

# The timing of action determines reward prediction signals in identified midbrain dopamine neurons

Luke T. Coddington<sup>ID</sup>\* and Joshua T. Dudman<sup>ID</sup>\*

**Animals adapt their behavior in response to informative sensory cues using multiple brain circuits. The activity of midbrain dopaminergic neurons is thought to convey a critical teaching signal: reward-prediction error. Although reward-prediction error signals are thought to be essential to learning, little is known about the dynamic changes in the activity of midbrain dopaminergic neurons as animals learn about novel sensory cues and appetitive rewards. Here we describe a large dataset of cell-attached recordings of identified dopaminergic neurons as naive mice learned a novel cue-reward association. During learning midbrain dopaminergic neuron activity results from the summation of sensory cue-related and movement initiation-related response components. These components are both a function of reward expectation yet they are dissociable. Learning produces an increasingly precise coordination of action initiation following sensory cues that results in apparent reward-prediction error correlates. Our data thus provide new insights into the circuit mechanisms that underlie a critical computation in a highly conserved learning circuit.**

The stomach of a hungry guest rumbles at the sound of a dinner bell; a child extends her hand for a piece of candy. Both sensory cues and initiation of actions can become associated with expectations about future outcomes. There are two broad conceptualizations of how associations between actions, events and their resultant outcomes are formed. The first is a gradual strengthening of an association between an action or a stimulus and a reliable outcome—often described as Hebbian learning<sup>1</sup>. The second is a gradual reduction in errors of prediction through comparison of expected and observed outcomes—often a critical component of reinforcement learning<sup>2</sup>. In a well-known example of such a representation, the activity of mammalian midbrain dopaminergic (mDA) neurons correlates with changes in reward expectation—or reward-prediction errors (RPEs)—following associative learning<sup>3</sup>. This has provided direct evidence in support of the notion that reinforcement learning proceeds through a progressive reduction in error that is represented in the RPE correlates of mDA neurons.

RPE correlates have been observed in mDA neurons almost exclusively within a specific experimental condition: well-trained animals learning new associations through the introduction of new cues or altered contingencies<sup>4–11</sup>. Behavioral performance and mDA reward signals can adapt to these new contingencies within tens of trials in rodents<sup>11</sup> and monkeys<sup>4,5</sup>. However, it has also been observed that mDA neuron correlates can lag behind adaptive changes in behavior<sup>5,7,11</sup>, calling into question the causal role of RPE correlates in the learning of associations. Additionally, even in well-studied overtrained conditions, the circuit mechanisms by which RPEs are computed remains unclear<sup>12</sup>. This could be due, at least in part, to the tight coordination of multiple learning systems that are characteristic of the trained state<sup>13,14</sup>.

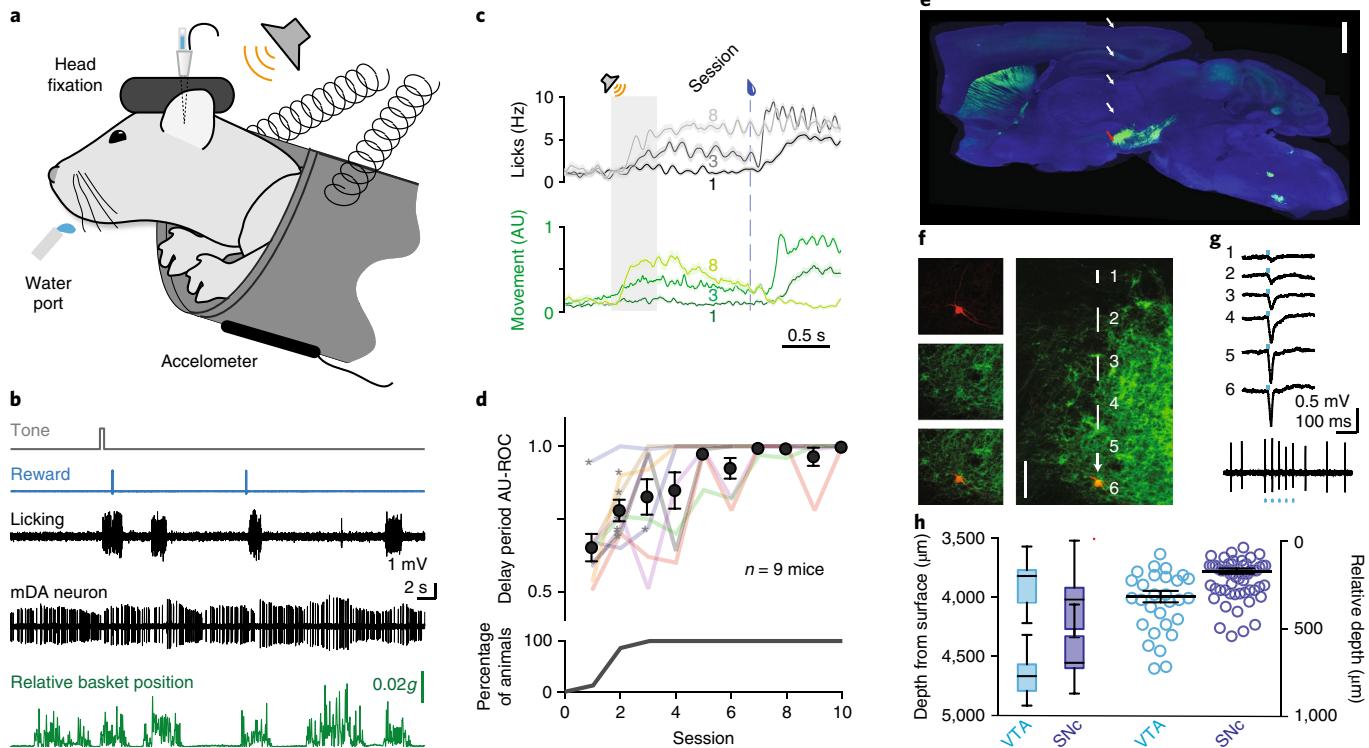
In sharp contrast to the rapid adaptation to changed contingencies in trained animals, naive animals learning a novel association may require many hundreds of trials for stable or asymptotic behavior to be observed—especially for trace-conditioning experiments. These more gradual adaptive changes in behavior allow for more observations of inherently probabilistic neural activity and behavior. Influential models of the RPE computation make quantitative

predictions about the emergence of RPE correlates that are reflected in mDA neuron activity during such novel learning<sup>3,15</sup>. However, to date, empirical data to quantitatively compare to these predictions is lacking. Moreover, whereas novel learning requires intact mDA signaling<sup>16</sup>, necessity has not been demonstrated for many aspects of adaptive changes in behavior following extensive training. Thus, we reasoned that observing the dynamics with which mDA neuron activity changes over training could provide important constraints on the possible circuit mechanisms that underlie RPE correlates.

Here we examine movement-related mDA activity as naive (but habituated) mice are first trained in a Pavlovian trace-conditioning experiment. We use optogenetic identification to unambiguously identify mDA neurons from two major midbrain populations in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc). By studying the initial mDA response in naive animals through responses to self-initiated movements, sweetened water rewards and sensory cues in well-trained mice, we describe the time courses with which each correlate emerges. We replicate the observation that RPE correlates become apparent in both VTA and SNc mDA neurons only after learning. We find that this late emergence is due to the fact that RPE correlates are a consequence of temporal integration of independent reward expectation signals that are associated with reward-predictive sensory cues and the initiation of appetitive actions.

## Results

Adult transgenic mice expressing a light-activated opsin, channelrhodopsin-2<sup>17</sup> (Chr2), in mDA neurons were exposed to a trace-conditioning framework in which a 0.5-s long auditory tone was presented 1.5 s before delivery of a sweetened water reward (Fig. 1a,b; see Methods). We chose to use a head-fixed framework<sup>18</sup> similar to canonical studies on representations in mDA neurons<sup>3,19</sup> and because of the desirable reduction in variability of behavior. Although no action was necessary to receive water, anticipation of available water is often accompanied by body movements (readily observed in freely moving animals) that are not typically assessed in head-fixed mice. To enable such measurements, we placed mice



**Fig. 1 | Juxtacellular recording from identified mDA neurons in awake, behaving mice.** **a**, Schematic of the head-fixed behavioral apparatus. Mice were head-fixed in front of water port positioned within tongue length and mice were supported by a spring-suspended basket monitored by an accelerometer. Small droplets of sweetened water ( $\sim 3\text{ }\mu\text{l}$ ) were delivered by opening a valve on a gravity-fed line; a suspended drop of liquid stayed on the port until it was collected by licking. **b**, Auditory stimuli, opening of a solenoid water valve (reward), vibration of the water port due to licking (licking), and juxtacellular recordings from identified mDA neurons could be recorded simultaneously with bodily movements that displaced basket position relative to its resting position (relative basket position). **c**, Mean licking (top) and basket displacement (bottom) in response to an auditory cue (shaded gray bar) and sweetened water delivery (dashed line) for the same mouse at training sessions 1, 3 and 8;  $n = 50$  trials for each trace. AU, arbitrary units. **d**, Top, area under the ROC curve analysis (AU-ROC) to assess whether the number of licks differed significantly from baseline (the first session to reach significance at  $P < 0.05$  for each animal is indicated with an asterisk) during the 1-s delay between cue and reward. Colored lines represent individual mice; data are shown as mean  $\pm$  s.e.m. Bottom, percentage of animals with  $P < 0.05$  for area under the ROC curve analysis as training progresses. **e**, mDA neurons in *Dat<sup>Cre/Cre</sup>;Ai32* mice selectively express a ChR2-eYFP fusion protein. A representative histological section shows labeling of mDA soma in the midbrain and terminals in the striatum (see Supplementary Fig. 1). White and red arrows indicate the trajectory of a recording pipette and approximate location of recorded neuron for the example experiment shown in **f**, **g**. Scale bar, 1 mm. **f**, Left, individual and merged images from pilot experiment (1 of 5 similar results) of a neurobiotin-labeled neuron (red) and eYFP-expressing mDA neurons (green). Right, pipette trajectory with locations of light stimulations numbered. Scale bar, 0.1 mm. **g**, Example traces of light-evoked field potentials corresponding to the recording locations (1–6) shown in **f**, demonstrating tight correspondence between ChR2-eYFP expression (green) and evoked field-potential amplitude. Bottom trace shows resulting juxtacellular recording with tight entrainment of action potentials to the light stimulus (optogenetic tagging). **h**, Spatial distribution of both VTA (cyan) and SNC (dark blue) optogenetically tagged mDA neurons relative to the dorsal-most position at which a light-evoked field potential was observed (left). Box-and-whisker plots show interquartile range (box), mean (center line), and minimum and maximum values (whiskers).

in a spring-suspended basket equipped with an accelerometer (Fig. 1a; see Methods). Body movements were correlated with, but not synonymous with, bouts of licking (Fig. 1b)—confirming that body movement can provide an additional metric to track learned behaviors in head-fixed mice.

Learning was characterized by the emergence of conditioned changes in behavior both to the delivery of water and to the presentation of the tone. With increasing experience, mice exhibited increasingly rapid reactions to the presence of available water (Fig. 1c). The mean reaction time to collect the water was monotonically reduced from hundreds of milliseconds in naive animals to a latency of  $\sim 60$  ms ( $\sim 1/2$  licking cycle) after  $\sim 1,000$  trials with water delivery. This decrease in the latency to collect available water was correlated with a monotonic increase in the number of licks at the water spout during the delay between tone offset and water delivery

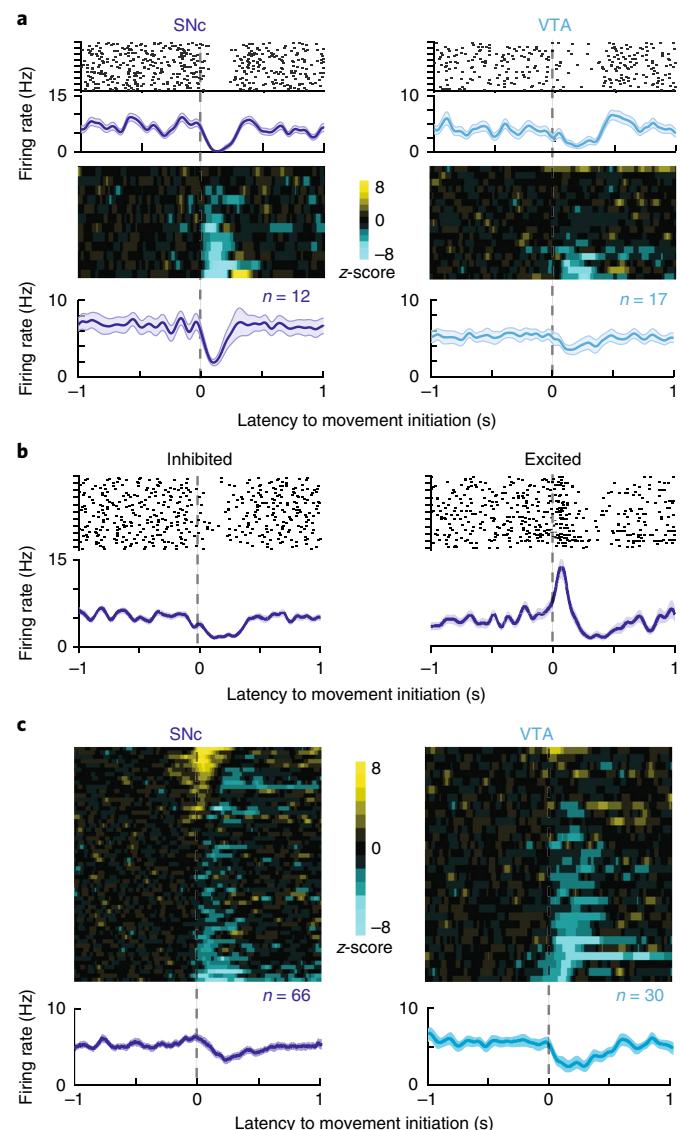
(Fig. 1c,d; latency to initiate licking following tone,  $r = -0.50$ ,  $P < 0.0001$ ; licks during tone-reward interval,  $r = 0.63$ ,  $P < 0.0001$ ). In addition, body movements became increasingly stereotyped and reliable (Fig. 1c; relative difference in basket displacement in response to the tone and reward,  $r = -0.43$ ,  $P < 0.0001$ ). Whereas mice overtrained on this same task rapidly reacquire and extinguish responses to the tone<sup>7</sup>, naive learning is characterized by an initial emergence of learned behavior ( $\sim 3$  sessions;  $\sim 100$  trials per session) and a gradual asymptotic stabilization of multiple aspects of learned behavior.

Given the results of behavioral analysis during the first  $\sim 1,000$  trials of training, we next investigated how responses to mDA neurons emerge over this same period. We obtained cell-attached electrophysiological recordings (Fig. 1e–h; see Methods) from optogenetically identified ('optotagged'<sup>20,21</sup>) mDA neurons in the

VTA ( $n=47$ ) and SNC ( $n=88$ ) of naive mice as they learned an auditory trace-conditioning task. Localization of the SNC or VTA on each penetration was determined by observing the magnitude of an optogenetically evoked field potential (Fig. 1g) known to reflect the presence of mDA neurons<sup>22</sup>. Cell-attached, loose-seal recordings were obtained in the region of the largest evoked potential (Fig. 1f,g) and mDA neuron identity was confirmed by optically driving trains and bursts of action potentials (Fig. 1g). The range of depths over which evoked potentials and positively identified mDA neurons were found agreed well with the anatomy of the VTA and SNC (Fig. 1h). Moreover, identified mDA neurons had properties that agreed with canonical electrophysiological criteria and post hoc confirmation of high-penetrance labelling of mDA neurons (Supplementary Fig. 1a,b).

**Phasic modulation of mDA activity at initiation of body movements and licking.** Recent studies have found that activity in at least some populations of mDA neurons correlates with movement initiation in naive or untrained animals<sup>23–25</sup>. In trained animals, the release of dopamine in the ventral striatum can correlate with the initiation of reward-related actions<sup>26,27</sup>. Therefore, we first examined activity in naive animals that were acclimatized to the head-fixed context, but before they had received any rewards in that context. In the absence of informative stimuli or expectation of water rewards, we observed significant inhibition in 10 of 12 mDA neurons from the SNC and 9 of 17 mDA neurons in the VTA upon initiation of body movement. Only one mDA neuron in the VTA and none in the SNC showed significant excitation (Wilcoxon signed-rank test, modulation window versus baseline, see Methods; Fig. 2a). Reports of excitatory peri-movement responses in mDA neurons<sup>23,25</sup> led us to more carefully examine activity around the time of movement initiation. We measured the interspike interval just before movement-related pauses in activity (see Methods). This analysis revealed an apparent ‘covert’ excitation in the form of a spike phase advance that is present in mDA neurons from both the SNC and VTA (Supplementary Fig. 2a,b). Thus, mDA neurons are modulated by a sequence of (covert) excitation followed by a more dominant inhibition prior to movement initiation in mice acclimated to head fixation.

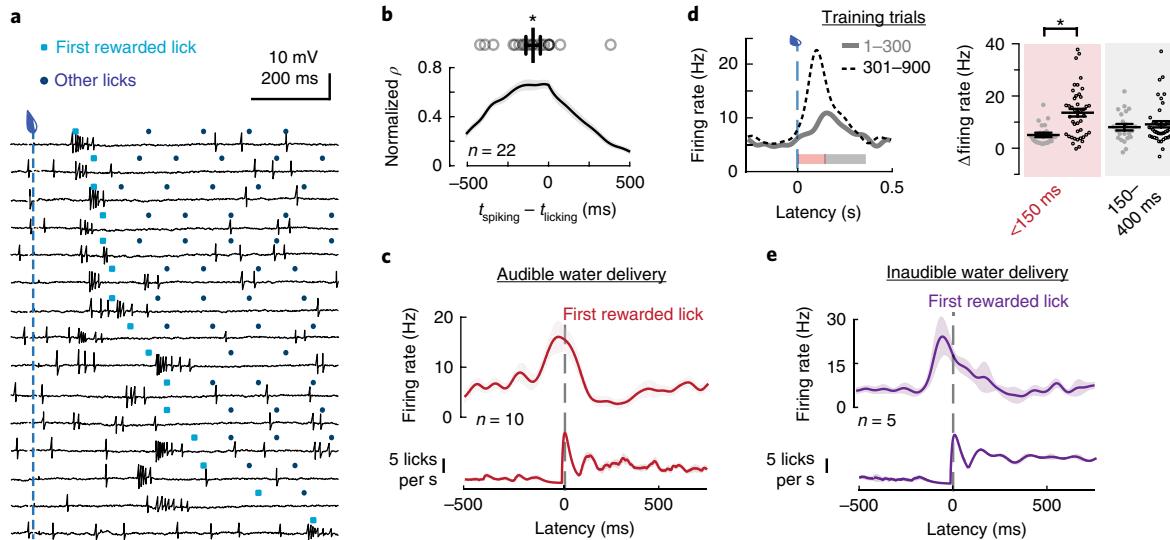
Given evidence for a more prominent positive modulation around purposive action initiation in rodents<sup>27–30</sup> and primates<sup>31</sup>, we next examined whether weak excitation observed in naive mice might be altered by the presence of available water rewards. Following the introduction of mice to auditory trace conditioning, we again examined peri-movement mDA neuron activity. We first considered movement bouts that were initiated during the long intertrial intervals that were well-separated from either auditory cues or water delivery. Such self-initiated movements occurred within the intertrial intervals in bouts (bout length:  $3.9 \pm 1.5$  s) separated by periods of stillness (interbout interval:  $6.4 \pm 1.9$  s). Around half ( $57 \pm 5\%$ ) of self-initiated movement bouts were accompanied by licking. In contrast to the net inhibition of mDA neuron activity in the naive context, in a rewarded context we observed more significant excitation of individual mDA neurons right before self-initiated movement onset (23 of 96 during training versus 1 of 29 pre-reward training,  $P=0.01$ ; SNC, 19 of 66 during training versus 0 of 12 pre-training,  $P=0.03$ ; VTA, 4 of 30 during training versus 1 of 17 pre-training,  $P=0.6$ , Fig. 2b,c). Thus, SNC dopaminergic neurons were more likely to be excited at movement initiation than VTA dopaminergic neurons, although substantial excitation was observed from individual neurons in both populations. As in naive animals, net inhibition at movement initiation was still common in both mDA neuron subpopulations (SNC, 34 of 66; VTA, 17 of 30) and appeared to be unchanged in its magnitude (Fig. 2c). Also similar to naive animals, covert excitation was present in non-excited neurons in the form of a spike phase advance right before



**Fig. 2 | Peri-movement excitation of mDA neurons, but not inhibition, depends on reward context.** **a**, Top, raster plots and peri-event time histograms (PETHs) of example SNC (left, dark blue) and VTA (right, cyan) dopaminergic neurons. Middle, heat maps of z-scored PETHs for the two populations. Bottom, mean PETHs recorded after acclimatization to the recording rig but before animal received sweetened water rewards. PETHs are aligned to the beginning of self-initiated movements. **b**, Single example mDA neurons from the SNC with clear negative (left) and positive (right) modulation of firing around self-initiated movement onset are shown for epochs during intertrial intervals after training had begun. **c**, Heat maps and PETH for population data as in a, but for all neurons recorded over the course of associative trace conditioning. **a–c**, Shaded areas indicate s.e.m.

movement-related pauses (Supplementary Fig. 2c). Together, these results suggest that a majority of mDA neurons receive both excitation and inhibition signals upon movement initiation; however, excitation can be enhanced by the introduction of rewards, whereas inhibition is not.

We next examined within-trial modulation of mDA neuron activity as naive mice first began to consume water rewards in our training context (sessions 1–3;  $n=22$  mDA neurons;  $n=6$  mice). Consistent with activity around self-initiated movements, we found that modulation of mDA activity after water delivery appeared to be



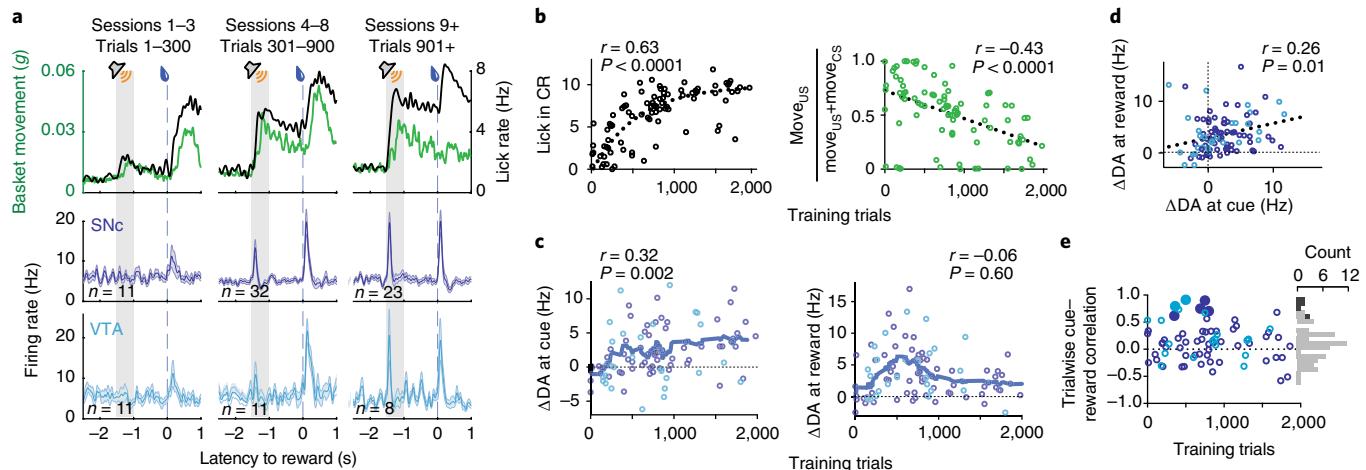
**Fig. 3 | Initial mDA reward responses encode reward-related movements.** **a**, Raw cell-attached recording from a mDA neuron in the SNC aligned to water delivery during the first 30 training trials experienced by the mouse. Trials are sorted according to the latency until the first lick (cyan) to collect the sweetened water droplet from the water port. **b**, Trial-by-trial normalized cross-correlation between lick rate and mDA neuron activity in the 1 s surrounding reward delivery ( $n=22$  neurons recorded early in learning, data are mean  $\pm$  s.e.m.). Points of maximum correlations are shown at the top (circles; with mean  $\pm$  s.e.m.). Lag was  $-96 \pm 44$  ms. \* $P=0.04$ ; two-sided Student's *t* test. **c**, Mean firing rate (top) and lick rates (bottom) aligned to the first contact of the tongue to the lick port after water delivery for 10 (of 22) mDA neurons with significant responses to water delivery (as determined by one-tailed Wilcoxon signed-rank test,  $P < 0.05$ ) early in training (sessions 1–3). Shaded areas indicate s.e.m. **d**, Left, mean firing rates aligned to water delivery for mDA neurons recorded in the first 300 trials of training (gray,  $n=22$ ) or the next 600 trials of training (black dashed,  $n=43$ ). Right, mean modulation of firing rate for mDA neurons recorded in trials 1–300 (gray points) versus trials 301–900 (black points) in the time windows indicated in pink (1–150 ms post-reward; two-sided Student's *t* test, \* $P=0.0002$ ) and gray (151–400 ms post-reward, two-sided Student's *t* test,  $P=0.6$ ) bars drawn below traces on the left. Black bars indicate mean  $\pm$  s.e.m. **e**, Data from two additional mice trained with an inaudible solenoid (see Supplementary Fig. 3) controlling water delivery. Mean firing rate for 5 (of 10) mDA neurons with significant responses to water delivery (as determined by one-tailed Wilcoxon signed-rank test,  $P < 0.05$ ) recorded early in training (session 1–3). Shaded areas indicate s.e.m.

associated with the initiation of consumptive licking (Fig. 3). This was particularly apparent in mDA neurons that were recorded in mice with variable latencies to initiate consumption (Fig. 3a). Mean mDA activity peaked prior to the first contact of the tongue to the water port (Fig. 3b,c), indicating that activity was related to movement initiation rather than sensation of the water. In trained animals, mDA neurons have a highly characteristic latency of response to auditory sensory cues—around 50 ms onset, 85 ms peak<sup>7</sup>. However, early in training we observed little specific modulation of mDA activity until ~150 ms after water delivery (Fig. 3a,d). This suggests that the audible solenoid valve did not provoke a sensory or ‘salience’ response in mDA neurons. As training progressed a large additional component of the mDA reward response emerged that was time-locked to reward delivery (within the first 150 ms post-reward,  $P=0.0002$  trials 1–300 versus trials 301–900; 151–400 ms post-reward,  $P=0.6$ ; Fig. 3d). Also consistent with the lack of a strong sensory component to mDA activity at this early stage of learning, we observed no response to the louder auditory cue (audible tone, 70 dB; solenoid valve, 52 dB; background noise, 55 dB; see Fig. 4) and none of our stimuli reached the intensities previously reported to produce salience responses (~90 dB<sup>3,2</sup>).

If it is the case that the modulation of mDA activity after reward delivery does not have a prominent sensory component early in training, we reasoned that substantially reducing a salient sensory component of water delivery—activation of the solenoid valve—would not diminish the mDA response. Therefore, we recorded mDA neurons in a subset of mice using a silent solenoid that was undetectable across the ultrasonic frequency range (Supplementary Fig. 3). The phasic modulation of mDA neuron activity with an undetectable solenoid remained similarly well-aligned to the initiation of consumptive licking (Fig. 3e). Notably, in primary thirst and

hunger centers in the rodent brain, even in naive animals, modulation of activity can be predictive of consumption rather than a reflection of physical, sensorial contact with water<sup>33,34</sup>. Our data therefore suggest that a movement initiation-associated phasic excitation–inhibition sequence is the earliest reward-related modulation of activity in SNC and VTA mDA neuron populations during novel learning.

**Emergence of sensory cue-related activity in mDA neurons during early learning.** We next sought to characterize the emergence of sensory cue-related activity of mDA neurons during early learning. Consistent with a large number of previous studies<sup>3</sup>, mDA neurons lacked a coherent response to the auditory cue early in learning, but developed a robust, monotonically increasing phasic response with training ( $r=0.32$ ,  $P=0.002$ ,  $n=96$ , Fig. 4a–c). Notably, the proportion of mDA neurons with a robust phasic response at the time of water delivery also increased over the first three sessions (1–3 sessions, 10 of 22; 4–8 sessions, 38 of 43, 9+ sessions, 27 of 31;  $P < 0.001$ ; Fig. 3d). In contrast to responses to the auditory cue, response amplitude at the time of water delivery was not significantly correlated with the number of trials experienced ( $r=-0.06$ ,  $P=0.6$ ,  $n=96$ ; Fig. 4c). The inference from existing data and influential models is that mDA neuron activity should directly ‘transfer’ from a reward to an earlier, predictive stimulus<sup>6,8,15,19,35–37</sup>. By contrast, our study—in which we examine the time course of the emerging mDA response to a predictive cue in naive animals—suggests that mDA responses to the predictive cue and the reward do not exhibit the predicted, negative trial-by-trial correlation<sup>15,36</sup>. Rather, auditory cue and water delivery responses had a small but significant positive correlation in individual mDA neurons across training ( $r=0.26$ ,  $P=0.01$ ,  $n=96$ ; Fig. 4d). Finally, we selectively



**Fig. 4 | mDA neuron responses to predictive cue and reward stimuli evolve independently during acquisition learning.** **a**, Behavioral and mean mDA firing rate aligned to the opening of the water valve and divided in to early (sessions 1–3, left) middle (sessions 4–8, middle) and late (sessions 9+, right) training sessions. Conditioned responses were assessed by measuring relative basket position (basket movement) and lick rate. Smoothed PETHs for identified mDA neuron populations recorded in the SNC (dark blue) and VTA (cyan). Shaded areas indicate s.e.m.; number of cells as labeled. **b**, Behavioral learning measures during each individual neuron recording period (*n*=96). Both the number of licks during the cue-reward trace interval (licks during delay period, left) and basket movement (fraction of movement following the reward) were significantly correlated with the number of training trials experienced at the time of recording. Results of Pearson's correlations are shown. **c**, The mean modulation of mDA neuron activity following the predictive tone (left, *n*=96) or reward (right, *n*=96) were plotted as a function of the number of training trials experienced at the time of recording, with 10-sample rolling mean overlaid (blue line),  $y=0$  is highlighted as thin-dashed lines. The SNC (dark blue) and VTA (cyan) populations are shown. Results of Pearson's correlation are shown. **d**, Mean modulation of mDA neuron firing rate in response to predictive tone and reward stimuli are positively correlated (dotted line represents best-fit trend). **e**,  $\rho$  values for within-neuron trial-by-trial Pearson's correlations between cue responses and reward responses for neurons with significant modulation by cues or water delivery (*n*=75). Only filled circles were significant at  $P < 0.05$ . Inset, histogram of cue-reward correlation  $\rho$  values, with significant correlations shaded dark gray.

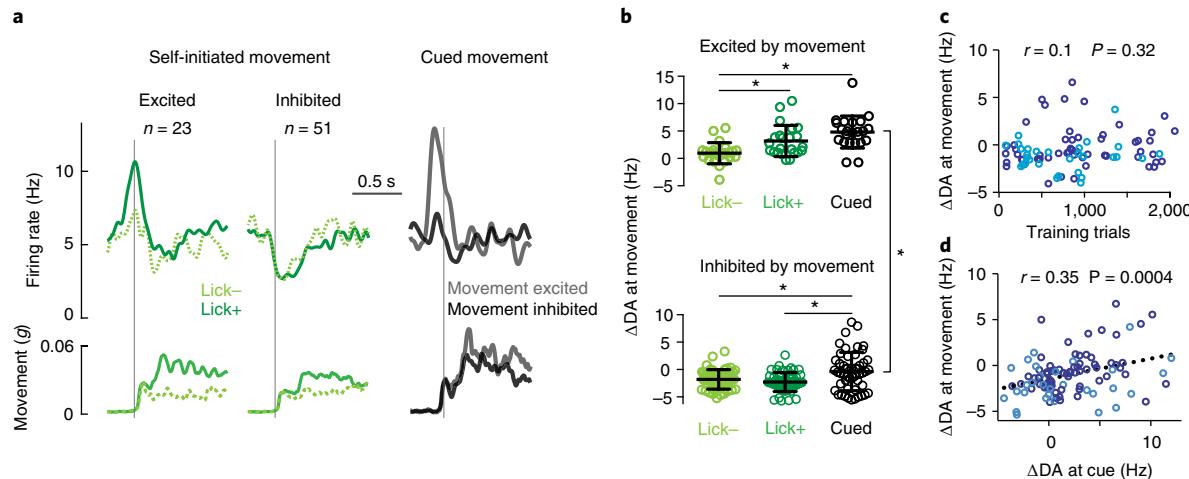
examined the relationship between cue and reward signals on a trial-by-trial basis within neurons that had significant sensory cue-related responses. Only 6 neurons of 75 analysed neurons showed significant correlations between auditory cue and reward responses and the correlations in these neurons were positive (Fig. 4e). Therefore, mDA responses to predictive sensory cues and water delivery appear to emerge independently while being scaled by a common cell-specific factor<sup>32,38</sup> (Supplementary Fig. 4).

We have described two apparently independent components of the phasic responses of mDA neurons: a movement-related excitatory-inhibitory sequence (Figs. 2 and 3) and phasic excitation associated with sensory stimuli that emerges as mice begin to exhibit learned behavior (Figs. 3d and 4a–c). In the sensory cue-related component, excitation is specifically enhanced as a function of the extent of learning (Figs. 3d and 4c). For the movement initiation-related component, excitation was pronounced once water rewards were first available even for self-initiated movements in the absence of cues or water rewards.

**Movement initiation-related activity of mDA neurons reflects reward expectation.** mDA movement-related activity was specific to the moment of movement initiation, as large accelerations within bouts of movement were not encoded by mDA neurons (Supplementary Fig. 5). We next sought to clarify whether the magnitude of movement initiation-related mDA excitation reflects the expectation of reward for the mice. We divided periods of self-initiated movement into those accompanied by licking (Lick+) and those without (Lick−), and examined mDA activity in neurons significantly excited or inhibited around movement initiation (Fig. 5a). We then compared these self-initiated actions to movements initiated following the reward predictive auditory cues ('cued movement'; Fig. 5a). We reasoned that this analysis should distinguish categories of action with relatively lower (Lick−), higher (Lick+) or the highest (Cued) degrees of reward expectation. A two-way

ANOVA (movement type  $\times$  movement response type) confirmed a main effect of movement response type (excited versus inhibited at movement,  $F_{1,216} = 149.8$ ,  $P < 0.0001$ ) as well as a main effect of movement type (Lick− versus Lick+ versus Cued,  $F_{2,216} = 18.1$ ,  $P < 0.0001$ ) and an interaction ( $F_{2,216} = 5.6$ ,  $P = 0.004$ ). Post hoc comparisons found that, in mDA neurons positively modulated by movement initiation, Lick+ movements were accompanied by greater excitation than Lick− movements ( $3.2 \pm 0.6$  versus  $1.0 \pm 0.4$  Hz,  $P = 0.036$ , *n*=23), consistent with the proposal that the excitatory action initiation-related component of mDA responses is sensitive to reward expectation. By contrast, inhibition that was apparent in negatively modulated mDA neurons was independent of expectation ( $-2.3 \pm 0.3$  versus  $-1.8 \pm 0.3$  Hz,  $P = 0.92$ , *n*=51; Fig. 5a,b). Furthermore, modulation of mDA neuron activity around cued movement was significantly greater than self-initiated movements in both response classes (positively modulated,  $4.8 \pm 0.6$  Hz,  $P < 0.0001$  versus Lick−; negatively modulated,  $-0.4 \pm 0.5$  Hz,  $P = 0.05$  versus Lick−,  $P = 0.002$  versus Lick+; Fig. 5a,b), with the exception that in movement-excited neurons, cued movement responses were not significantly greater than self-initiated Lick+ responses ( $3.2 \pm 0.6$  versus  $4.8 \pm 0.6$  Hz,  $P = 0.25$ ; Fig. 5b). Notably, responses to self-initiated movement were not correlated with training (Fig. 5c;  $r = 0.1$ ,  $P = 0.32$ ), but were correlated with their response to auditory cues (Fig. 5d;  $r = 0.35$ ,  $P < 0.001$ ), as can be appreciated in the blunted cue responses of mDA neurons that were significantly inhibited around movement (Fig. 5a, right;  $P < 0.0001$ ). Together, these results suggest that mDA neuron activity reflects the summation of action initiation and reward-predictive sensory cue inputs.

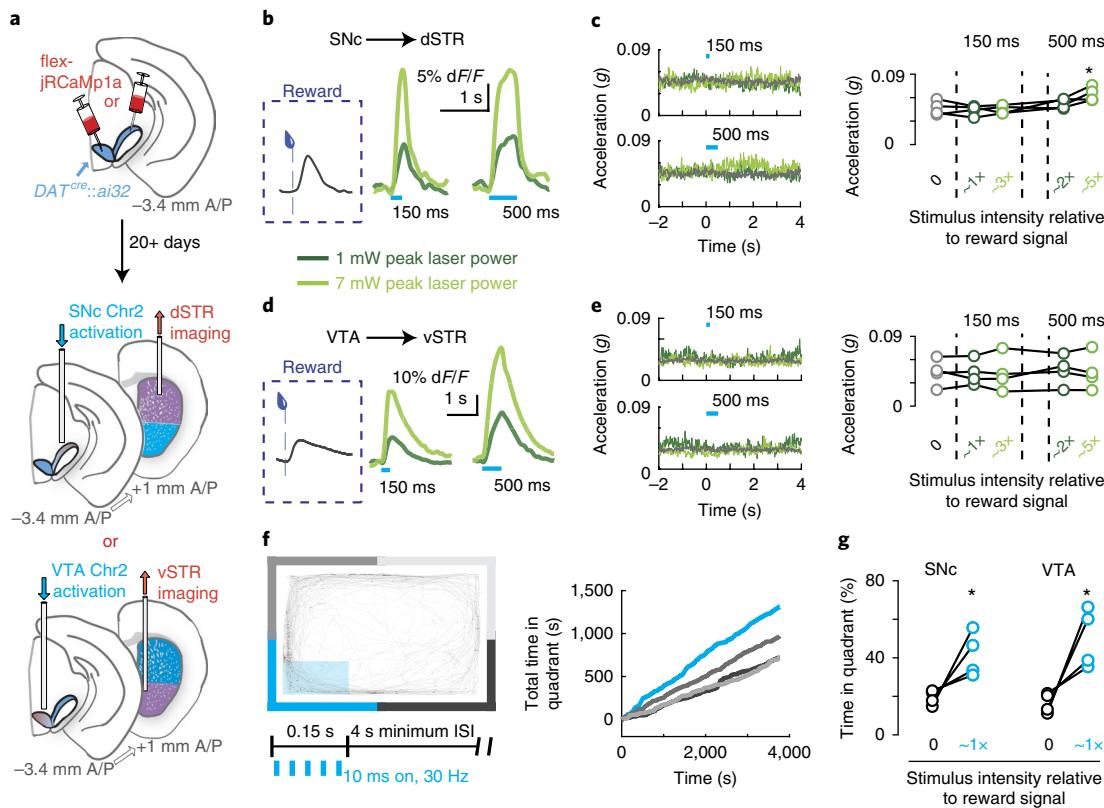
Excitation and inhibition were balanced such that there was a negligible self-initiated movement signal on average and net inhibition during movements in naive mice (Fig. 2). This indicates that in our experiment, mDA activity is not obligatory for movement. However, it has recently been proposed that excitation in select SNC mDA neurons is specifically sufficient<sup>24</sup> or perhaps even necessary<sup>25</sup>



**Fig. 5 | Peri-movement activity reflects reward expectation and sums with cue responses.** **a**, PETHs aligned to initiation time of movements from the intertrial intervals (self-initiated) or following reward-predictive cue (cued) from the combined population of SNC and VTA mDA neurons. mDA neurons were sorted into populations with significant excitation or inhibition around time of movement initiation (thin lines). Each set of PETHs is separated into self-initiated movements without licking (dashed light green, Lick−), self-initiated movements accompanied by licking (dark green, Lick+). In the same neurons, activity aligned to movement initiations in the trace period following auditory cue onset separated according to whether mDA neurons were excited (gray) or inhibited (black) around self-initiated movements. Corresponding PETHs of the basket movement signal are shown. **b**, Quantification of individual mDA neurons for populations represented in **a**. **c**, Mean modulation of mDA neuron activity ( $\Delta DA$ ) during movement initiation ( $\Delta DA$  at movement) is not correlated with training trials experienced (Pearson's correlation,  $n=96$ ). **d**,  $\Delta DA$  at movement is correlated with mean mDA response to the auditory cue ( $\Delta DA$  at cue, Pearson's correlation,  $n=96$ ). SNC (dark blue) and VTA (cyan) mDA neurons are shown. Dotted line represents best-fit trend.

for movement initiation. Therefore, we directly tested whether bursts of mDA activity of the duration and magnitude observed in our task might be responsible for initiating movement. We bilaterally injected the red-shifted  $\text{Ca}^{2+}$ -indicator jRCaMP1a directly under stimulation fiber sites in the SNC or VTA of separate cohorts of mice ( $n=4$  per group). We then performed fiberometry recordings in the primary targets of those regions in the ventral (vSTR) or dorsal (dSTR) striatum, respectively (Fig. 6a). Following 2–4 days of trace conditioning, we used our ability to monitor the activity of axonal populations to calibrate somatic optogenetic stimulation with observed reward signals at the same recording sites. Bursts of 10-ms pulses at 30 Hz were given in four conditions: 1 mW for 150 ms, 1 mW for 500 ms, 7 mW for 150 ms, and 7 mW for 500 ms, producing axonal  $\text{Ca}^{2+}$  responses that were, respectively,  $1.1 \pm 0.3$ ,  $2.8 \pm 0.4$ ,  $2.1 \pm 0.5$  and  $5.2 \pm 1$  times the magnitude of cued reward responses (see Methods; Fig. 6a–e). Neither SNC nor VTA stimulation matched to physiological signals was sufficient to elicit significant movement initiation (Fig. 6c,e). Indeed, only one condition, the strongest stimulation in the SNC, resulted in significant modulation of movement (VTA: one-way ANOVA  $P=0.8$ ,  $n=4$ ; SNC: one-way ANOVA  $P=0.01$ , baseline versus 150-ms, 1-mW stimulation,  $P=0.8$ ; baseline versus 150-ms, 7-mW stimulation,  $P=1$ ; baseline versus 500-ms, 1=-mW stimulation,  $P=0.6$ ; baseline versus 500-ms, 7-mW stimulation,  $P<0.0001$ ,  $n=4$ ; Fig. 6b–e and Supplementary Fig. 6). As a positive control, stimulation of mDA neurons matched to the magnitude of physiological reward signals (~1 $\times$ ; 150-ms, 30-Hz burst at 1 mW peak power) was able to induce conditioned place preference in a single 60-min session (Fig. 6f,g; VTA:  $50 \pm 8\%$  in the stimulated quadrant versus  $17 \pm 3\%$  in the other quadrants,  $n=4$ ,  $P=0.02$ ; SNC:  $42 \pm 6\%$  in the stimulated quadrant versus  $22 \pm 8\%$  in the other quadrants,  $n=4$ ,  $P=0.03$ ). Taken together, these data indicate that in the context of our task and within the physiological range of activity that we observe in both electrophysiological and calcium imaging, phasic mDA neuron activity at the time of movement does not (exclusively) reflect activity that is causal for movement initiation<sup>39</sup>.

**Emergence of RPE correlates in mDA activity during novel learning.** How do these sensory cue-related and movement initiation-related response components contribute to the quantitative representation of reward, and specifically RPE correlates, in mDA neurons? RPE correlates are well-known to be present in animals after extensive conditioning<sup>3</sup>; however, it is unclear when RPE correlates emerge during initial trace conditioning. To address this question, we began to omit either the auditory cue or water delivery on a subset of trials ( $\leq 10\%$  of trials in any session) once animals exhibited criterion learning (session 4; see Fig. 1c,d). This produced trials with ‘unpredicted (uncued) rewards’ and ‘omitted rewards’ and, therefore, trials with positive prediction errors and negative prediction errors, respectively<sup>3,19</sup>. mDA responses to water delivery remained unaffected by the presence or absence of a reward-predictive auditory cue in middle training (predicted versus unpredicted,  $4.3 \pm 0.6$  versus  $4.1 \pm 0.7$  Hz,  $P=0.65$ ,  $n=30$ ) despite clear evidence that mice had learnt that the auditory cue was predictive of reward (Figs. 1 and 4) and had a reduced latency to consume the reward when it was preceded by a cue (Fig. 7a). In late training, cued water delivery began to evoke smaller mDA responses than uncued water delivery ( $3.1 \pm 0.6$  versus  $4.5 \pm 1.0$  Hz,  $P=0.03$ ,  $n=24$ , Fig. 7b–d). Therefore, in late learning, we observed clear evidence of positive prediction error correlates in mDA neuron activity. We also examined the response of mDA neurons to omitted rewards to look for a negative prediction error correlate. Brief pauses in mDA neuron activity at the time of omitted rewards have provided the most direct evidence to date for negative prediction errors mediated by inhibition of mDA neurons at the moment of an expected (but omitted) delivery of reward<sup>40,41</sup>. As with positive prediction errors, we observed a significant pause in mDA activity to omitted rewards that emerged during late training, but was not present earlier (middle training,  $5.4 \pm 0.7$ – $5.7 \pm 0.9$ ,  $P=0.73$ ,  $n=16$ ; late training,  $5.3 \pm 0.6$ – $2.6 \pm 0.4$ ,  $P<0.0001$ ,  $n=17$ ; Fig. 7d and Supplementary Fig. 7). Therefore, the eventual emergence of signed RPE correlates (uncued > cued > 0 > omitted) and the specific shape of the mDA neuron response to water delivery that we observed is remarkably



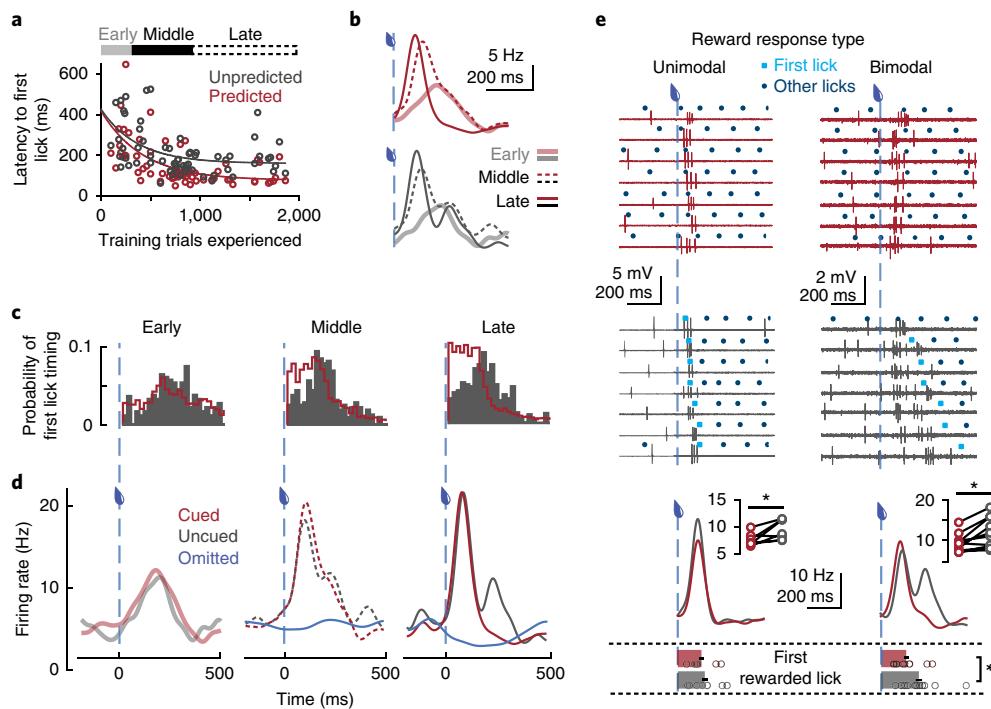
**Fig. 6 | Physiological mDA stimulation supports learning but is insufficient to provoke movement initiation.** **a**, Illustration of strategy used to record projection-specific mDA activity while optogenetically stimulating mDA somata (see Methods for details). In brief, adeno-associated viruses (AAVs) carrying a Cre-dependent transgene for a red genetically encoded calcium indicator (jRCamp1a) were injected into either the VTA or SNC. Fibers were implanted in either the SNC and dSTR (top) or VTA and vSTR (bottom) to allow simultaneous optogenetic stimulation and fiberometry of mDA axons. A/P, anteroposterior. **b**, Comparison of reward responses recorded in mDA projections from SNC to dSTR to responses to optogenetic burst stimulations with a duration of 150 and 500 ms at low (1mW peak, dark green) and high (7mW peak, light green) laser power. This yielded axonal  $\text{Ca}^{2+}$  responses  $\sim 1\times$ ,  $\sim 3\times$ ,  $\sim 2\times$ , and  $\sim 5\times$  the magnitude of endogenous reward responses. **c**, Left, for an example mouse (1 of 4), mean basket movement is plotted relative to time of stimulation onset. Stimulation regimes are indicated by color and correspond to data in **b**. Right, mean modulation of movement for all mice ( $n=4$ ) and stimulation conditions ( $n=4$ ). One-way ANOVA with post hoc Bonferroni's test,  $P<0.0001$ . **d**, **e**, Same as **b**, **c**, but for mDA projections from VTA to vSTR. One-way ANOVA (**e**) did not show significant differences between groups,  $F=0.3$ ,  $P=0.8$ . **f**, Example conditioned place preference session during which the mouse received 150-ms, low-power burst (1mW peak) stimulation in the blue-shaded area, showing raw movement data (left) and cumulative time spent in each quadrant as the session progressed (right). Quadrants are color-coded corresponding to enclosure edges. ISI, interspike interval. **g**, Mean time spent in stimulated versus nonstimulated quadrants for the final 15 min of the session.  $n=4$ , \* $P=0.03$  for mDA projections from the SNC to dSTR,  $n=4$ , \* $P=0.02$  for mDA projections from the VTA to vSTR.

consistent with previous observations in mice with  $\sim 1,000$ – $2,000$  trials of training<sup>21,38</sup>.

The unpredicted mDA response had two clear peaks (bimodal distribution) consistent with many previous reports at later stages of training in a variety of species<sup>3</sup>, including mice<sup>25,36</sup>. What has not been appreciated to date is that these two peaks occur at very similar latencies to canonical sensory cue-related responses ( $\sim 50$ – $150$  ms) and the latency to initiate consumptive licking ( $>150$  ms) after uncued water delivery, respectively (Fig. 7a–d). Additionally, we found that only a subset of individual mDA neurons displayed bimodal response time courses (Fig. 7e). When analyzed separately, both unimodal and bimodal populations encoded positive prediction errors in the form of larger responses to uncued rewards (unimodal,  $7.6\pm 0.5$  versus  $9.4\pm 0.7$ ,  $P=0.03$ ,  $n=7$ ; bimodal,  $9.2\pm 0.7$  versus  $11.6\pm 1.1$ ,  $P=0.006$ ,  $n=11$ ), indicating that unimodal responses did not reflect the absence of a component of the mDA response; rather, bimodal and unimodal response time courses could reflect differences in temporal integration. One key difference in action timing could dictate these changes in integration: whereas during bimodal responses, the latency from reward

delivery to the first rewarded lick event was significantly longer for cued versus uncued rewards ( $155\pm 9$  versus  $213\pm 27$  ms,  $n=11$ ,  $P=0.002$ ), this difference was smaller and nonsignificant during the recording of unimodal responses ( $142\pm 14$  versus  $162\pm 1$  ms,  $n=7$ ,  $P=0.06$ ). The timing of action therefore correlates with the time course of the mDA RPE correlate, indicating that temporal integration of sensory cue-related and action initiation-related components determines mDA reward-expectation signals.

**A model sufficient to account for the emergence of RPE correlates.** To further assess whether temporal integration was quantitatively consistent with our data, we developed a model with the minimal features that we propose explains mDA reward responses. mDA responses integrate: (i) excitation associated with reward-predictive sensory cues that strengthens with learning and (ii) an excitation–inhibition sequence around initiation of appetitive actions (for example, lick bouts) that remains constant throughout training. Predicted mDA neuron reward responses were generated by convolving these impulse response functions with observed behavior distributions (Fig. 7c; the distribution of latencies to



**Fig. 7 | Time course of RPE correlates in mDA neurons is determined by the timing of action initiation.** **a**, Mean latency from water delivery to first lick during each behavioral session in which mDA neurons were recorded as a function of number of training trials experienced at session start. Latencies are separately plotted for uncued water delivery (unpredicted, black) or water delivery preceded by auditory cue (predicted, red). Trend lines are single exponential fits. **b**, Mean mDA neuron PETHs from early (sessions 1–3;  $n=12$ ), middle (sessions 4–8,  $n=29$ ) and late (sessions 9+,  $n=24$ ) training overlaid for cued (top, red) and uncued (bottom, black) water delivery. **c**, Histograms of the time of first contact of the tongue to the lick port following water delivery in early, middle and late training. Uncued water delivery trials shown in gray and cued trials in red. **d**, PETHs of mDA activity aligned to cued (red), uncued (dark gray), or omitted (cue followed by no reward, blue) water delivery (dashed blue line) for early ( $n=12$ ), middle ( $n=30$ ) and late ( $n=24$ ) training sessions. Note, omitted trials were only present in a subset of recordings in middle ( $n=16$ ) and late ( $n=17$ ) training. **e**, Raw traces from two example cell-attached mDA neuron recordings in late training. Top, trials with cued water delivery aligned to the time of water delivery (dashed blue line). Bottom, trials with uncued water delivery aligned to the time of water delivery and sorted by the latency to the first lick of a consummative bout (cyan squares), distinguished from within-bout licks (dark blue circles). Left, example neuron with a unimodal PETH. Right, example neuron with a bimodal PETH. **e**, Bottom, PETHs from mDA neurons in late sessions divided into unimodal and bimodal response types for uncued (dark gray) and cued (red) water delivery. Latency to the first rewarded lick for each population is shown. Data are mean  $\pm$  s.e.m. Two-tailed Student's t-test \* $P=0.02$ . Insets show mean reward responses, with significant effects of prediction for both response types (two-sided Student's t-tests, unimodal:  $7.6 \pm 0.5$  versus  $9.4 \pm 0.7$ ,  $*P=0.03$ ,  $n=7$ ; bimodal:  $9.2 \pm 0.7$  versus  $11.6 \pm 1.1$ ,  $*P=0.006$ ,  $n=11$ ).

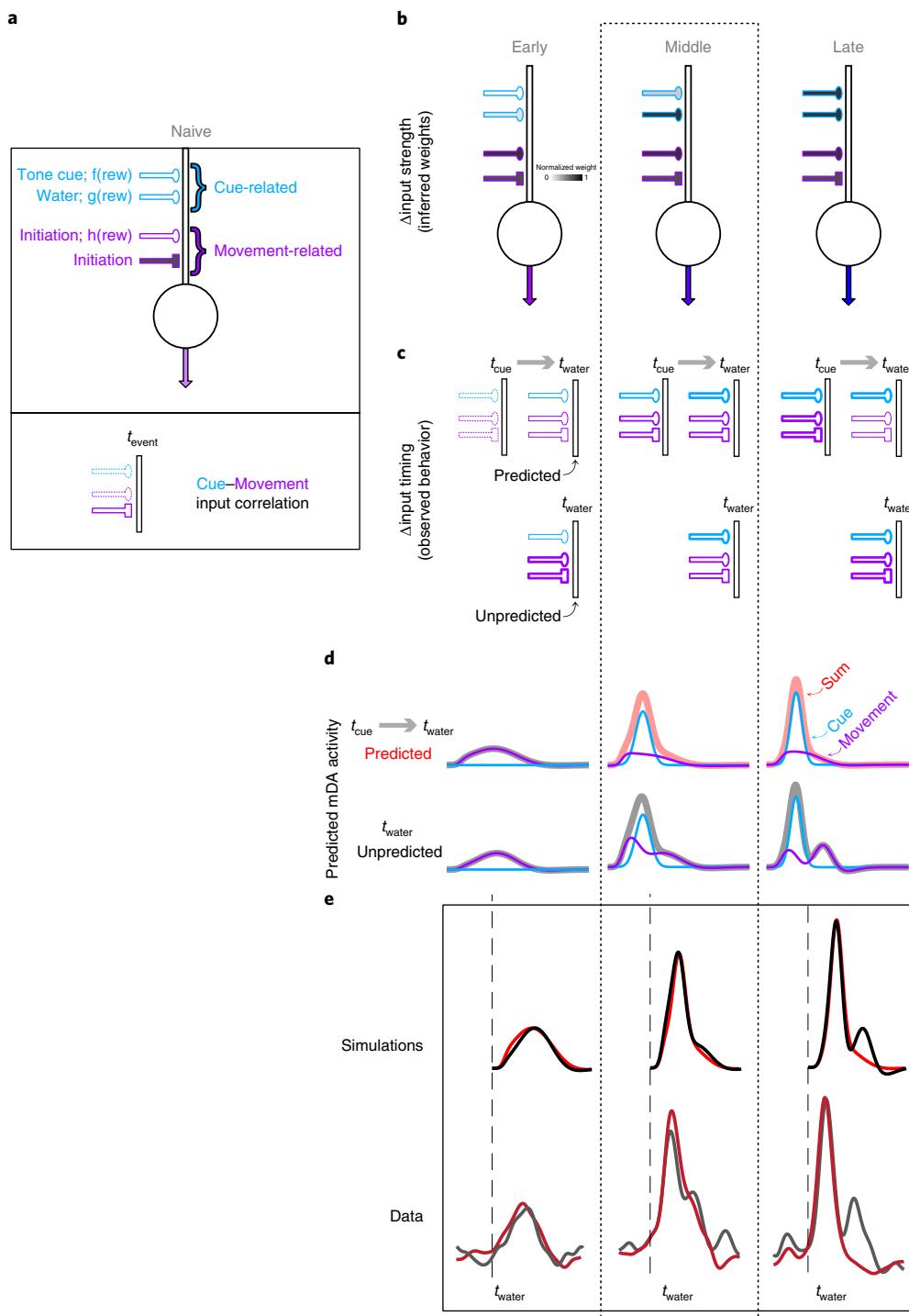
initiate consummatory lick bouts) across training phases (Fig. 8 and Supplementary Methods). This reduced model was sufficient to account for the detailed time course of mDA neuron activity throughout novel learning (Fig. 8).

Canonical implementation models of RPE predict that the magnitude of inhibition to an omitted reward is necessarily correlated with the difference between predicted and unpredicted rewards (for example, equations 3 and 4 in a previous study<sup>15</sup>). As noted above, our data exhibit a positive, rather than negative, correlation inconsistent with predictions from such models (Fig. 4c–e). In the model that we propose here (Fig. 8), negative trial-by-trial correlations are not predicted nor is inhibition required to account for the difference between predicted and unpredicted reward responses. Consistent with our implementation model, suppression of activity on omitted trials was uncorrelated with either the time course or integrated difference between predicted and unpredicted reward responses (Supplementary Fig. 7a–c). By contrast, our model predicts that the magnitude of inhibition that is observed during omitted rewards reflects inhibition following self-initiated movement that lacks the expectation of reward (that is, similar to ‘Lick−’ in Fig. 5). The magnitude of suppression of mDA neurons on omitted trials was indeed correlated with the magnitude of inhibition during self-initiated movements (Supplementary Fig. 7d,e).

## Discussion

Here we show that the positive, phasic responses of mDA neurons reflect the summation of an excitation–inhibition sequence that is related to the initiation of appetitive actions and excitation associated with sensory cues that predict future reward. In both components, excitation (but not inhibition) is modulated by expectation of reward; however, only sensory cue-related excitation depends on the extent of training. These two sources of input appear to summate. We therefore propose that these components of the mDA neuron response may reflect separate afferent pathways consistent with the diverse excitatory inputs to mDA neurons<sup>42</sup>. All inputs, albeit in different ratios<sup>43</sup>, appear to be present in the majority of VTA and SNC mDA neurons, consistent with a number of projections from frontal cortex and midbrain that target both mDA neuron populations<sup>42</sup>. Specific inactivation of colliculonigral projections has been observed to block some, but not all, components of mDA neuron responses in trained monkeys<sup>44</sup>, consistent with our observations that multiple, dissociable components underlie mDA neuron responses during conditioning.

**Action determines the timing of mDA signals.** Phasic mDA activity associated with the initiation of action has been observed many times in the literature; however, the sign, prevalence, timing and



**Fig. 8 | mDA neuron responses are consistent with temporal summation of sensory cue and action initiation components.** Overview of simulation data from conceptual model (top) through simulation (bottom). **a**, Top, simulations assumed two classes of excitation onto mDA neurons: sensory cue-related (cyan) and movement initiation-related (purple). All excitatory inputs are independent functions of reward expectation ( $\bullet(\text{rew})$ ) with a shared history of reward experience. In addition, movement-related input contained delayed inhibition. Weight of input is schematized by fill shading in schematic. Consistent with experimental data (Fig. 2) in the naive condition, a predominant inhibition is observed around movement initiation and reflected in the increased weight of movement-related inhibition. Bottom, afferent activation of pathways at time  $t$  are indicated by the saturation of pathway outline. **b**, Changes in the strength of movement-related and cue-related inputs are schematized for the three main learning stages. Darker shading represents a greater input strength. **c**, Changes in the relative timing of cue-related and movement-related activation for tone cue to water delivery (predicted) trials and unpredicted water delivery. The thickness of the line indicates weight of input. **d**, Impulse response functions were derived from data using a single excitatory Gaussian function for cue-related activity or a sequence of excitatory and delayed inhibitory Gaussian functions for movement-related activity (Figs. 3, 5, 7). Impulse response functions scaled by input strength for cue-related activity or scaled and convolved with a probability distribution matching observed behavior distributions for movement-related activity (distributions were fitted to Fig. 7). This yielded two components of mDA activity (cue-related in cyan and movement-related in magenta). **e**, Simulated mDA responses are the sum of these components for predicted (red) and unpredicted (black) water delivery. Simulated mDA responses and mDA neuron recording data are shown for comparison. See Methods for details to recreate figure.

learning-related properties of these correlates are variable across studies<sup>23–25,29,31,45</sup>. Here we provide a set of observations obtained in a large set of identified mDA neurons from two major midbrain nuclei during associative learning that can help to reconcile these observations. We find that a majority of mDA neurons receive a sequence of excitation followed by inhibition around movement initiation, even in naive animals. Movement-related inhibition appears more constant than excitation, as the introduction of rewards to the context significantly increased excitation at least in the SNC subpopulation. However, even in a rewarded context, self-initiated movement did not result in large excitation across the entire population, but only within a minority of individual neurons in both the VTA and SNC.

Arguing against a movement-specific subpopulation in our data, of 21 individual mDA neurons lacking excitation to reward cues, only 1 exhibited significant excitation around self-initiated movement (Supplementary Fig. 1d), whereas across the population, movement initiation-related and sensory cue-related activity was positively correlated (Fig. 5d,  $n=96$ ,  $r=0.35$ ,  $P<0.001$ ). We note that our recordings were localized to relatively lateral VTA and medial SNC. These subregions contain a substantial proportion of mDA neurons, and are less diverse in their connectivity, response patterns and biophysical properties than the medial VTA and lateral SNC, which might exhibit more specialization<sup>43,46</sup>. Although we did not observe a clear dissociation between mDA neurons that respond to action initiation and sensory cues, we note that the characteristic response to movement initiation was correlated with a difference in baseline firing rate (Supplementary Fig. 1c), consistent with the mDA heterogeneity reflected in the biophysics of individual neurons<sup>46</sup>.

Optogenetic manipulation can drive a substantially larger and more sustained activation of mDA populations than observed during behavior. Thus, we tested the sufficiency of mDA excitation for the control of movement using stimulation patterns calibrated (within the same animal) to physiological reward signals (largest transients observed during behavior). Stimulation calibrated to reward responses was sufficient to produce conditioned place preference, yet effects on movement were not apparent—arguing for a primary role of phasic dopamine signals in learning, rather than real-time behavioral control. A net effect on movement required stimulation at least 5× greater in magnitude and 2× greater in duration than that observed in response to reward and many-fold greater than the observed correlate of movement initiation. These data highlight the value of using calibrated exogenous stimulation and suggest that the typical phasic bursts of mDA activity associated with action initiation are not causal for movement. Rather, our data suggest that these mDA bursts may be a corollary signal that is associated with certain instances of action initiation under the expectation of a reward. The tight temporal synchronization of certain actions with phasic mDA signals (for example, Fig. 3b) suggests that the circuit mechanisms for generating phasic movement-related bursts in mDA neurons may fundamentally rely on the transition to action<sup>25,27,29</sup>, not only on motivationally relevant sensory stimuli.

**RPE correlates emerge following behavioral adaptation.** Our study is consistent with the fundamental insight that mDA neuron activity reflects aspects of subjective reward expectation<sup>3,31</sup>. We generated a parsimonious model that quantitatively accounts for the precise time course of the mDA response and the slow emergence of RPE correlates during initial learning. Our model contrasts with qualitative proposals that mDA responses in naive mice are primarily salience or novelty responses and/or reflect the direct sensation of a water reward<sup>3,15</sup>. Our data and model are consistent with causal evidence that exogenous induction of RPE correlates in mDA neuron activity is sufficient for learning<sup>47,48</sup> and that even prolonged suppression of mDA neuron activity within trials does not impair motor components of task performance<sup>39,49</sup>.

Detailed analysis of the emergence of mDA activity correlates revealed that both sensory cue-related and action initiation-related mDA neuron activity (at least in part) reflect expectation of reward. Learning drives both changes in behavior (action timing) and enhanced sensory cue-related components that lead the integrated mDA response to correlate with RPE after learning. However, our data and our proposed model call for a revision of the proposed circuit implementation of RPE signaling during initial learning<sup>2,15,36</sup>. The additive combination of sensory-related (input) and movement-related (output) signals in mDA neurons suggests that dopamine release could mediate a Hebbian-like teaching signal that reflects moments when sensory input coincides with action output. Hebbian rules are often sufficient for learning and can be equivalent to error-based rules<sup>50</sup>. Thus, an initial correlation-based teaching signal in the activity of mDA neurons may allow a novice animal to rapidly learn from its successes and later produce RPE correlates that could be useful for reducing errors or adapting to changes in contingency.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41593-018-0245-7>.

Received: 26 February 2018; Accepted: 20 August 2018;

Published online: 15 October 2018

### References

- Hebb, D. O. *The Organization of Behavior: a Neuropsychological Theory* (Wiley, New York, USA, 1949).
- Sutton, R. S. & Barto, A. G. *Reinforcement Learning: an Introduction* (MIT Press, Boston, MA, USA, 1998).
- Schultz, W. Neuronal reward and decision signals: from theories to data. *Physiol. Rev.* **95**, 853–951 (2015).
- Hollerman, J. R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* **1**, 304–309 (1998).
- Lak, A., Stauffer, W. R. & Schultz, W. Dopamine neurons learn relative chosen value from probabilistic rewards. *eLife* **5**, e18044 (2016).
- Pan, W.-X., Schmidt, R., Wickens, J. R. & Hyland, B. I. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J. Neurosci.* **25**, 6235–6242 (2005).
- Pan, W. X., Brown, J. & Dudman, J. T. Neural signals of extinction in the inhibitory microcircuit of the ventral midbrain. *Nat. Neurosci.* **16**, 71–78 (2013).
- Schultz, W., Apicella, P. & Ljungberg, T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* **13**, 900–913 (1993).
- Roesch, M. R., Calu, D. J. & Schoenbaum, G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.* **10**, 1615–1624 (2007).
- Kawagoe, R., Takikawa, Y. & Hikosaka, O. Reward-predicting activity of dopamine and caudate neurons—a possible mechanism of motivational control of saccadic eye movement. *J. Neurophysiol.* **91**, 1013–1024 (2004).
- Menegas, W., Babayan, B. M., Uchida, N. & Watabe-Uchida, M. Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. *eLife* **6**, e21886 (2017).
- Watabe-Uchida, M., Eshel, N. & Uchida, N. Neural circuitry of reward prediction error. *Annu. Rev. Neurosci.* **40**, 373–394 (2017).
- Bornstein, A. M. & Daw, N. D. Multiplicity of control in the basal ganglia: computational roles of striatal subregions. *Curr. Opin. Neurobiol.* **21**, 374–380 (2011).
- Hart, G., Leung, B. K. & Balleine, B. W. Dorsal and ventral streams: the distinct role of striatal subregions in the acquisition and performance of goal-directed actions. *Neurobiol. Learn. Mem.* **108**, 104–118 (2014).
- Suri, R. E. & Schultz, W. A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* **91**, 871–890 (1999).
- Darvas, M., Wunsch, A. M., Gibbs, J. T. & Palmiter, R. D. Dopamine dependency for acquisition and performance of Pavlovian conditioned response. *Proc. Natl Acad. Sci. USA* **111**, 2764–2769 (2014).

17. Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
18. Osborne, J. E. & Dudman, J. T. RIVETS: a mechanical system for *in vivo* and *in vitro* electrophysiology and imaging. *PLoS One* **9**, e89007 (2014).
19. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
20. Lima, S. Q., Hromadka, T., Znamenskiy, P. & Zador, A. M. PINP: a new method of tagging neuronal populations for identification during *in vivo* electrophysiological recording. *PLoS One* **4**, e6099 (2009).
21. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
22. Pan, W. X. & Dudman, J. T. A specific component of the evoked potential mirrors phasic dopamine neuron activity during conditioning. *J. Neurosci.* **35**, 10451–10459 (2015).
23. Dodson, P. D. et al. Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *Proc. Natl Acad. Sci. USA* **113**, E2180–E2188 (2016).
24. Howe, M. W. & Dombeck, D. A. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* **535**, 505–510 (2016).
25. da Silva, J. A., Tecuapetla, F., Paixao, V. & Costa, R. M. Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* **554**, 244–248 (2018).
26. Hamid, A. A. et al. Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016).
27. Syed, E. C. et al. Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat. Neurosci.* **19**, 34–36 (2016).
28. Barter, J. W. et al. Beyond reward prediction errors: the role of dopamine in movement kinematics. *Front. Integr. Neurosci.* **9**, 39 (2015).
29. Jin, X. & Costa, R. M. Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* **466**, 457–462 (2010).
30. Collins, A. L. et al. Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. *Sci. Rep.* **6**, 20231 (2016).
31. Romo, R. & Schultz, W. Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.* **63**, 592–606 (1990).
32. Fiorillo, C. D., Yun, S. R. & Song, M. R. Diversity and homogeneity in responses of midbrain dopamine neurons. *J. Neurosci.* **33**, 4693–4709 (2013).
33. Betley, J. N. et al. Neurons for hunger and thirst transmit a negative-valence teaching signal. *Nature* **521**, 180–185 (2015).
34. Zimmerman, C. A. et al. Thirst neurons anticipate the homeostatic consequences of eating and drinking. *Nature* **537**, 680–684 (2016).
35. Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* **10**, 1020–1028 (2007).
36. Pan, W.-X., Schmidt, R., Wickens, J. R. & Hyland, B. I. Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. *J. Neurosci.* **28**, 9619–9631 (2008).
37. Stuber, G. D. et al. Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science* **321**, 1690–1692 (2008).
38. Eshel, N., Tian, J., Bukwicz, M. & Uchida, N. Dopamine neurons share common response function for reward prediction error. *Nat. Neurosci.* **19**, 479–486 (2016).
39. Soares, S., Atallah, B. V. & Paton, J. J. Midbrain dopamine neurons control judgment of time. *Science* **354**, 1273–1277 (2016).
40. Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412**, 43–48 (2001).
41. Eshel, N. et al. Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* **525**, 243–246 (2015).
42. Watabe-Uchida, M., Zhu, L., Ogawa, S. K., Vamanrao, A. & Uchida, N. Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* **74**, 858–873 (2012).
43. Lammel, S., Lim, B. K. & Malenka, R. C. Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology* **76**, 351–359 (2014).
44. Takakuwa, N., Kato, R., Redgrave, P. & Isa, T. Emergence of visually-evoked reward expectation signals in dopamine neurons via the superior colliculus in V1 lesioned monkeys. *eLife* **6**, e24459 (2017).
45. Wood, J., Simon, N. W., Koerner, F. S., Kass, R. E. & Moghaddam, B. Networks of VTA neurons encode real-time information about uncertain numbers of actions executed to earn a reward. *Front. Behav. Neurosci.* **11**, 140 (2017).
46. Lammel, S. et al. Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. *Neuron* **57**, 760–773 (2008).
47. Steinberg, E. E. et al. A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* **16**, 966–973 (2013).
48. Chang, C. Y. et al. Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nat. Neurosci.* **19**, 111–116 (2016).
49. Fischbach-Weiss, S., Reese, R. M. & Janak, P. H. Inhibiting mesolimbic dopamine neurons reduces the initiation and maintenance of instrumental responding. *Neuroscience* **372**, 306–315 (2018).
50. Xie, X. & Seung, H. S. Equivalence of backpropagation and contrastive Hebbian learning in a layered network. *Neural Comput.* **15**, 441–454 (2003).

## Acknowledgements

We thank members of the J.T.D. laboratory, K. Bittner, C. Grienberger, D. Hunt, J. Macklin, J. Cohen, and R. Egnor for technical guidance; members of the J.T.D. laboratory and members of the V. Jayaraman laboratory, B. Mensh, A. Lee, G. Rubin, and J. Day for project feedback; R. Rogers, J. Arnold, and C. Loper for assistance with behavioral rig design and implementation; and S. Lindo for assistance with surgeries. This work was supported by the Howard Hughes Medical Institute. J.T.D. is supported by Janelia.

## Author contributions

Data collection and analysis were performed by L.T.C. with input from J.T.D. Simulations were implemented by J.T.D. with input from L.T.C. All other aspects of the work were the product of both authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41593-018-0245-7>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to L.T.C. or J.T.D.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018

## Methods

**Mice.** All procedures and animal handling were performed in strict accordance with protocols (11–39) that were approved by the Institutional Animal Care and Use Committee and consistent with the standards set forth by the Association for Assessment and Accreditation of Laboratory Animal Care. For behavior and cell-attached recordings, we used 11 adult male *Dat<sup>cre/+</sup>:Ai32* (*Dat* is also known as *Slc6a3*) mice (10–24 weeks old) resulting from the cross of *Dat<sup>REScre</sup>* (The Jackson Laboratory, 006660) and *Ai32* (The Jackson Laboratory, 012569) lines of mice, such that a Chr2-enhanced yellow-fluorescent protein (eYFP) fusion protein was expressed under the control of the endogenous dopamine-transporter *Slc6a3* locus to specifically label dopaminergic neurons. For combined fiber photometry and Chr2-stimulation experiments, we used 8 *Dat<sup>cre/+</sup>:Ai32* male mice (14–24 weeks old). The specificity of labelling in the *Dat<sup>cre</sup>* mice has been previously characterized with *Rosa26*-reporter mice<sup>31</sup> and the *Dat<sup>cre/+</sup>:Ai32* double transgenic mouse line has been used previously to obtain specific activation of mDA neurons<sup>32</sup>. All animals were handled in accordance with guidelines approved by the Institutional Animal Care and Use Committee of Janelia Research Campus. Animals were housed on a 12-hour dark:light cycle (08:00–20:00) and recording sessions were all done between 09:00–13:00. Following at least four days recovery from headcap implantation surgery, the water consumption of the mice was restricted to 1 ml per day for at least three days before training. Mice underwent daily health checks, and water restriction was eased if mice fell below 75% of their original body weight. No randomization procedure was used to determine inclusion of mice in the study—mice for each experiment type were chosen arbitrarily from among available litters.

**Behavioral training.** Mice were habituated to head fixation in a separate area from the recording rig in multiple sessions of increasing length over ≥3 days, including manual water administration through a syringe. Mice were then habituated to head fixation while resting in a spring-suspended basket in the recording rig for at least two sessions of 30 min or more before recordings were attempted or training commenced. No liquid rewards were administered during the acclimatization to the recording rig. Mice head-fixed while resting in a spring-suspended basket were then trained to learn a classical (Pavlovian) auditory trace-conditioning task. The reward consisted of 3 µl of water sweetened with the non-caloric sweetener acesulfame potassium delivered through a lick port under control of a solenoid. In the first training session, a 0.5-s, 10-kHz tone preceded reward delivery by 1.5 s on 100% of trials. In subsequent training sessions, unpredicted reward responding was assessed by randomly omitting the predictive tone on 30% of trials during select blocks of trials, with the result that the proportion of unpredicted reward trials never exceeded 10% of any given training session. Omitted reward responding was assessed by randomly omitting the reward following a predictive tone on 30% of trials during select blocks of trials, such that the proportion of omitted reward trials never exceeded 10% of any given training session. Intertrial intervals were chosen from randomly permuted exponential distributions (means of ~10, 25 or 50 s) every 20 trials in order to fully disrupt reliable estimation of intertrial interval while keeping the mean interval tractable for recording. Ambient room noise was 50–55 dB, while an audible click of ~53 dB was associated with solenoid opening upon water delivery and the predictive tone was ~65 dB loud. It should be noted that these stimuli are all quieter than the 72–90 dB stimuli that were reported to activate mDA neurons through their salience in primates<sup>32</sup>. Indeed, in naïve animals no significant modulation was apparent to the tone (Fig. 2a, left), and the time course of modulation by reward delivery early in training (Fig. 5a,b) was not consistent with proposed salience signaling in mDA neurons. However, to control for such signaling as an alternative explanation for the predictive nature of the mDA reward response, we recorded from additional mice trained with an inaudible solenoid (The Lee Company, LHQA1231220H).

**Behavioral and electrophysiological measurements.** Individual licks were timestamped according to deflections of a piezo strip supporting the lick port. Piezo signals were high-pass filtered at 0.1 Hz, then rectified and smoothed by convolving with a square wave in order to facilitate identification of individual tongue strikes during high-frequency lick bouts.

Basket movements were recorded by a triple-axis accelerometer (Adafruit, ADXL335) attached to the underside of a custom-designed 3D-printed basket suspended from springs (Century Spring Corp, ZZ3-36) well-suited to allow robust movements while fully supporting the ~20–25-g body weight of adult mice. Basket height relative to the point of head fixation was set so that the back of a mouse was at a ~30° angle to the plane of its headcap. This positioning was comfortable for the animals and minimized the translation of body movement to movement of the brain with respect to the skull, allowing for more stable recordings. Raw accelerometer signals summed across all axes were used to identify transitions from rest to movement (movement initiations). Relative basket position was tracked by low-pass filtering accelerometer data at 2.5 Hz to enrich for the signal corresponding to forces due to the angle of the accelerometer with respect to the ground. Both lick and movement events were detected by threshold crossing but were timestamped according to the their earliest deviation from the previous baseline signal.

Electrophysiological recordings were made with a Multiclamp 700B amplifier (Molecular Devices) interfaced to a computer by an analog-to-digital converter

(National Instruments, PCI 6259) controlled by Axograph X recording software ([www.axograph.com](http://www.axograph.com)). Spikes were recorded in current-clamp mode, AC-coupled at 1 Hz, then high-pass filtered at 300 Hz to facilitate simple threshold crossing analysis to generate spike timestamps. Data were smoothed by convolving with a Gaussian function with a 20-ms decay.

The above signals as well as the command signals that spanned the predictive tone length, the reward solenoid open time and the activation time of the laser used for optotagging were synchronously recorded and digitized (at 1 kHz for behavioral data, 30 kHz for electrophysiology data) with a Cerebus Signal Processor (Blackrock Microsystems). Stimulations and cue deliveries were coordinated with custom-written software using Arduino Mega hardware ([www.arduino.cc](http://www.arduino.cc)). Data was analyzed using MATLAB software (Mathworks). Data analysis was done in batches following days or weeks of recording sessions. During initial analysis, including timestamping of action potentials and alignment to task and behavioral variables, the experimenter was blinded to the recording location (VTA versus SNC) of the individual neuron as well as the training progression of the animal at the time of recording. Blinding to the above variables during data acquisition was not attempted.

**Cell-attached recording.** A small craniotomy (<200-µm diameter) was made over the recording site (from bregma: −3.2 posterior, 0.5 lateral for VTA; −3.0 posterior, 1.5 lateral for SNC) at least 4 h before recording. Exposed brain tissue was kept moist with phosphate-buffered saline (PBS) at all times, and craniotomy sites were covered with Kwik-Sil elastomer (WPI) outside of the recording session. Borosilicate glass pipettes (Sutter, BF165-120-10) were pulled to a long taper (~10 mm taper) with a ~1–2-µm tip (resistance 4–14 MΩ) with a P-97 micropipette puller (Sutter). Pipettes were filled with 0.5 M NaCl solution and mounted in a holder with a side port (Warner, PE30W-T17P) to allow insertion of a fiber (105-µm core, 0.22 numerical aperture (NA), Thorlabs) that was coupled to a 473-nm laser (50 mW, OEM Laser Systems) to carry light to the pipette tip. An AgCl pellet reference electrode was placed in the well of saline that covered the craniotomy site. Experimental power ranges out of the assembly were estimated by painting the taper of a pipette with nail polish in order to simulate brain tissue attenuation of light escaping the pipette, and were found to be 0.5–1.5 mW across the laser powers used during experiments.

Pipettes were lowered through the brain with a micromanipulator (Luigs and Neumann) while a small cycling current injection enabled monitoring of resistance changes across the pipette tip. In addition to stereotactic coordinates, dopaminergic regions could be further targeted empirically by monitoring extracellular potentials in response to brief (<5 ms) flashes of blue light through the pipette tip (Fig. 1e). The amplitude of population responses closely corresponded to patterns of dopaminergic cell body enrichments predicted from standard coordinates<sup>32</sup>. Deviations from that correspondence represented errors introduced through individual anatomical variation or the plane of headcap attachment to the skull, and referencing of the population response enabled correction on subsequent passes, greatly increasing the yield of identified mDA recordings across the recording sessions for each animal. Once within a dopaminergic region, the pipette tip was advanced by 1–2-µm steps until a steep increase in resistance was detected. The pipette was then advanced 5–10 µm until positive-going spikes were resolved well above noise (>0.5 mV). Chr2 expression was assayed with 0.5-s 473-nm light stimulations, either in bursts of 1-ms flashes at 10 Hz or continuous pulses. Cells were not re-identified for the remainder of the recording session so as not to interfere with physiological responses. Cells were held for a median of 11 trials (interquartile range, 7–19) and a median of 27 self-initiated movements (interquartile range, 14–47). During pilot experiments, 0 of 20 neurons in the first 2-mm of tissue overlying the targeted areas were significantly modulated by stimulation with maximum laser power (~26 mW out of the fiber tip), indicating that activation due to heating or other nonspecific processes was not a concern.

This technique is commonly referred to as ‘juxtacellular’ recording, however that term can suggest that transmembrane labelling was always performed. In order to increase yield from each animal, we did not label and recover a majority of cells, instead relying on the depth soundings reported by the extracellular responses to light for a reliable estimation of recording position. However, in some pilot experiments, identified mDA cells were labeled in the juxtacellular configuration for future recovery by including neurobiotin (Life Technologies) in the pipette solution and entraining spiking with 2 Hz, 50% duty cycle current injections of the minimum amplitude required to entrain spiking (>1 nA).

We quantified tone and reward responses by averaging firing rates over a 500-ms window following the cue and subtracting the mean baseline firing rate in the 1,000 ms preceding tone delivery<sup>31</sup>. We quantified movement responses by averaging firing rates over a 300-ms window beginning 50 ms before movement initiation and subtracting baseline firing rates—this window was chosen in order to capture the bulk of both excitatory and inhibitory responses across different neurons. Slightly different windows were used for one-tailed tests of significant modulations of activity by movement initiation in single cells: for excitation, a 200-ms window beginning 100 ms before initiation was used; for inhibition, a 300-ms window beginning at the moment of movement initiation was used. These parameters better described the activity during the optimal windows for each signal.

The numbers of cells recorded in each trial type for each animal is included in Supplementary Table 1.

**Combined fiber photometry and optogenetic stimulation.** During the course of a single surgery session, *Dat<sup>Cre/Cre</sup>;Ai32* mice received:

1. Bilateral injections of AAV2/1-CAG-FLEX-jRCaMP1a<sup>53</sup> in the VTA (50–100 nl at the coordinates –3.1 mm anteroposterior (A/P), 1.3 mm mediolateral (M/L) from bregma, at depths of 4.6 and 4.3 mm) or in the SNC (100 nl at the coordinates –3.2 mm A/P, 0.5 mm M/L, depth of 4.3 mm).
2. Custom 0.39 NA, 200-μm fiber cannulas implanted bilateral directly above the injection sites (VTA, –3.2 mm A/P, 0.5 mm M/L, depth of –4.1 mm; SNC: –3.2 mm A/P, 1.4 mm M/L, depth of 4 mm).
3. A fiber cannula implanted unilaterally in either dSTR (0.9 mm A/P, 1.5 mm M/L, depth of 2.5 mm) or vSTR (1.5 mm A/P, 0.85 mm M/L, depth of 4.1 mm). Hemisphere choice was counterbalanced across individual mice.

Imaging began 20 days post-injections using a fiber photometry system custom-built around a five-port filter cube (FMC5, Doric Lenses). The system was designed with two parallel excitation–emission channels to allow for simultaneous measurement of RCaMP1a and eYFP fluorescence, the latter channel having the purpose of controlling for the presence of movement artifacts. The 470-nm and 565-nm fiber-coupled LEDs (M470F3, M565F3, Thorlabs) were connected to excitation ports with acceptance bandwidths of 465–490 nm and 555–570 nm, respectively, with 200-μm, 0.22-NA fibers (Doric Lenses). Light was conveyed between the sample port of the cube and the animal by a 200-μm core, 0.39-NA fiber (Doric Lenses) that terminated in a ceramic ferrule that was connected to the implanted fiber cannula by a ceramic mating sleeve (ADAL1, Thorlabs) using index matching gel to improve coupling efficiency (G608N3, Thorlabs). Light collected from the sample fiber was measured at separate output ports (emission bandwidths 500–540 nm and 600–680 nm) by 600-μm core, 0.48-NA fibers (Doric Lenses) connected to silicon photoreceivers (2151, Newport). Signals from the receivers were streamed into the Blackrock Cerebus Signal Processor described above at 30 kHz. In order to avoid influencing physiological RCaMP signals, blue excitation for the measurement of the static eYFP signal was only used intermittently outside of recording periods to verify the absence of significant movement artifacts aligned to task variables or somatic mDA stimulation. LEDs were controlled at a frequency of 50 Hz (3 ms on, 17 ms off) by TTL signals, which were recorded so that excitation-specific signals could be resolved in analysis using custom-written MATLAB code.

Somatic Chr2 excitation was performed with a 473-nm laser (50 mW, OEM Laser Systems) coupled by a branching fiber patch cord (200 μm, Doric Lenses) to the SNC or VTA-implanted fiber cannulas using a ceramic mating sleeve. All 30-Hz burst activations (10 ms on, 23 ms off) were delivered with durations of either 150 ms or 500 ms. Two levels of laser power were used, resulting in 1 mW or 7 mW peak power with constant illumination, measured at the tip of the ferrule of the patch cable. The lower level of 1 mW was chosen, because when given in a 150-ms-duration burst it resulted in axonal Ca<sup>2+</sup> signals equivalent to those measured in response to reward delivery after 2–4 training sessions. For movement-initiation sufficiency tests (Fig. 7b–d), burst stimulations were delivered with interstimulus intervals randomly chosen from an exponential distribution with a mean of 32 s. One of the four burst options (low or high power, short or long burst) were chosen at random for each trial.

**Conditioned place preference.** Mice were placed in an open arena with a clear floor through which video data were monitored, and their VTA- or SNC-targeted fibers were mated to the laser patch cable tethered to a commutator suspended above the arena. One quadrant was chosen at random in which the mouse would receive bilateral laser burst stimulation (calibrated to a reward-equivalent level: 150 ms long, 30 Hz burst with 1 mW peak power), and stimulation was triggered by entry into an area within one-half of the length of the quadrant from the

corner of the quadrant (see Fig. 7f). Stimulation was delivered with a minimum interstimulus interval of 4 s while the mouse was in the designated stimulation area. This intermittent schedule was chosen rather than the more common constant burst illuminations used in order to be comparable to the isolated bursts used in the tests for movement initiation sufficiency. Sessions lasted 60 min and time spent in each quadrant was summed over the entirety of the session.

**Statistical analysis.** No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous publications<sup>7,21,23</sup>. Data distribution was assumed to be normal but was not formally tested for normality. Paired comparisons were made using two-tailed Student's *t*-tests, except where significance was tested for individual neuron activity aligned to movement or reward, in which case one-tailed Wilcoxon signed-rank tests were performed. Multiple comparisons were made using ANOVAs with Tukey's post hoc multiple-comparisons test (Graphpad Prism). Contingency testing was done with the χ<sup>2</sup> test. Errors are reported as standard errors of the mean (s.e.m.). All sample sizes refer to the number of distinct neurons summarized in the data, or for behavioral quantification (for example, Fig. 2b), it refers analogously to the number of distinct neuron recording sessions during which behavior was quantified. For the receiver-operator characteristic (ROC) analysis of differences between baseline and delay period (anticipatory) licking, number of licks were counted per 100-ms bins in the baseline period 1,000 ms before cue delivery, and compared to lick counts in 100-ms bins during the 1,000-ms delay period following cessation of the auditory cue.

**Histology.** Mice were killed by anesthetic overdose (isoflurane, >3%) and perfused with ice-cold PBS, followed by paraformaldehyde (4% wt/vol in PBS). Brains were post-fixed for 2 h at 4 °C and then rinsed in saline. Whole brains were then sectioned (100 μm thickness) using a vibrating microtome (VT-1200, Leica Microsystems). For recovery of juxta-labelled cells from pilot experiments exemplified in Fig. 1e–g, slices were incubated with Alexafluor594-conjugated streptavidin to visualize neurobiotin-labeled neurons against the backdrop of eYFP-expressing mDA neurons.

**Model simulations.** Simulations were implemented in MATLAB as described in Fig. 8 and the main text. Example simulation code related to Fig. 8 is available at <http://www.dudmanlab.org/html/learnda.html>.

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

**Code availability.** The custom codes used to generate the results that support the findings of this study are available from the corresponding authors upon reasonable request. The code for the model simulation in relation to Fig. 8 is available at the following website: <http://www.dudmanlab.org/html/learnda.html>.

## Data availability

The data used to generate the results that support the findings of this study are available from the corresponding authors upon reasonable request.

## References

51. Zhuang, X., Masson, J., Gingrich, J. A., Rayport, S. & Hen, R. Targeted gene expression in dopamine and serotonin neurons of the mouse brain. *J. Neurosci. Methods* **143**, 27–32 (2005).
52. Tritsch, N. X., Oh, W. J., Gu, C. & Sabatini, B. L. Midbrain dopamine neurons sustain inhibitory transmission using plasma membrane uptake of GABA, not synthesis. *eLife* **3**, e01936 (2014).
53. Hod, D. et al. Sensitive red protein calcium indicators for imaging neural activity. *eLife* **5**, e12727 (2016).

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

## Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated
- Clearly defined error bars  
*State explicitly what error bars represent (e.g. SD, SE, CI)*

*Our web collection on [statistics for biologists](#) may be useful.*

## Software and code

Policy information about [availability of computer code](#)

Data collection

Axograph X was used to monitor neural recordings online and to control laser stimulation, while data was recorded through a Blackrock Cerebus system under control of Cerebus Central Suite (7.0.3.0) software. Task variables were controlled by custom built Arduino-based hardware programmed with Arduino IDE (1.8.3) software.

Data analysis

Matlab r2016a and Graphpad Prism 6 were used for data analysis and statistical tests.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data and custom code used to generate results supporting the findings of this study are available from the corresponding authors upon reasonable request.

# Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. Sample sizes compared favorably with existing studies of dopamine neuron activity in awake-behaving animals, exceeding all previously reported data sets in the number of unambiguously-identified dopamine neurons included.
Data exclusions	No data were excluded from the analysis.
Replication	We have reported all methods and analysis details required to replicate our experiment.
Randomization	This is not relevant to our study, as care was taken that all mice used experienced identical training conditions.
Blinding	Data analysis was done in batches following days or weeks of recording sessions. During initial analysis including timestamping of action potentials and alignment to task and behavioral variables, the experimenter was blinded to the recording location (ventral tegmental area vs substantia nigra) of the individual neuron as well as the training progression of the animal at the time of recording. Blinding to the above variables during data acquisition was not attempted.

## Reporting for specific materials, systems and methods

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Animals and other organisms

Policy information about [studies involving animals; ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	We used male adult DAT-cre x ai32 mice (10-24 weeks old) resulting from the cross of DAT-IREScre (The Jackson Laboratory stock 006660) and Ai32 (The Jackson Laboratory stock 012569) lines of mice.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve field-collected samples.