speed
Acceleration
Time from event (s)
E
cluster # = 1
n = 4
cluster # = 2
n = 3
cluster # = 3
n = 6
cluster # = 4
n = 25
cluster # = 5
n = 8
cluster # = 6
n = 15
cluster # = 7
n = 80
cluster # = 8
n = 13
cluster # = 9
n = 7
Firing rate (Hz)
Time from CS onset (s)

Figure 2

Figure 2. Emergence of action-related neuronal clusters in freely moving mice.

(A) Schematics of the single-unit recordings. Custom-made optrodes were implanted into the left VTA of adult DAT-cre mice that were injected with virus expressing floxed ChR2.

(B) Example of four simultaneously recorded units. Left: raster plots for each individual neuron. Right: average spiking histogram aligned to the CS showing the diversity of firing patterns during the same session along with the speed of the mouse, its spatial position and licking (below).

(C) Optogenetic identification of DA neurons. Left: 5ms blue light pulses at 5, 10 and 20 Hz. Neurons exhibiting a response rate ≥0.8 spikes/pulse during the 6 ms window after light onset and showing a waveform correlation ≥0.85 were identified as light-responsive (bin size opto=1ms).

(D) auROC spiking profiles around events of interest as well as the mean firing rate (FR) were used to reveal functional clusters of single-units. All optogenetically identified neurons fell into cluster 7, which exhibited large phasic responses around CS and US.

(E) Average continuous spike density function (cSDF) aligned to the CS (mean ± s.e.m.) for each functional cluster highlighted response patterns with clear modulation to task-related actions such as locomotion (clusters 3, 4), licking (clusters 1 & 2) as well as more complex behaviors (clusters 5, 6, 8 & 9). Colors of the bounding boxes correspond to colors of the clusters in the dendrogram in panel (D).

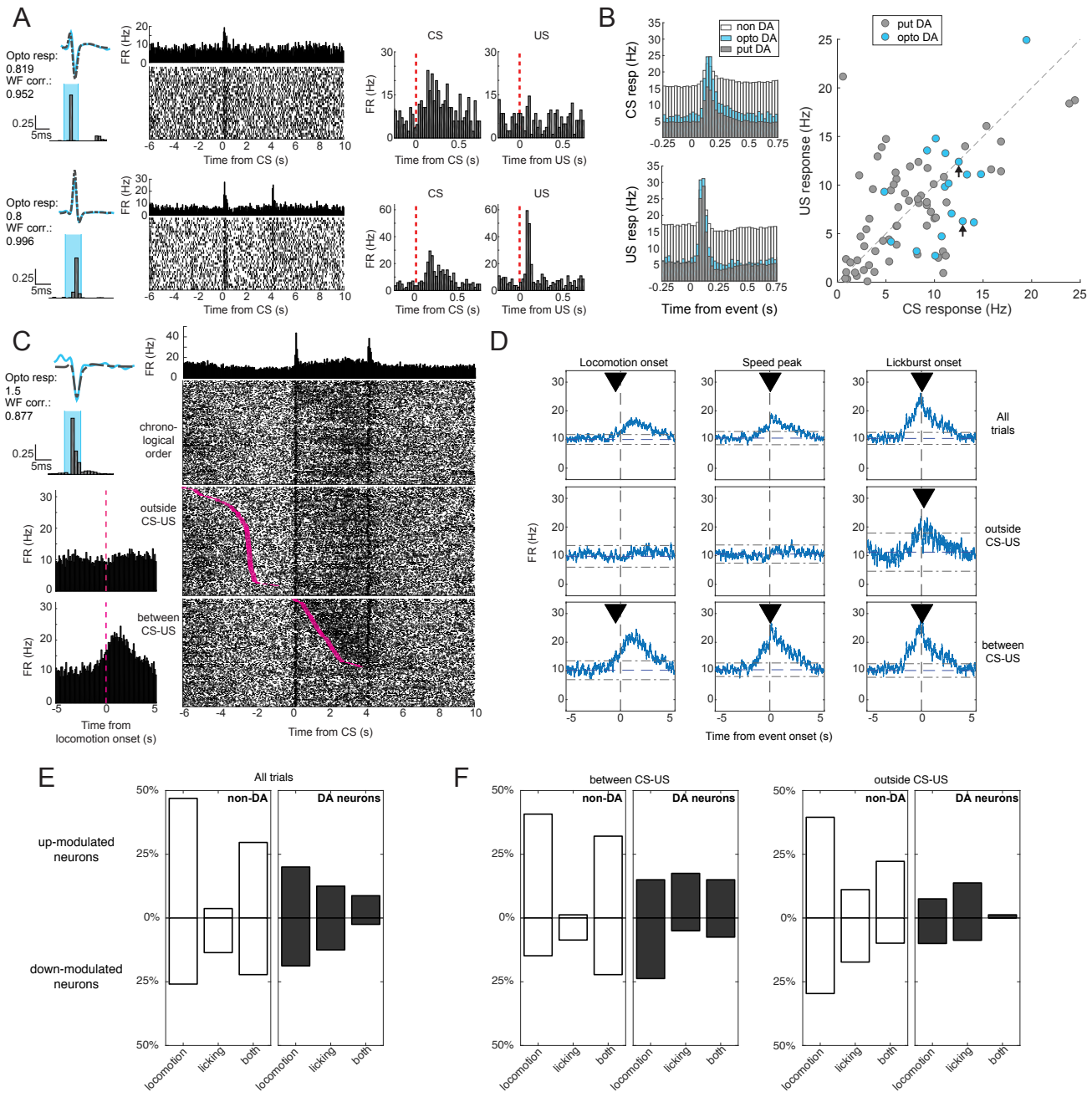
A
Opto resp: 0.819
WF corr.: 0.952
Opto resp: 0.8
WF corr.: 0.996
FR (Hz)
Time from CS (s)
CS
US
Time from US (s)
B
non DA
opto DA
put DA
CS resp (Hz)
US resp (Hz)
Time from event (s)
US response (Hz)
CS response (Hz)
C
Opto resp: 1.5
WF corr.: 0.877
chrono-logical order
outside CS-US
between CS-US
Time from locomotion onset (s)
D
Locomotion onset
Speed peak
Lickburst onset
All trials
Time from event onset (s)
E
All trials
non-DA
DA neurons
up-modulated neurons
down-modulated neurons
locomotion
licking
both
F
between CS-US
outside CS-US

Figure 3

Figure 3. DA neurons exhibit phasic responses to CS and US and simultaneously encode motor parameters.

(A) Two optogenetically identified DA neurons, simultaneously recorded, one showing canonical RPE responding exclusively to the CS (top) while the second neuron exhibited also a large response to the predicted US (below; bin size opto=1ms; bin size PSTH=25ms).

(B) Left: Average response histograms of all neurons to CS (top) and US (below) showing large average responses to both events. Right: Average US vs CS responses during the 500ms post-event window for all DA neurons (bin size PSTH=25ms). Black arrows highlight the neurons in (A).

(C) Example of an optogenetically identified DA neuron exhibiting sustained increased firing during locomotion and licking. Top: average histogram (bin size=50ms). Raster plots show rewarded trials sorted in chronological order (top), by time interval between locomotion onset (magenta dots) preceding zone entry and CS (middle) and by time interval between locomotor onset and CS between CS and US (below). Below left: Histograms show average firing aligned to locomotion onset (bin size PSTH=100ms; bin size opto=1ms).

(D) Response detection on the same neuron, using the firing rate aligned to locomotion onset, speed peak and lick burst onset. Rows indicate spiking to all events (top), outside CS-US (middle) and between CS-US (below). Black arrows indicate time of significant response to the event of interest (STAR methods).

(E) Distribution of non-DA and DA neurons modulated by motor parameters, cells for locomotion onset or speed were grouped together.

(F) Distribution of non-DA and DA neurons modulated by locomotor parameters confined to the time interval between CS and US or outside this time window.

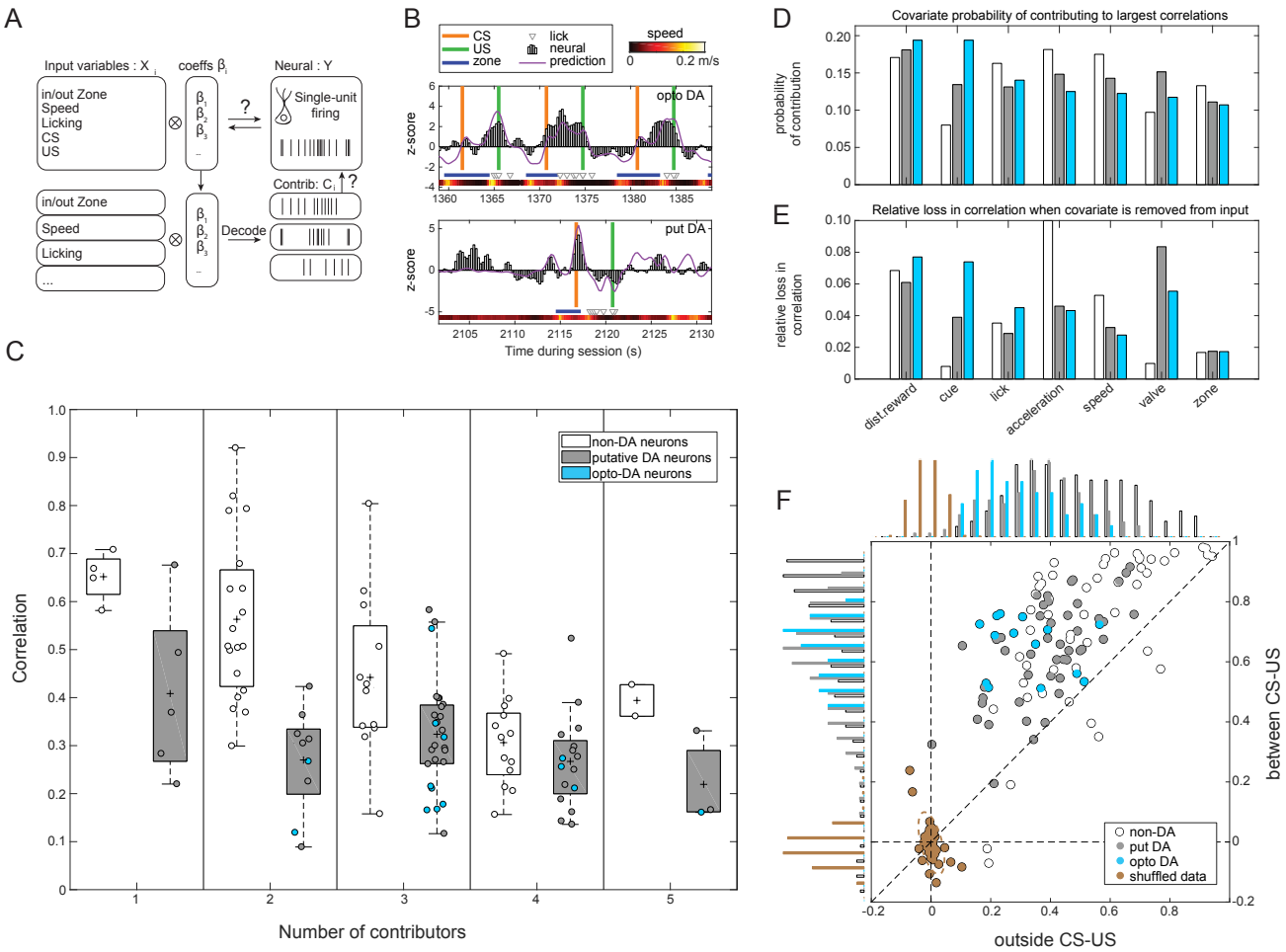
A
Input variables : X
coeffs β
Neural : Y
in/out Zone
Speed
Licking
CS
US
Single-unit firing
Contrib: C
Decode
B
CS
US
zone
lick
neural
prediction
speed
0
0.2 m/s
opto DA
put DA
z-score
Time during session (s)
C
Correlation
Number of contributors
non-DA neurons
putative DA neurons
opto-DA neurons
D
Covariate probability of contributing to largest correlations
probability of contribution
E
Relative loss in correlation when covariate is removed from input
relative loss in correlation
dist.reward
cue
lick
acceleration
speed
valve
zone
F
between CS-US
outside CS-US
non-DA
put DA
opto DA
shuffled data

Figure 4

Figure 4. DA neurons reflect multiplexed action components during task execution.

(A) Schematics of our encoding approach. All recorded variables were used to construct the encoder and to predict the firing of each neuron individually. Then the correlation between the prediction and the neural signal was used to assess the contribution of each behavioral variable by either adding or removing them one-by-one from the input to the encoder.

(B) Two example DA neurons: the z-scored recorded signal (white histogram) is compared to the encoder's prediction (purple line). All external events (vertical lines) and behavioral variables (below; triangles: licks; heat map: speed; blue: presence in zone) are also indicated.

(C) For non-DA and DA neurons, we first defined the contributing variables required to reach 90% of the maximum correlation (also see Fig.S3 & Star methods). The plot shows the correlation between the recorded neural signal and the signal predicted using only those variables as a function of the number of contributors.

(D) Ranking of input variables according to the average probability to be among the strongest contributors of individual neurons and grouped by cell-type. The higher the value, the more often the variable is among the most informative contributors. Input variables are displayed in decreasing order with respect to the optogenetically identified DA neurons.

(E) Average loss in amount of correlation for each variable when individually removed from the input plotted as a function of the cell type. The higher the value, the more information is lost (and the more important the contributor) when the variable is removed from the input.

(F) Scatterplot comparing the average correlation between the CS-US vs outside CS-US time interval. Note that the neural activity is better predicted between the CS-US interval compared to outside the CS-US interval, indicating encoding of behavioral variables during task execution. Plotted in brown are the shuffled controls for each neuron representing the average correlation between the predicted signal and 100 randomly shifted neural recordings.

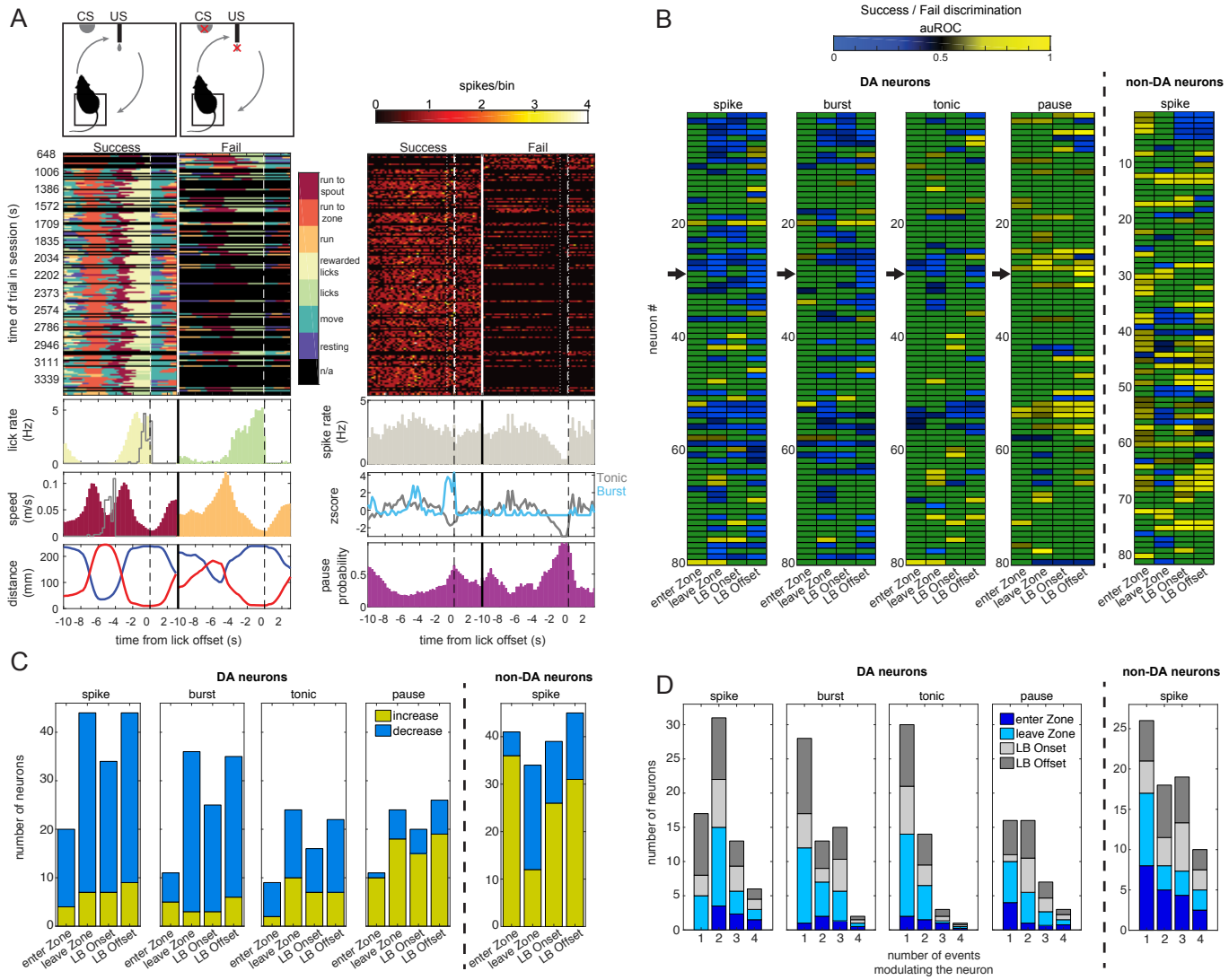
A
CS US
CS US
spikes/bin
0 1 2 3 4
Success
Fail
time of trial in session (s)
run to spout
run to zone
run
rewarded licks
licks
move
resting
n/a
lick rate (Hz)
speed (m/s)
distance (mm)
spike rate (Hz)
zscore
Tonic
Burst
pause probability
time from lick offset (s)
B
Success / Fail discrimination
auROC
DA neurons
non-DA neurons
spike
burst
tonic
pause
neuron #
enter Zone
leave Zone
LB Onset
LB Offset
C
DA neurons
non-DA neurons
increase
decrease
number of neurons
D
enter Zone
leave Zone
LB Onset
LB Offset
number of events modulating the neuron

Figure 5

Figure 5. Action related spiking of DA neurons discriminates between successful and failed trials.

(A) Successful vs failed trial analysis of one example session. Left: trial-by-trial analysis of behavioral modules used to define successful (left) and failed (right) trials in temporal order as performed during the session. Below: lick rate, speed and the animal's distance to the RZ (blue) and to the spout (red) averaged across success and fail trials. Grey bars on lick and speed histograms represent US and CS distribution respectively. Right: Neuronal activity of a DA neuron split into successful and failed trials as defined from the behavioral analysis. Below: average spiking histogram, z-scored tonic & bursting activity and probability of being in a spike pause. All plots are aligned to the lick burst offset (bin size=250ms).

(B) Left: Success vs fail discrimination of all DA neurons (n=80). Significantly discriminating neurons are represented by their auROC change in spiking, burst firing, tonic firing and pausing. Black arrows highlight the example neuron in (A). Neurons whose activity did not change significantly are represented in green (Wilcoxon two-sided rank sum tests with α=0.05). Right: Discrimination of non-DA neurons is shown for their spiking (n=81) (burst/tonic/pause modes were undefined for non-DA neuron).

(C) Left: Number of DA neurons significantly discriminating success vs fail trials either with an increase or a decrease in the corresponding firing modality. The negative error signal, which manifests as a pause in the tonic activity, is carried by 22% of DA neurons (18/80 for leave zone event, 15/80 for lick burst onset and 19/80 for lick burst offset). Right: Spiking modulation for non-DA neurons.

(D) Left: Number and distribution of actions, which significantly modulate DA neurons in a specific firing modality. Right: Same representation for non-DA neurons.

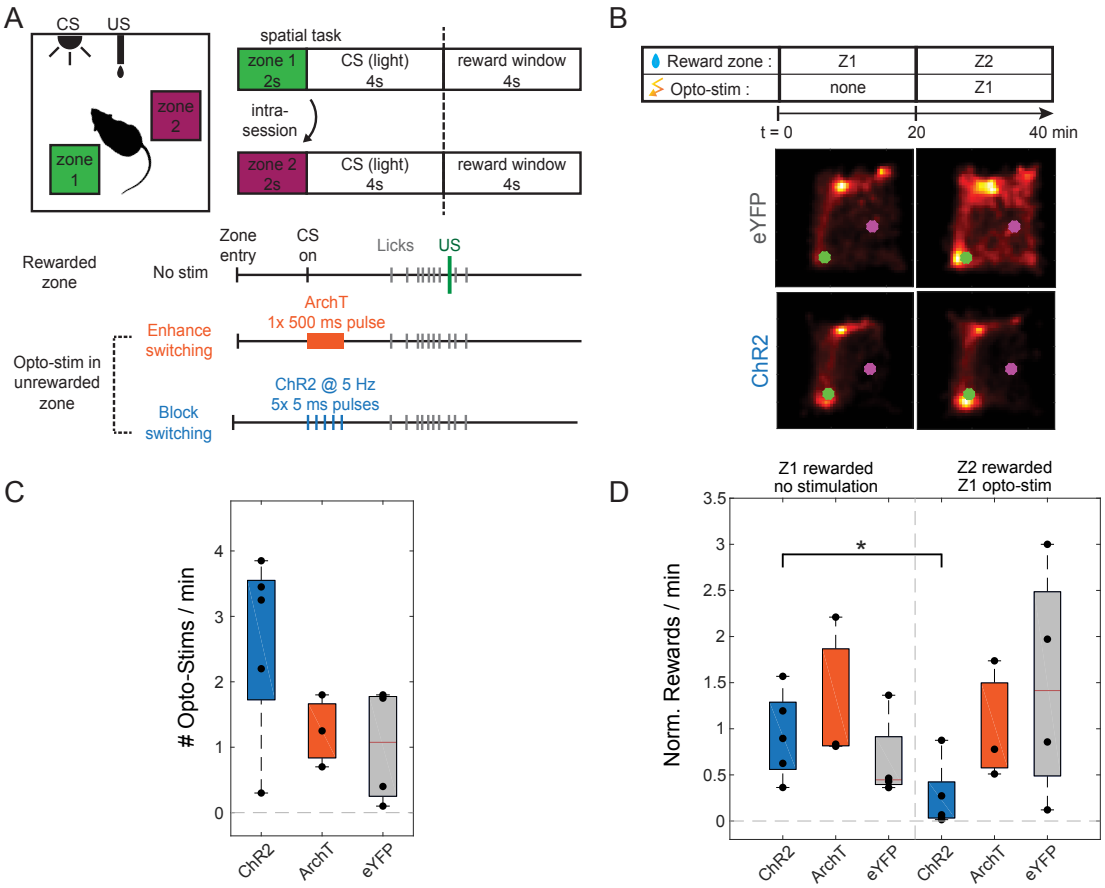
A
CS
US
zone 1
zone 2
spatial task
zone 1 2s
CS (light) 4s
reward window 4s
intra-session
zone 2 2s
CS (light) 4s
reward window 4s
Rewarded zone
No stim
Zone entry
CS on
Licks
US
Opto-stim in unrewarded zone
Enhance switching
ArchT
1x 500 ms pulse
Block switching
ChR2 @ 5 Hz
5x 5 ms pulses
B
Reward zone :
Z1
Z2
Opto-stim :
none
Z1
t = 0
20
40 min
eYFP
ChR2
C
Opto-Stims / min
ChR2
ArchT
eYFP
D
Z1 rewarded no stimulation
Z2 rewarded Z1 opto-stim
Norm. Rewards / min
*
ChR2
ArchT
eYFP
ChR2
ArchT
eYFP

Figure 6

Figure 6. Optogenetic jamming of the error signal prevents learning from mistakes.

(A) Experimental rationale. Expert mice were first exposed to the regular spatial task. After 20 mins, the location of the RZ was moved from location 1 (RZ1) to location 2 (RZ2). During correct trials, no optogenetic stimulation occurred, but when the mouse returned to RZ1 (now unrewarded) optogenetic stimulation or inhibition was delivered at the time the CS would have been presented. ChR2-mice received 5x 5ms blue light pulses to fill the pause with optogenetic tonic firing, eArch3.0-animals received 500ms of continuous orange light to enhance the pause.

(B) Track plots of 1 control (eYFP) and 1 stimulated (ChR2) animal exhibiting similar performance and behavior during the 20 min baseline period left. Once the change in location of the zone had occurred, the control mouse quickly explored the chamber to activate the CS. A brief tonic optogenetic stimulation jamming the error signal prevented learning.

(C) Average optogenetic stimulation rate received per animal, indicating how often animals returned to the unrewarded RZ1.

(D) Normalized reward rate for each animal during baseline and during stimulation periods. The ChR2-group showed a significant decrease in reward rate due to the jamming of the error signal (Wilcoxon rank sum test; $p=0.03$ for the ChR2 group; $p=0.40$ and $p=0.49$ for eArch3.0 and control animals respectively).

STAR METHODS

KEY RESOURCES TABLE

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed and will be fulfilled by the Lead Contact, Prof. Christian Lüscher (christian.luscher@unige.ch).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

All experiments were reviewed by the institutional ethics committee and approved by the relevant authorities of the Canton of Geneva. Animals were housed individually after implantation to avoid damage to the optrode. During training, they were food-restricted to a single pellet of chow per day (~2g). The animal's weight did not drop below 85%. All animals used in this study were adult DAT-cre mice. Only males were used for the electrophysiological recordings because of the weight of the optrode implant. For optogenetic experiments, male and female mice were trained.

METHOD DETAILS

Surgical procedures and viral injections

For recordings, adult male DAT-cre mice (n=10) were implanted with a custom-built 16-channel optrode mounted into a microdrive (Anikeeva et al., 2011). 16 NiCr-microwires were assembled into an electrode bundle, glued to a 250 µm large NA (0.66) optic fiber (Prizmatix) and mounted into the microdrive assembly. The tips were electroplated in gold solution containing multi-wall carbon nanotubes (1mg/ml)(Ferguson et al., 2009) to reach a resistance of 100-150 kΩ using the NanoZ impedance tester. The completed implant did not exceed 2g in weight. During the surgery, mice were anesthetized using isoflurane (~2%), the scalp was injected with lidocaine and the skull exposed. Three skull screws were inserted and a craniotomy performed above

the VTA. The dura was carefully removed and 400-600nl of cre-dependent ChR2 were slowly injected into the VTA (AP: -3.2 ± 0.2; ML: 0.8 ± 0.2; DV: -4.2). Then, the optrode was lowered slowly into the brain using a micromanipulator and implanted just above the VTA (AP: -3.2 ± 0.2; ML: 0.8 ± 0.2; DV: -3.8 to -4.0). The space between the brain surface and the implant was first sealed using silicone elastomer (Kwik-Cast, WPI), then secured using super glue and dental cement. The ground electrode was inserted into the posterior part of the contralateral hemisphere devoid of brain vessels.

For optogenetic causality experiments, adult male and female DAT-cre mice (n=12) underwent the same surgical procedures. Then they were injected and implanted bilaterally into the VTA (AP: -3.2; ML: ±1.75; DV: -4.2 injection; DV: -4.0 fiberoptic implant; at an angle of 20°) with either 400-600nl of cre-dependent ChR2 (n=5), eArch3.0 (n=3) or eYFP (n=4) and large NA fiber optic implants (Plexon).

Behavioral training

After the surgery, animals were given at least one week to recover. For recording experiments, DAT-cre mice were progressively food-restricted down to 1 pellet of chow per day to reach approximately 85-90% of their initial weight. During one habituation session in the behavioral apparatus (MedAssociates) they located the drinking spout and had free access to the fat solution (solution of lipofundin 5% in water). Then mice started pre-training to associate a light cue (CS) with the availability of a liquid reward (US). If during the reward window mice licked against a custom-made piezo-based lickometer, they obtained a single drop of liquid reward. The analog piezo lickometer signal was sent to an Arduino board which applied a threshold and simultaneously sent a digital signal to the MedAssociates software and to our electrophysiology recording system for high temporal resolution timestamping of lick detection (Plexon). Lick bursts were defined as starting when the inter-lick interval (ILI) dropped below 0.5s and ended once the ILI exceeded 2s. The duration of the CS was 4s and the reward window lasted 4s. Both remained unchanged across all training stages. During pre-training, the average random inter-trial interval between two consecutive CSs was increased across pre-training sessions from an average of 45s to 65s during the last session. Animals underwent pre-training until they exhibited good success rates (>0.8) of US per CS while simultaneously reducing their licking between subsequent cues. During both, the pre-training and the spatial

task, AnyMaze (Stoelting) or CinePlex (Plexon), were used to online video track the animal's position (defined by its center of gravity). A 4x4 cm zone was defined inside the 22x22 cm operant chamber. Digital signals were sent to MedPC (MedAssociates) to trigger the light cue and reward delivery, if the mouse spent 2s inside the reward zone as required by the task. Additionally, all zone entries, licks, CSs and USs were recorded with high temporal resolution by our electrophysiology recording system.

Electrophysiological recordings and spike sorting

We started recordings during the last pre-training sessions when mice reached the criterion to move to the spatial task. If a neuron exhibited DA-like firing patterns or responded to the opto-identification protocol, the mouse was switched to the spatial task on the next day. Neuronal activity was recorded at 40 MHz sampling rate and all behavioral variables timestamped by our recording system (Plexon, Omniplex Neural Data Acquisition System). To allow for unrestrained locomotion and to reduce weight, we used a dual LED and 16-channel commutator (Plexon) and the distance between the headstage and the commutator was optimized for each animal. Recording sessions lasted 60-90 minutes, simple online sorting was performed to monitor recording quality. For analysis, offline spike sorting was performed using WaveClus (Quiroga et al., 2004). A frequency of 150Hz was used to separate the wideband signal into LFP and spike bands. Positive or negative thresholds were applied depending on the observed waveforms. Spike sorting results were visually inspected, the optimal temperature parameter in WaveClus selected and minimal manual correction applied. Only units with a L-ratio<0.05 were further analyzed (Hill et al., 2011). After each session, the microdrive was used to lower the optrode by ~50µm and the same procedure applied on the next day. Once the electrode had crossed the VTA along the dorso-ventral direction, animals were killed and perfused to recover the location of the electrode track (Fig S1C). Cell-type specific expression of ChR2 was also verified (Fig S1B).

Optogenetic identification of single-units

After each recording session and before the optrode was lowered for the next day, we delivered blocks of 10 pulses of 5ms using a LED (Plexbright LED modules, Plexon driven by an A.M.P.I. Master-8 stimulator) at frequencies of 1, 2, 5, 10 and 20 Hz. Light intensities ranged from 1-5mW before the optrode. Light evoked spikes were detected during a 6ms window after light

onset. A single-unit was validated as an optogenetically identified DA neuron if its response rate was ≥0.8 and the correlation between the average light-evoked waveform and the average of all other spike waveforms produced during the entire recording session was ≥0.85.

Optogenetic causality experiments

For optogenetic manipulation experiments, mice were implanted and trained as described above. During the training to the spatial task, they were bilaterally connected to habituate to the fibers. The mice were recorded using AnyMaze (Stoelting Co.) or Cineplex software (Plexon) and events time-stamped to a synchronized recording system (Omniplex; Plexon). The behavioral apparatus was controlled by custom scripts in MedPC (MedAssociates).

Mice were not stimulated optogenetically until the test day and each mouse experienced the switch in reward zones only once during the test session. After the first 20 minutes of the regular spatial task on reward zone 1 (RZ1), the zone was automatically switched to reward zone 2 (RZ2) without any hint to the animal. After this switch, mice would receive the CS after 2s spent in RZ2 and optogenetic stimulation if they returned and stayed 2s in RZ1. For inhibition experiments, the light stimulus was delivered bilaterally by 2 LEDs (Plexbright LED modules; Plexon) mounted on a commutator and consisted of 500ms continuous light stimulation of 550nm light (power: 8-10mW) whereas in stimulation experiments 5Hz tonic firing was mimicked using 5 pulses lasting 5ms of 465nm blue light (power: 10-12mW). Both light stimulation protocols were delivered to the animal only if it stayed for 2s in the unrewarded RZ1 after the switch in location. Mice were given 20 minutes to locate RZ2 after which the test was finished.

Heat maps were generated by dividing the field of view into a 50x50 grid and accumulating all spatial coordinates into this grid. This matrix was then normalized, smoothed using a Gaussian filter and a hyperbolic arcsine function applied to enhance contrast. To normalize for different behavioral performances between animals, reward rates during the test day were normalized using the reward rates from the day preceding the test day.

Post-mortem histology & imaging

After the optrode had crossed the VTA, mice were sacrificed with a lethal injection of pentobarbital (150mg/kg) and perfused with 4% of paraformaldehyde in cold PBS. After fixation,

the optrode was slowly retracted using the microdrive and the brain extracted. Coronal brain slices of the VTA were cut 80-120-µm thick.

To quantify the specificity of ChR2-expression, slices were immunostained for tyrosine hydroxylase (TH). Briefly, slices were permeabilized with 0.1% Triton in TBS-Tween 20 (TBST), blocked using 1% BSA in TBST and incubated with primary anti-TH antibody (dilution 1:500; Millipore AB152). 24h later, slices were rinsed and incubated with secondary antibody (dilution 1:500; Millipore AP182C) diluted in 1% BSA TBST for 1-2h. After each step, samples were washed with TBST.

After mounting using Fluoroshield mounting medium with DAPI (Abcam), low-resolution imaging was performed using either a 5x/NA 0.25 or 10x/NA 0.45 objective in a slide scanner (Mirax or AxioScan Z1; Zeiss). High-resolution stacks were acquired in a confocal fluorescence microscope (Nikon A1r Spectral 40x/NA1.3 or Zeiss LSM800 using 40x/NA1.4). In recorded animals, the electrode track could easily be located in 1-3 brain slices and was mapped to the stereotaxic coordinates of the slice where the track was most prominent. For cell counting, confocal high-resolution imaging stacks (z-steps of 2.5µm) were acquired, then cells expressing ChR2 and those stained for TH were counted manually using ImageJ.

QUANTIFICATION AND STATISTICAL ANALYSIS

Event-related neuronal responses

For each neuron, spikes were binned into a regular time grid of 250ms time bins. Neurons with a mean firing rate <1Hz as well as maximal firing rate <3Hz were excluded leaving us with 214 neurons. For each neuron, we constructed five event-related profiles around CS (cue), US (valve), the mouse's licks, speed and acceleration. The increase or decrease of each time bin from the baseline (defined for each event as [-3.5;-3.25] for the cue, [2.5;3.0] for valve, [-1.0;-0.75] for lick, [-2.0;-1.75] for the speed and the acceleration) was quantified using an area under a receiver-operating characteristic curve (auROC) method (Cohen et al., 2012). auROC values are between 0 and 1, values <0.5 indicate a decrease, while values >0.5 reflect increases in firing relative to the baseline. To visualize the absolute neuronal firing rate, for each neuron we also computed the corresponding continuous spike density function (cSDF) by convolving the

spikes with an excitatory post-synaptic potential (ePSP) kernel (Wallisch et al., 2014) using the algorithm proposed by the Schall Lab (Hanes et al., 1995).

Clustering of neuronal firing patterns

Prior to the clustering, we applied a dimensional reduction to each set of auROC profiles using an independent component analysis (ICA) (fastICA Matlab package). Mean firing rate was added to the 18 combined ICA components for each neuron. Then, we applied a hierarchical clustering with the standardized Euclidean distance metric and ward linkage method. The clustering quality was assessed with the silhouette index and Cophenetic correlation coefficient. We identified the putative dopamine (DA) cluster by the typical phasic responses to CS and US and without using the optogenetic identification during the clustering. All optogenetically identified DA neurons (n=17) were assigned to this cluster, thus confirming its identity. Since our optrode was moved across the VTA, we further refined our set of neurons to those located in the VTA by selecting single-units only between the first and the last session containing a DA neuron. This resulted in 161 VTA neurons. We repeated the clustering procedure on the resulting set of VTA neurons which yielded the 9 clusters in Fig. 2D.

Action related modulation of neuronal spiking

To determine which neurons were significantly modulated by goal-directed actions (licking and locomotion), we analyzed peri-event time histograms around lick burst onsets, locomotion onsets and speed peaks (Alves et al., 2018). In a window of [-5;5] seconds around these time points, spikes were averaged in a sliding window of 100ms shifted by 1ms steps. The baseline was defined as the [-5;-4] time interval. If the neuronal activity during the response window ([-0.5;0.5] for locomotion onset, [0;0.5] for speed peak and lick burst onset) exceeded 3 standard deviations and remained above for at least 50ms, the neuron was considered significantly modulated by the action. When a minimum time interval was imposed between external events (CS and US) and action onsets, we first removed all movement-related time points closer than 1250ms to an external event (CS and US) and the same analysis was applied to the remaining behavioral events.

Definition of DA neuron firing modes

Since in the freely moving condition, we found elevated firing rates compared to head-fixed animals, we could not directly apply the 80/160ms rule (Grace and Bunney, 1984) to separate

spike bursts from tonic firing. We adapted the 80/160 rule by first calculating the mean inter-spike interval (ISI) across the entire session. The onset was then defined by the first interval shorter than a 1/3 of the mean ISI and the offset was detected by using an interval twice as large as the onset threshold. This resulted in burst detection thresholds similar to previous studies when DA neurons fired at 4-5Hz, but scaled for neurons with increased firing rates. All spikes not attributed to a spike burst were tagged as tonic spikes. To detect spike pauses, we plotted the histogram of ISIs of all tonic spikes (ISI_{tonic}), fitted a Gaussian in the interval [0;3] of this distribution and the time at which the difference between the fit and the actual distribution was largest, was used as the threshold for pause detection. If this threshold was smaller than $2<ISI_{tonic}>$, it was set to $2<ISI_{tonic}>$. In subsequent analyses, time bins were considered to be in a pause if they were between the spike starting a pause and the spike ending a pause.

Encoding analysis of the spiking

To determine which events and behaviors contributed most to the spiking output, we applied a Generalized Linear Model (GLM) to predict spiking using a linear-nonlinear-Poisson (LNP) model (Paninski et al., 2007).

To improve between session comparison, this analysis was restricted to a homogeneous subset of 21 sessions (117 out of the 161 neurons). The design matrix contained seven covariates, three events (cue, valve and zone) and four behaviors (licks, speed, acceleration and distance to reward) resampled using a 250ms bin size as for the spiking. The presence inside the reward zone was represented in a binary vector as were cue and valve events, indicated by binary events lasting 2 bins. The remaining continuous variables (licks, speed, acceleration and distance to reward) were normalized and z-scored. We used a sliding window of 10 bins to predict the spiking at a given time bin. To avoid over-fitting due to the large number of parameters used to predict the spiking, we resorted to a Lasso and elastic-net regularization (glmnet for Matlab). Pearson's correlation coefficient was used to compare the recorded and predicted spiking. To determine the contribution of each behavior, we used 3 different approaches (Fig. S3).

Firstly, for each neuron, the prediction was performed for each of the possible 127 combinations of the seven covariates (i.e. 7+21+35+35+21+1 from 1 to 7 covariates). In the design matrix, removed covariates were set as their mean over the entire session and the regressors

constructed from all seven covariates. We selected the combinations displaying the maximum correlation for each number of covariates and counted the number of times each covariate appeared in these best combinations. We estimated the occurrence probability of each covariate from this histogram by normalizing it with the total number of occurrences (from 1 to 7 covariate we get 1+2+3+4+5+6+7=28 possible occurrences). We also determined the minimum number of covariates (or contributors) necessary to cross 90% of the largest value between the previously computed maximum correlations.

Secondly, we looked at the relative loss in correlations after removing a single variable from the input to the model constructed from all seven covariates, by computing:

$$\text{Relative loss in correlation} = (\text{correlation}_{7\text{covariates}} - \text{correlation}_{6\text{covariates}})/\text{correlation}_{7\text{covariates}}.$$

Finally, we also computed the relative loss by considering only six covariates for the fit (i.e. constructing the design matrix and regressors with six covariates only).

To compare the difference between CS-US and outside CS-US, we extracted corresponding periods from recorded and predicted spiking and calculated the median correlations for these two sets. We temporally shuffled the neural signal 100x by a randomized time offset (50-500s) to check that the prediction is better than chance level (threshold of 2.326 standard deviations to the shuffled distribution) and only neurons meeting this criterion were used in the main figure.

Behavioral modules

To more precisely characterize behaviors during each session, we defined 7 modules based on the recorded behavioral variables: at rest, (slow) move, run, licks, rewarded licks, run toward zone and run toward reward. Transition points between resting and moving periods used a threshold of 2.1cm/s. We then computed the mean speed value for move intervals and used a move-run threshold of 2.8 cm/s thereby defining the resting, moving and running modules. Lick bursts were detected as described above, starting with a ILI $\leq$ 0.5s (lick burst onset) and ending when ILI $\geq$ 2s (lick burst offset). We defined rewarded lick bursts as those whose onsets were occurring between a CS and a US. All other lick bursts were considered as unrewarded lick bursts. Finally, we computed the distance to the reward and to the zone in order to further specify if run modules were directed towards the zone or towards the reward.

Comparing neural signal underlying correct and false trials

Correct trials lasted approximately 10s, the mouse entered the zone, waited for 2s in the zone until the CS turned on, then ran toward the valve where the US was delivered between 4-8s after the CS. The mouse spent a few seconds near the spout to consume the reward and then returned to the zone. We sought to detect failed trials that had motor output which was similar to correct trials but during which neither CS nor US were delivered. For example, mice frequently left the zone before the CS came on, ran toward the spout, licked but did not receive the reward. Using the behavioral modules defined above, we defined false trials with similar speeds and trajectories to correct trials. In a failed trial, we required the time from the run toward the zone to the run toward the valve to last at least 3s to avoid the case when mice just randomly crossed the zone and ran to the valve. To be included in failed trials, mice had to lick and test for reward delivery. Similarly, we used these same behavioral modules to identify correct trials and subsequently validated the procedure using the CS- and a US-events, showing that this procedure was capable of extracting correct and false trials based on similar motor output.

Using the definition of successful and failed trials, we then compared their associated neural signal (spike, tonic, burst spiking and pause) around 4 events of interest (zone entry, zone exit, lick burst onset and lick burst offset). We applied an auROC analysis (Brodersen et al., 2010) to compare the mean signal in false trials to activity in correct trials for the four events. The response windows for the different events were: zone entry [-0.75;1.5], zone exit [-0.5;1], lick burst onset [0;1] and lick burst offset [-0.75;0.25]). To determine if the relative change was significant, we performed a Wilcoxon two-sided rank sum test (α=0.05). While for DA neurons the analyses were carried out on all spikes, tonic firing, burst spiking and pauses, non-DA neurons were analyzed only with all spikes. All analyses were performed using custom written code in Matlab (Mathworks).

REFERENCES

Allen, W.E., Kauvar, I. V., Chen, M.Z., Richman, E.B., Yang, S.J., Chan, K., Gradinaru, V., Deverman, B.E., Luo, L., and Deisseroth, K. (2017). Global representations of goal-directed behavior in distinct cell types of mouse neocortex. Neuron *94*, 891–907.

Anikeeva, P., Andalman, A.S., Witten, I., Warden, M., Goshen, I., Grosenick, L., Gunaydin, L.A., Frank, L.M., and Deisseroth, K. (2011). Optetrode: a multichannel readout for optogenetic control in freely moving mice. Nat. Neurosci. *15*, 163–170.

Atallah, H.E.E., McCool, A.D.D., Howe, M.W.W., and Graybiel, A.M.M. (2014). Neurons in the ventral striatum exhibit cell-type-specific representations of outcome during learning. Neuron *82*, 1145–1156.

Barter, J.W., Li, S., Lu, D., Bartholomew, R. a., Rossi, M. a., Shoemaker, C.T., Salas-Meza, D., Gaidis, E., and Yin, H.H. (2015). Beyond reward prediction errors: the role of dopamine in movement kinematics. Front. Integr. Neurosci. *9*, 1–22.

Beier, K.T., Steinberg, E.E., DeLoach, K.E., Xie, S., Miyamichi, K., Schwarz, L., Gao, X.J., Kremer, E.J., Malenka, R.C., and Luo, L. (2015). Circuit architecture of VTA dopamine neurons revealed by systematic input-output mapping. Cell *162*, 622–634.

Brodersen, K.H., Ong, C.S., Stephany, K.E., and Buhmann, J.M. (2010). The binormal assumption on precision-recall curves. Proc. - Int. Conf. Pattern Recognit. 4263–4266.

Chang, C.Y., Esber, G.R., Marrero-Garcia, Y., Yau, H.J., Bonci, A., and Schoenbaum, G. (2015). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. Nat. Neurosci. *19*, 111–116.

Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature *482*, 85–88.

Costa, R.M. (2011). A selectionist account of de novo action learning. Curr. Opin. Neurobiol. *21*, 579–586.

Dreyer, J.K., Herrik, K.F., Berg, R.W., and Hounsgaard, J.D. (2010). Influence of phasic and tonic dopamine release on receptor activation. J. Neurosci. *30*, 14273–14283.

Ellwood, I.T., Patel, T., Wadia, V., Lee, A.T., Liptak, A.T., Bender, K.J., and Sohal, V.S. (2017). Tonic or phasic stimulation of dopaminergic projections to prefrontal cortex causes mice to maintain or deviate from previously learned behavioral strategies. J. Neurosci. *37*, 1221–17.

Ferguson, J.E., Boldt, C., and Redish, A.D. (2009). Creating low-impedance tetrodes by electroplating with additives. Sensors Actuators, A Phys. *156*, 388–393.

Fiorillo, C.D., Tobler, P.N., and Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. Science. *299*, 1898–1902.

Floresco, S.B., West, A.R., Ash, B., Moorel, H., Grace, A.A., Moore, H., and Grace, A.A. (2003). Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. Nat. Neurosci. *6*, 968–973.

Fuhrmann, F., Justus, D., Sosulina, L., Kaneko, H., Beutel, T., Friedrichs, D., Schoch, S., Schwarz, M.K., Fuhrmann, M., and Remy, S. (2015). Locomotion, theta oscillations, and the speed-correlated firing of hippocampal neurons are controlled by a medial septal glutamatergic circuit. Neuron *86*, 1253–1264.

Gadagkar, V., Puzerey, P.A., Chen, R., Baird-Daniel, E., Farhang, A.R., and Goldberg, J.H. (2016). Dopamine neurons encode performance error in singing birds. Science. *354*, 1278–1282.

Gilat, M., Lígia Silva de Lima, A., Bloem, B.R., Shine, J.M., Nonnekes, J., and Lewis, S.J.G. (2018). Freezing of gait: promising avenues for future treatment. Park. Relat. Disord. *52*, 7–16.

Ginis, P., Nackaerts, E., Nieuwboer, A., and Heremans, E. (2017). Cueing for people with parkinson's disease with freezing of gait: a narrative review of the state-of-the-art and novel perspectives. Ann. Phys. Rehabil. Med.

Grace, A., and Bunney, B. (1984). The control of firing pattern in nigral dopamine neurons: burst firing. J. Neurosci. *4*, 2877–2890.

Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Vander Weele, C.M., Kennedy, R.T., Aragona, B.J., and Berke, J.D. (2016). Mesolimbic dopamine signals the value of work. Nat. Neurosci. *19*, 117–126.

Hanes, D.P., Thompson, K.G., and Schall, J.D. (1995). Relationship of presaccadic activity in frontal eye field and supplementary eye field to saccade initiation in macaque: poisson spike train analysis. Exp. Brain Res. *103*, 85–96.

Hill, D.N., Mehta, S.B., and Kleinfeld, D. (2011). Quality metrics to accompany spike sorting of extracellular signals. J. Neurosci. *31*, 8699–8705.

Hiroyuki, N. (2014). Multiplexing signals in reinforcement learning with internal models and dopamine. Curr. Opin. Neurobiol. *25*, 123–129.

Hollerman, J.R., and Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. Nat. Neurosci. *1*, 304–309.

Howe, M.W., and Dombeck, D.A. (2016). Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature *535*, 505–510.

Howe, M.W., Tierney, P.L., Sandberg, S.G., Phillips, P.E.M., and Graybiel, A.M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature *500*, 575–579.

Keiflin, R., and Janak, P.H. (2015). Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. Neuron *88*, 247–263.

Klein, T.A., Neumann, J., Reuter, M., Hennig, J., von Cramon, D.Y., and Ullsperger, M. (2007). Genetically determined differences in learning from errors. Science. *318*, 1642–1645.

Lammel, S., Hetzel, A., Häckel, O., Jones, I., Liss, B., and Roeper, J. (2008). Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. Neuron *57*, 760–773.

Lammel, S., Ion, D.I., Roeper, J., and Malenka, R.C. (2011). Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. Neuron *70*, 855–862.

Lammel, S., Lim, B.K., Ran, C., Huang, K.W., Betley, M.J., Tye, K.M., Deisseroth, K., and Malenka, R.C. (2012). Input-specific control of reward and aversion in the ventral tegmental area. Nature *491*, 212–217.

Marcott, P.F.F., Mamaligas, A.A.A., and Ford, C.P.P. (2014). Phasic dopamine release drives rapid activation of striatal D2-receptors. Neuron *84*, 164–176.

Matsumoto, M., and Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. Nature *447*, 1111–1115.

Nakahara, H., and Hikosaka, O. (2012). Learning to represent reward structure: a key to adapting to complex environments. Neurosci. Res. *74*, 1–7.

Niv, Y., Daw, N.D., and Dayan, P. (2005). How fast to work: Response vigor, motivation and tonic dopamine. In Advances in Neural Information Processing Systems 18, pp. 1019–1026.

Niv, Y., Daw, N.D., Joel, D., and Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology (Berl). *191*, 507–520.

Nomoto, K., Schultz, W., Watanabe, T., and Sakagami, M. (2010). Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. J. Neurosci. *30*, 10692–10702.

Oster, A., Faure, P., and Gutkin, B.S. (2015). Mechanisms for multiple activity modes of VTA dopamine neurons. Front. Comput. Neurosci. *9*, 1–17.

Panigrahi, B., Martin, K.A.A., Li, Y., Graves, A.R.R., Vollmer, A., Olson, L., Mensh, B.D.D., Karpova, A.Y.Y., and Dudman, J.T.T. (2015). Dopamine is required for the neural representation and control of movement vigor. Cell *162*, 1418–1430.

Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. Prog. Brain Res. *165*, 493–507.

Pascoli, V., Terrier, J., Hiver, A., and Lüscher, C. (2015). Sufficiency of mesolimbic dopamine neuron stimulation for the progression to addiction. Neuron *88*, 1054–1066.

Paxinos, G., and Franklin, K.B.J. (2001). The mouse brain in stereotaxic coordinates 2nd edition (Academic Press).

Puryear, C.B., Kim, M.J., and Mizumori, S.J.Y. (2010). Conjunctive encoding of movement and reward by ventral tegmental area neurons in the freely navigating rodent. Behav. Neurosci. *124*, 234–247.

Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput. *16*, 1661–1687.

Sadacca, B.F., Jones, J.L., and Schoenbaum, G. (2016). Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework. Elife *5*, 1–13.

Saunders, B.T., Richard, J.M., Margolis, E.B., and Janak, P.H. (2018). Dopamine neurons create pavlovian conditioned stimuli with circuit-defined motivational properties. Nat. Neurosci. *21*, 1072-1083.

Schultz, W. (1986). Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. J. Neurophysiol. *56*, 1439–1461.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. J. Neurophysiol. *80*, 1–27.

Schultz, W. (2007). Multiple dopamine functions at different time courses. Annu. Rev. Neurosci. *30*, 259–288.

Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J. Neurosci. *13*, 900–913.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science. *275*, 1593–1599.

Shen, W., Flajolet, M., Greengard, P., and Surmeier, D.J. (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. Science. *321*, 848–851.

da Silva, J.A., Tecuapetla, F., Paixão, V., and Costa, R.M. (2018). Dopamine neuron activity before action initiation gates and invigorates future movements. Nature *554*, 244–248.

Soares, S., Atallah, B. V., and Paton, J.J. (2016). Midbrain dopamine neurons control judgment of time. Science. *354*, 1273–1277.

Starkweather, C.K., Babayan, B.M., Uchida, N., and Gershman, S.J. (2017). Dopamine reward prediction errors reflect hidden-state inference across time. Nat. Neurosci. *20*, 581–589.

Surmeier, D.J., Obeso, J.A., and Halliday, G.M. (2017). Selective neuronal vulnerability in parkinson disease. Nat. Rev. Neurosci. *18*, 101–113.

Syed, E.C.J., Grima, L.L., Magill, P.J., Bogacz, R., Brown, P., and Walton, M.E. (2015). Action initiation shapes mesolimbic dopamine encoding of future rewards. Nat. Neurosci. *19*, 34–36.

Tian, J., and Uchida, N. (2015). Habenula lesions reveal that multiple mechanisms underlie dopamine prediction errors. Neuron *87*, 1304–1316.

Tian, J., Huang, R., Cohen, J.Y., Osakada, F., Kobak, D., Machens, C.K., Callaway, E.M., Uchida, N., and Watabe-Uchida, M. (2016). Distributed and mixed information in monosynaptic inputs to dopamine neurons. Neuron *91*, 1374–1389.

Tobler, P.N., Fiorillo, C.D., and Schultz, W. (2005). Adaptive coding of reward value by dopamine neurons. Science. *307*, 1642–1645.

Tsai, H.-C.C., Zhang, F., Adamantidis, A., Stuber, G.D., Bond, A., de Lecea, L., Deisseroth, K., Bonci, A., de Lecea, L., Deisseroth, K., et al. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science. *324*, 1080–1084.

Wallisch, P., Lusignan, M., Benayoun, M., Baker, T., Dickey, A., and Hatsopoulos, N. (2014). MATLAB for Neuroscientists (Elsevier).

Watabe-Uchida, M., Zhu, L., Ogawa, S.K., Vamanrao, A., and Uchida, N. (2012). Whole-brain mapping of direct inputs to midbrain dopamine neurons. Neuron *74*, 858–873.

Xiao, C., Cho, J.R., Zhou, C., Treweek, J.B., Chan, K., McKinney, S.L., Yang, B., and Gradinaru, V. (2016). Cholinergic mesopontine signals govern locomotion and reward through dissociable midbrain pathways. Neuron *90*, 333–347.

Yagishita, S., Hayashi-Takagi, A., Ellis-Davies, G.C.R.R., Urakubo, H., Ishii, S., and Kasai, H. (2014). A critical time window for dopamine actions on the structural plasticity of dendritic spines. Science. *345*, 1616–1620.

Yttri, E.A., and Dudman, J.T. (2016). Opponent and bidirectional control of movement velocity in the basal ganglia. Nature *533*, 402–406.

Van Zessen, R., Phillips, J.L., Budygin, E.A., and Stuber, G.D. (2012). Activation of VTA GABA neurons disrupts reward consumption. Neuron *73*, 1184–1194.