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

Luke T. Coddington¹ and Joshua T. Dudman¹

¹Address: Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147

Abstract

Animals adapt behavior in response to informative sensory cues using multiple brain circuits. The activity of midbrain dopamine (mDA) neurons is thought to convey a critical teaching signal: reward prediction error (RPE). Although RPE signals are thought to be essential to learning, little is known about the dynamic changes in mDA neuron activity 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 naïve mice learned a novel cue-reward association. During learning mDA neuron activity results from summation of sensory cue-related and movement initiation-related response components. These components are both a function of reward expectation yet dissociable. Learning produces an increasingly precise coordination of action initiation following sensory cues that results in apparent RPE correlates. Our data thus provide new insights into circuit mechanisms underlying a critical computation in a highly conserved learning circuit.

Introduction

A hungry guest's stomach rumbles at the sound of a dinner bell; a child extends her hand for a piece of candy. Both sensory cues and the 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 learning¹. The second is a gradual reduction in errors

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence and requests for materials should be sent to: luketc82@gmail.com (L.T.C.) or dudmanj@janelia.hhmi.org (J.T.D.).

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.

Author Information:

The authors declare no competing interests.

Data availability statement:

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

Code availability statement:

The custom code used to generate results supporting the findings of this study are available from the corresponding authors upon reasonable request. The code for the model simulation in relation to Figure 8 are available at the following website:

www.dudmanlab.org/html/learnnda.html

of prediction through comparison of expected and observed outcomes - often a critical component of reinforcement learning². In a well-known example of such a representation, the activity of mammalian midbrain dopamine (mDA) neurons correlates with changes in reward expectation - or reward prediction errors (RPEs) - following associative learning³. This has provided direct evidence in support of the notion that reinforcement learning proceeds via 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 contingencies⁴⁻¹¹. Behavioral performance and mDA reward signals can adapt to these new contingencies within tens of trials in rodents¹¹ and monkeys^{4,5}. However, it has also been observed that mDA neuron correlates can lag adaptive changes in behavior^{5,7}, calling into question the causal role of RPE correlates in the learning of associations. Additionally, even in the well-studied overtrained condition the circuit mechanisms by which reward prediction errors are computed remains unclear¹². This could be due, at least in part, to the tight coordination of multiple learning systems characteristic of the trained state^{13,14}.

In sharp contrast to the rapid adaptation to changed contingencies in trained animals, for naive animals learning a novel association many hundreds of trials can be required to observe stable or asymptotic behavior - especially for trace conditioning paradigms. 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 reflected in mDA neuron activity during such novel learning^{3,15}. However, to date, empirical data to quantitatively compare to these predictions is lacking. Moreover, whereas novel learning requires intact mDA signaling¹⁶, 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 possible circuit mechanisms underlying RPE correlates.

Here we examine movement-related mDA activity as naïve (but habituated) mice are first trained in a Pavlovian trace-conditioning paradigm. 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 to responses to self-initiated movements, sweetened water rewards, and sensory cues in well trained mice we detail 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 that fact that RPE correlates are a consequence of temporal integration of independent reward expectation signals associated with reward-predictive sensory cues and initiation of appetitive actions.

Results

Adult transgenic mice expressing a light-activated opsin, channelrhodopsin-2¹⁷ (Chr2), in mDA neurons were exposed to a trace conditioning paradigm in which a 0.5 sec long auditory tone was presented 1.5 sec before delivery of a sweetened water reward (Fig. 1a-b; see Methods). We chose to use a head fixed paradigm¹⁸ as in canonical studies on representations in mDA neurons^{3,19} and for the desirable reduction in variability of behavior. While no action was necessary to receive water, anticipation of available water is often accompanied by body movements (readily observed in freely moving animals) not typically assessed in head-fixed subjects. To enable such measurement, we supported mice 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 presence of available water (Fig. 1c). The mean reaction time to collect the water was monotonically reduced from hundreds of milliseconds in naïve animals to a latency of ~60 ms (~1/2 licking cycle) after ~1000 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 steadily increased (Fig. 1c-d; latency to initiate licking following tone, $r = -0.50$, $p < 0.0001$; licks in the CR 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 paradigm rapidly re-acquire and extinguish responses to the tone⁷, naïve learning is characterized by a rapid emergence of learned behavior (~3 sessions; ~100 trials/session) and a gradual asymptotic stabilization of multiple aspects of learned behavior (decreased latency to collect water, increased anticipatory licking and stereotyped body movements).

Given this examination of behavior during the first ~1000 trials of training we next explored 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'^{20,21}) mDA neurons in the VTA ($n = 47$) and SNc ($n = 88$) of naïve mice as they learned an auditory trace conditioning paradigm. 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 neurons²². 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 to movement initiation in naïve or untrained animals^{23–25}. In trained animals, DA release in ventral striatum can correlate with initiation of reward-related actions^{26,27}. Thus, we first examined activity in naïve animals acclimated to the head-fixed context, but prior to receiving 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 SNc and 9 of 17 in VTA upon initiation of body movement. Only one mDA neuron in VTA and none in SNc exhibited significant excitation (Wilcoxon's signed rank test, modulation window vs baseline, see Methods; Fig. 2a). Reports of excitatory peri-movement responses in mDA neurons^{23,25} led us to more carefully examine activity around the time of movement initiation. We measured the interspike interval (ISI) just before movement-related pauses in activity (see Methods). This analysis revealed an apparent "covert" excitation in the form of a spike phase advance present in mDA neurons from both 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 more prominent positive modulation around purposive action initiation in rodents^{28,29,30} and primates³¹, we next examined whether weak excitation in naïve 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 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 ± 1.5 s) separated by periods of stillness (interbout interval: 6.4 ± 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 naïve context, in a rewarded context we observed more significant excitation of individual mDA neurons just prior to self-initiated movement onset (23 out of 96 during training vs 1 out of 29 pre-reward training, $p = 0.01$, SNc: 19 of 66 during training vs 0 of 12 pre-training, $p = 0.03$; VTA: 4 of 30 during training vs 1 of 17 pre-training, $p = 0.6$, Fig. 2b, c). Thus, SNc DA neurons were more likely to be excited at movement initiation than VTA DA neurons although significant excitation was observed from individual neurons in both populations. As in naïve 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 naïve animals, covert excitation was present in non-"excited" neurons in the form of a spike phase advance just prior to movement-related pauses (Supplementary Fig. 2c). Together these results suggest a majority of mDA neurons receive both excitation and inhibition upon movement initiation; however, excitation can be enhanced by the presence of rewards, while inhibition is not.

We next examined within-trial modulation of mDA neuron activity as naïve 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 associated with the initiation of consumptive licking (Fig. 3). This was particularly apparent in mDA neurons recorded in mice with variable latencies to initiate consumption (Fig. 3a). On average, 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 50ms onset, 85ms peak⁷. 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 robust 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 vs 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: 52dB; background noise: 55 dB; see Fig. 4) and none of our stimuli reached the intensities previously reported to produce salience responses (~90 dB^{3,32}).

If it is the case that the modulation of mDA activity after reward delivery does not have a significant 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. Thus, we recorded mDA neurons in a subset of animals 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 naïve animals, modulation of activity can be predictive of consumption rather than a reflection of physical, sensorial contact with water^{33,34}. Our data thus 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 over early learning. Consistent with a large body of prior work³, 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). Surprisingly, 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^{6,8,15,19,35–37}. By contrast, our data – the first to examine the time course of the emerging mDA response to a predictive cue in naïve animals – suggest that mDA responses to the predictive cue and the reward do not exhibit the predicted, negative trial-by-trial correlation^{15,36}. 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). We finally examined the relationship of cue and reward signals on a trial-by-trial basis selectively within neurons that had significant sensory cue-related responses. Only 6 of 75 such neurons exhibited significant correlations between auditory cue and reward responses, and they were *positive* (Fig. 4e). mDA responses to predictive sensory cues and water delivery thus appear to emerge independently while being scaled by a common cell-specific factor^{32,38} (Supplementary Fig. 4).

We have described two apparently independent components of the phasic responses of mDA neurons: a movement-related excitatory-inhibitory sequence (Fig. 2, 3) and a phasic excitation associated with sensory stimuli that emerges as mice begin to exhibit learned behavior (Fig. 3d, 4a-c). In the sensory cue-related component, excitation is specifically enhanced as a function of the extent of learning (Fig. 3d, 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 for 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 animal's reward expectation. 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 mvmt"; Fig 5a). We reasoned that this analysis should distinguish categories of action with relatively less (Lick-), more (Lick +), or most (Cued) degrees of reward expectation. A 2-way ANOVA (movement type x movement response type) confirmed a main effect of movement response type (excited vs inhibited at mvmt, $F(1,216) = 149.8$, $p < 0.0001$) as well as a main effect of movement type (Lick- vs Lick+ vs 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 indeed, in mDA neurons positively modulated by movement initiation, Lick+ movements were accompanied by greater excitation than Lick-movements (3.2 ± 0.6 vs 1.0 ± 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. In contrast, inhibition apparent in negatively modulated mDA neurons was independent of expectation (-2.3 ± 0.3 vs -1.8 ± 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 ± 0.6 Hz, $p < 0.0001$ vs Lick-; negatively modulated: -0.4 ± 0.5 Hz, $p = 0.05$ vs Lick-, $p = 0.002$ vs 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 ± 0.6 vs 4.8 ± 0.6 Hz, $p = 0.25$, Fig 5b). Importantly, 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 for mDA neurons significantly inhibited around movement (Fig 5a, right; $p < 0.0001$). Together, these results suggest that mDA neuron activity reflects the sum 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 a net inhibition during movements in naïve mice (Fig. 2). This indicates that in our paradigm mDA activity is not obligatory for movement. However, it has recently been proposed that excitation in select mDA^{SNC} neurons is specifically sufficient²⁴ or perhaps even necessary²⁵ for movement initiation. Thus, we sought to directly test whether bursts of mDA activity of the duration and magnitude observed in our task might be responsible for initiating movement. We injected the red-shifted Ca^{2+} -indicator jRCaMP1a bilaterally 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 made use of our ability to monitor axonal population activity to calibrate somatic optogenetic stimulation to observed reward signals at the same recording sites (4 total conditions: axonal Ca^{2+} responses that were, respectively, 1.1 ± 0.3 , 2.8 ± 0.4 , 2.1 ± 0.5 , and 5.2 ± 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 a significant modulation of movement (VTA: 1-way ANOVA $p = 0.8$, $n = 4$; SNC: 1-way ANOVA $p = 0.01$, baseline vs 150 ms, 1 mW stim: $p = 0.8$, baseline vs 150 ms, 7 mW stim: $p = 1$, baseline vs 500 ms, 1 mW stim: $p = 0.6$, baseline vs 500 ms, 7 mW stim: $p < 0.0001$, $n = 4$; Fig. 6b-e; Supplementary Fig. 6). As a positive control, stimulation of mDA neurons matched to the magnitude of physiological reward signals (150 ms, 30 Hz burst at 1 mW peak power) was able to induce conditioned place preference in a single 60 minute session (Fig. 6f-g; VTA: $50 \pm 8\%$ in stim quadrant vs $17 \pm 3\%$ in others, $n = 4$, $p = 0.02$; SNC: $42 \pm 6\%$ in stim quadrant vs $22 \pm 8\%$ in others, $n = 4$, $p = 0.03$). Thus, taken together these data argue that in the context of our task and within the physiological range of activity 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 initiation³⁹.

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 conditioning³; 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 thus trials with positive prediction errors and negative prediction errors, respectively^{3,19}. mDA responses to water delivery remained unaffected by the presence or absence of a reward-predictive auditory cue in middle training (pred. vs unpred.: 4.3 ± 0.6 vs 4.1 ± 0.7 Hz, $p = 0.65$, $n = 30$) despite clear evidence that mice had

learnt the auditory cue was predictive of reward (Fig. 1, 4) and had a reduced latency to consume 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 ± 0.6 vs 4.5 ± 1.0 Hz, $p = 0.03$, $n = 24$, Fig 7b-d). Thus, 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^{40,41}. 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 ± 0.7 to 5.7 ± 0.9 , $p = 0.73$, $n = 16$; late training: 5.3 ± 0.6 to 2.6 ± 0.4 , $p < 0.0001$, $n = 17$; Fig. 7d; Supplementary Fig. 7). Thus, the eventual emergence of signed RPE correlates (uncued > cued > 0 > omitted), and the specific shape of the mDA neuron response to water delivery we observed is remarkably consistent with previous observations in mice with ~1000–2000 trials of training^{21,38}.

The unpredicted mDA response had two clear peaks ('bimodal') consistent with many previous reports at later stages of training in a variety of species³ including mice^{25,36}. What has not been appreciated to date is that these two peaks occur at very similar latencies to canonical sensory cue-related responses (~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 timecourses (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 ± 0.5 vs 9.4 ± 0.7 , $p = 0.03$, $n = 7$; bimodal: 9.2 ± 0.7 vs 11.6 ± 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 timecourses could reflect differences in temporal integration. One key difference in action timing could dictate these changes in integration: while during bimodal responses the latency from reward delivery to the first rewarded lick event was significantly longer for cued vs uncued rewards (155 ± 9 vs 213 ± 27 ms, $n = 11$, $p = 0.002$), this difference was smaller and nonsignificant during the recording of unimodal responses (142 ± 14 vs 162 ± 1 ms, $n = 7$, $p = 0.06$). The timing of action thus correlates with the timecourse 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 we propose to explain mDA responses during early learning. mDA responses integrate: (1) an excitation associated with reward predictive sensory cues that strengthens with learning and (2) an excitation-inhibition sequence around initiation of appetitive actions (*e.g.* lick bouts) constant throughout learning. Predicted mDA neuron responses were generated by convolving these impulse response functions with observed behavior distributions (Fig. 7c; the distribution of latencies

to initiate consummatory lick bouts) across training phases (Fig. 8; Supplemental 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 (*e.g.* equations 3 and 4 in ¹⁵). As noted above, our data exhibit a positive, rather than negative, correlation inconsistent with predictions from such models (Fig. 4c-e). In the implementation 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, suppression of activity on omitted trials was uncorrelated with either the timecourse or integrated difference between predicted and unpredicted reward responses (Supplementary Fig. 7a-c). In contrast, our account predicts that the magnitude of inhibition observed during omitted rewards reflects inhibition following self-initiated movement lacking expectation of reward (*i.e.* 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 related to initiation of appetitive actions and an excitation associated with sensory cues that predict future reward. In both components, the excitation (but not inhibition) is modulated by expectation of reward; however, only the sensory cue-related excitation depends upon the extent of training. These two sources of input appear to summate and have distinct dynamics during learning. We thus propose that these components of the mDA neuron response may reflect separate afferent pathways consistent with the diverse excitatory inputs to mDA neurons ⁴². All inputs, albeit in different ratios ⁴³, 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 ⁴². Specific inactivation of colliculonigral projections has been observed to block some, but not all, components of mDA neuron responses in trained monkeys ⁴⁴ 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 learning-related properties of these correlates are variable across studies ^{23–25,29,31,45}. 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 naïve 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 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), while over 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 portion of midbrain dopamine 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^{43,46}. While 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 mDA heterogeneity reflected in the biophysics of individual neurons⁴⁶.

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 DA signals in learning, rather than online behavioral control. A net effect on movement required stimulation at least 5x greater in magnitude and 2x greater in duration than that observed in response to reward and many fold greater than the observed correlate of movement initiation. This data both highlights the value of using calibrated exogenous stimulation and suggests that the typical phasic bursts of mDA activity associated with action initiation are not causal for movement. Rather, our data suggests that these mDA bursts may be a corollary signal associated with certain instances of action initiation under the expectation of a reward. The tight temporal synchronization of certain actions with phasic mDA signals (*e.g.* Fig 3b) suggests that the circuit mechanisms for generating phasic movement-related bursts in mDA neurons may fundamentally rely on the transition to action^{25,27,29} not solely 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^{3,31}. We elaborate a parsimonious model that is the first model to quantitatively account for the precise timecourse of mDA response and the slow emergence of RPE correlates during initial learning. Our model contrasts with qualitative proposals that mDA responses in naïve mice are primarily salience or novelty responses and/or reflect direct sensation of a water reward^{3,15}. Our data and model are consistent with causal evidence that exogenous induction of RPE correlates in mDA neuron activity is sufficient for learning^{47,48} and that even prolonged suppression of mDA neuron activity within trials does not impair motor components of task performance^{39,49}.

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 calls for a revision of the proposed circuit implementation of RPE signaling during initial learning^{2,15,36}. 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 reflecting moments when sensory input coincides with action output. Hebbian rules are often sufficient for learning and can be equivalent to error-based rules⁵⁰. 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.

Methods

Animals.

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 (IACUC) and consistent with the standards set forth by the Association for Assessment and Accreditation of Laboratory Animal Care (AALAC). For behavior and cell-attached recordings we used 11 adult male DAT-Cre::ai32 mice (10–24 weeks old) resulting from the cross of DAT^{IRESc} (The Jackson Laboratory stock 006660) and Ai32 (The Jackson Laboratory stock 012569) lines of mice, such that a Chr2/EYFP fusion protein was expressed under 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-Cre::ai32 male mice (14–24 weeks old). The specificity of labelling in the DAT-cre mice has been previously characterized with Rosa26 reporter mice⁵¹ and the DAT-cre::Ai32 double transgenic has been used previously to obtain specific activation of mDA neurons⁵². 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 (8am–8pm) and recording sessions were all done between 9am–1pm. Following at least 4 days recovery from headcap implantation surgery, animals' water consumption was restricted to 1 mL per day for at least 3 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 employed 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 30+ minute sessions before recordings were attempted or training commenced. No liquid rewards were administered during this

recording rig acclimation. Mice head-fixed while resting in a spring-suspended basket were then trained to learn a classical (Pavlovian) auditory trace-conditioning paradigm. The reward consisted of 3 μ L 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% over 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% over 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 attended 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 via their salience in primates³². Indeed, in naïve animals no significant modulation was apparent to the tone (Fig. 2a, left), and the timecourse 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 but fully support the ~20–25 g body weight of adult mice. Basket height relative to the point of head fixation was set so that a mouse’s back was at a ~30° angle to the plane of the 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 earth. 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). 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 time stamps. 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 on 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). 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 (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.

Cell-attached recording.

A small craniotomy (<200 μ m 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 hours prior to recording. Exposed brain tissue was kept moist with phosphate-buffered saline 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 μ m tip (resistance 4–14 mOhm) 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 μ m core, 0.22 NA, Thorlabs) that was coupled to a 473 nm laser (50mW, 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 allowed 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 coordinates²². 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 allowed correction on subsequent passes, greatly

increasing the yield of identified mDA recordings across the life of 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 ($> \sim 0.5$ 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 (IQR 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 max 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 imply 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 average baseline firing rate in the 1000 ms preceding tone delivery²¹. 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 detailed in Supplementary Table 1.

Combined fiber photometry and optogenetic stimulation.

In the course of a single surgery session, DAT-Cre::ai32 mice received:

- 1) Bilateral injections of AAV2/1-CAG-FLEX-jRCaMP1a in the VTA (50–100 nL at the coordinates -3.1 mm A/P, 1.3 mm 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) Fiber cannula implanted unilaterally in either dorsal striatum (0.9 mm A/P, 1.5 mm M/L, depth of 2.5 mm) or ventral striatum (1.5 mm A/P, 0.85 mm M/L, depth of 4.1 mm). Hemisphere choice was counterbalanced across individuals.

Imaging began 20 days post-injections using a fiber photometry system custom-built around a 5-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. 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 μ m, 0.22 NA fibers (Doric Lenses). Light was conveyed between the sample port of the cube and the animal by a 200 μ m core, 0.39 NA fiber (Doric Lenses) terminating 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 μ m 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 (50mW, OEM Laser Systems) coupled by a branching fiber patch cord (200 μ m, Doric Lenses) to the SNc or VTA-implanted fiber cannulas using a ceramic mating sleeve. 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 Ca²⁺ 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 seconds. 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 was 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 quadrant's length from the corner of the quadrant (see Fig 7f). Stimulation was delivered with a minimum interstimulus interval

of 4 seconds 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 minutes and time spent in each quadrant was totaled 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^{7,21,23}. Data distribution was assumed to be normal but was not formally tested for normality. Paired comparisons were made using two-tailed Students' 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 chi-squared 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 (e.g. 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 1000 ms before cue delivery, and compared to lick counts in 100 ms bins during the 1000 ms delay period following cessation of the auditory cue.

Histology.

Mice were killed by anesthetic overdose (isoflurane, >3%) and perfused with ice-cold phosphate-buffered saline (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 Alexafluor 594-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 Figure 8 and the main text. Example simulation code related to Figure 8 is available at www.dudmanlab.org/html/learnnda.html.

Reporting Summary.

Further information on study design is available in the attached Life Science Reporting Summary.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

This work was supported by the Howard Hughes Medical Institute. J.T.D. is a Janelia Group Leader. Thanks to Dudman lab members, Kate Bittner, Christine Grienberger, David Hunt, John Macklin, Jeremy Cohen, and Roian Egnor for technical guidance. Thanks to the Dudman lab, Vivek Jayaraman lab, Brett Mensh, Albert Lee, Gerry Rubin, and Jeremy Day for project feedback. Thanks to Roger Rogers, Jon Arnold, and Cameron Loper for assistance with behavioral rig design and implementation. Thanks to Sarah Lindo for assistance with surgeries.

References

1. Hebb DO The organization of behavior; a neuropsychological theory. (Wiley, 1949).
2. Sutton RS & Barto AG Reinforcement learning : an introduction. (MIT Press, 1998).
3. Schultz W Neuronal Reward and Decision Signals: From Theories to Data. *Physiol Rev* 95, 853–951, doi:10.1152/physrev.00023.2014 (2015). [PubMed: 26109341]
4. Hollerman JR & Schultz W Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature neuroscience* 1, 304–309, doi:10.1038/1124 (1998). [PubMed: 10195164]
5. Lak A, Stauffer WR & Schultz W Dopamine neurons learn relative chosen value from probabilistic rewards. *Elife* 5, doi:10.7554/eLife.18044 (2016).
6. Pan W-X, Schmidt R, Wickens JR & Hyland BI Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25, 6235–6242 (2005). [PubMed: 15987953]
7. Pan WX, Brown J & Dudman JT Neural signals of extinction in the inhibitory microcircuit of the ventral midbrain. *Nature neuroscience* 16, 71–78, doi:10.1038/nn.3283 (2013). [PubMed: 23222913]
8. 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. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 13, 900–913 (1993). [PubMed: 8441015]
9. Roesch MR, Calu DJ & Schoenbaum G in *Nature neuroscience* Vol. 10 1615–1624 (2007). [PubMed: 18026098]
10. Kawagoe R, Takikawa Y & Hikosaka O Reward-predicting activity of dopamine and caudate neurons—a possible mechanism of motivational control of saccadic eye movement. *Journal of neurophysiology* 91, 1013–1024, doi:10.1152/jn.00721.2003 (2004). [PubMed: 14523067]
11. Menegas W, Babayan BM, Uchida N & Watabe-Uchida M Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. *Elife* 6, doi:10.7554/eLife.21886 (2017).
12. Watabe-Uchida M, Eshel N & Uchida N Neural Circuitry of Reward Prediction Error. *Annual review of neuroscience*, doi:10.1146/annurev-neuro-072116-031109 (2017).
13. Bornstein AM & Daw ND Multiplicity of control in the basal ganglia: computational roles of striatal subregions. *Curr Opin Neurobiol* 21, 374–380, doi:10.1016/j.conb.2011.02.009 (2011). [PubMed: 21429734]
14. Hart G, Leung BK & Balleine BW Dorsal and ventral streams: the distinct role of striatal subregions in the acquisition and performance of goal-directed actions. *Neurobiology of learning and memory* 108, 104–118, doi:10.1016/j.nlm.2013.11.003 (2014). [PubMed: 24231424]
15. Suri RE & Schultz W in *Neuroscience* Vol. 91 871–890 (1999). [PubMed: 10391468]
16. Darvas M, Wunsch AM, Gibbs JT & Palmiter RD Dopamine dependency for acquisition and performance of Pavlovian conditioned response. *Proceedings of the National Academy of Sciences of the United States of America* 111, 2764–2769, doi:10.1073/pnas.1400332111 (2014). [PubMed: 24550305]
17. Boyden ES, Zhang F, Bamberg E, Nagel G & Deisseroth K Millisecond-timescale, genetically targeted optical control of neural activity. *Nature neuroscience* 8, 1263–1268, doi:10.1038/nn1525 (2005). [PubMed: 16116447]

18. Osborne JE & Dudman JT RIVETS: a mechanical system for in vivo and in vitro electrophysiology and imaging. *PloS one* 9, e89007, doi:10.1371/journal.pone.0089007 (2014). [PubMed: 24551206]
19. Schultz W, Dayan P & Montague PR A neural substrate of prediction and reward. *Science* 275, 1593–1599 (1997). [PubMed: 9054347]
20. Lima SQ, Hromadka T, Znamenskiy P & Zador AM PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. *PloS one* 4, e6099, doi:10.1371/journal.pone.0006099 (2009). [PubMed: 19584920]
21. Cohen JY, Haesler S, Vong L, Lowell BB & Uchida N Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482, 85–88 (2012). [PubMed: 22258508]
22. Pan WX & Dudman JT A Specific Component of the Evoked Potential Mirrors Phasic Dopamine Neuron Activity during Conditioning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 35, 10451–10459, doi:10.1523/JNEUROSCI.4096-14.2015 (2015). [PubMed: 26203140]
23. Dodson PD et al. Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *Proceedings of the National Academy of Sciences of the United States of America* 113, E2180–2188, doi:10.1073/pnas.1515941113 (2016). [PubMed: 27001837]
24. Howe MW & Dombeck DA Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature*, doi:10.1038/nature18942 (2016).
25. da Silva JA, Tecuapetla F, Paixao V & Costa RM Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* 554, 244–248, doi:10.1038/nature25457 (2018). [PubMed: 29420469]
26. Hamid AA et al. Mesolimbic dopamine signals the value of work. *Nature neuroscience* 19, 117–126, doi:10.1038/nn.4173 (2016). [PubMed: 26595651]
27. Syed EC et al. Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nature neuroscience* 19, 34–36, doi:10.1038/nn.4187 (2016). [PubMed: 26642087]
28. Barter JW et al. Beyond reward prediction errors: the role of dopamine in movement kinematics. *Frontiers in integrative neuroscience* 9, 39, doi:10.3389/fnint.2015.00039 (2015). [PubMed: 26074791]
29. Jin X & Costa RM Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature* 466, 457–462 (2010). [PubMed: 20651684]
30. Collins AL et al. Dynamic mesolimbic dopamine signaling during action sequence learning and expectation violation. *Sci Rep* 6, 20231, doi:10.1038/srep20231 (2016). [PubMed: 26869075]
31. Romo R & Schultz W Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *Journal of neurophysiology* 63, 592–606 (1990). [PubMed: 2329363]
32. Fiorillo CD Two dimensions of value: dopamine neurons represent reward but not aversiveness. *Science* 341, 546–549, doi:10.1126/science.1238699 (2013). [PubMed: 23908236]
33. Betley JN et al. Neurons for hunger and thirst transmit a negative-valence teaching signal. *Nature* 521, 180–185, doi:10.1038/nature14416 (2015). [PubMed: 25915020]
34. Zimmerman CA et al. Thirst neurons anticipate the homeostatic consequences of eating and drinking. *Nature* 537, 680–684, doi:10.1038/nature18950 (2016). [PubMed: 27487211]
35. Day JJ, Roitman MF, Wightman RM & Carelli RM Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nature neuroscience* 10, 1020–1028, doi:10.1038/nn1923 (2007). [PubMed: 17603481]
36. Pan W-X, Schmidt R, Wickens JR & Hyland BI Tripartite mechanism of extinction suggested by dopamine neuron activity and temporal difference model. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28, 9619–9631 (2008). [PubMed: 18815248]
37. Stuber GD et al. in *Science* Vol. 321 1690–1692 (2008). [PubMed: 18802002]
38. Eshel N, Tian J, Bukwich M & Uchida N Dopamine neurons share common response function for reward prediction error. *Nature neuroscience* 19, 479–486, doi:10.1038/nn.4239 (2016). [PubMed: 26854803]

39. Soares S, Atallah BV & Paton JJ Midbrain dopamine neurons control judgment of time. *Science* 354, 1273–1277, doi:10.1126/science.aah5234 (2016). [PubMed: 27940870]
40. Waelti P, Dickinson A & Schultz W Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412, 43–48 (2001). [PubMed: 11452299]
41. Eshel N et al. Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* 525, 243–246, doi:10.1038/nature14855 (2015). [PubMed: 26322583]
42. Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A & Uchida N Whole-brain mapping of direct inputs to midbrain dopamine neurons. *Neuron* 74, 858–873, doi:10.1016/j.neuron.2012.03.017 (2012). [PubMed: 22681690]
43. Lammel S, Lim BK & Malenka RC Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology* 76 Pt B, 351–359, doi:10.1016/j.neuropharm.2013.03.019 (2014). [PubMed: 23578393]
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, doi:10.7554/eLife.24459 (2017).
45. Wood J, Simon NW, Koerner FS, Kass RE & Moghaddam B Networks of VTA Neurons Encode Real-Time Information about Uncertain Numbers of Actions Executed to Earn a Reward. *Frontiers in behavioral neuroscience* 11, 140, doi:10.3389/fnbeh.2017.00140 (2017). [PubMed: 28848408]
46. Lammel S et al. Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. *Neuron* 57, 760–773 (2008). [PubMed: 18341995]
47. Steinberg EE et al. A causal link between prediction errors, dopamine neurons and learning. *Nature neuroscience* 16, 966–973, doi:10.1038/nn.3413 (2013). [PubMed: 23708143]
48. Chang CY et al. Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nature neuroscience* 19, 111–116, doi:10.1038/nn.4191 (2016). [PubMed: 26642092]
49. Fischbach-Weiss S, Reese RM & Janak PH Inhibiting Mesolimbic Dopamine Neurons Reduces the Initiation and Maintenance of Instrumental Responding. *Neuroscience* 372, 306–315, doi:10.1016/j.neuroscience.2017.12.003 (2018). [PubMed: 29248527]
50. Xie X & Seung HS Equivalence of backpropagation and contrastive Hebbian learning in a layered network. *Neural Comput* 15, 441–454, doi:10.1162/089976603762552988 (2003). [PubMed: 12590814]
51. Zhuang X, Masson J, Gingrich JA, Rayport S & Hen R Targeted gene expression in dopamine and serotonin neurons of the mouse brain. *Journal of neuroscience methods* 143, 27–32 (2005). [PubMed: 15763133]
52. Tritsch NX, Oh WJ, Gu C & Sabatini BL Midbrain dopamine neurons sustain inhibitory transmission using plasma membrane uptake of GABA, not synthesis. *Elife* 3, e01936 (2014). [PubMed: 24843012]

Fig. 1. Juxtacellular recording from identified midbrain dopamine neurons in awake, behaving mice.

a) Schematic of the head-fixed behavioral apparatus. Mice were head-fixed in front of water port tube positioned within tongue-length and supported by a spring-suspended basket monitored by an accelerometer. Small droplets of sweetened water (~3 μ L) were delivered by opening a valve on a gravity fed line; a suspended drop of liquid stayed on port until collected by licking.

b) Auditory stimuli, openings of a solenoid water valve (reward), vibration of the water port due to licking (licking), and juxtacellular recordings from identified midbrain dopamine (mDA) neurons could be recorded simultaneously with bodily movements that displaced basket position relative to its resting position (relative basket position).

c) Licking (upper) and basket displacement (lower, 'mvmt') in response to auditory cue (shaded gray bar) and sweetened water delivery (dashed line) averaged for the same mouse at training sessions 1, 3, and 8, $n = 50$ trials for each average trace.

d) (upper) Area under the receiver operating characteristic (AU-ROC) curve to assess whether number of licks differed significantly from baseline (first session to reach significance at $p < 0.05$ for each animal indicated with asterisk) during the 1 second delay between cue and reward. Colored lines represent individual mice, mean data shown in bolded circles with SEM. (lower) % of animals with $p < 0.05$ for AU-ROC analysis as training progresses.

e) mDA neurons in DAT-cre::ai32 mice selectively express a channelrhodopsin-YFP (ChR2-YFP) fusion protein. Representative histological section shows labeling of mDA soma in the midbrain and terminals in the striatum (also 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. White scale bar = 1 mm.

f) (left) Individual and merged images from pilot experiment (1 of 5 similar results) of neurobiotin-labeled neuron (red) and YFP-expressing mDA neurons (green). (right) Pipette trajectory with locations of light stimulations numbered. White 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-YFP expression (green) and evoked field potential amplitude. Lower 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 (LEFP) was observed (right), referenced to the dorsal and ventral limits at which LEFP was observed (left, interquartile box, center line at mean, whiskers indicating min and max values).

Fig. 2. Peri-movement excitation of mDA neurons, but not inhibition, depends on reward context. **a** (upper) Raster plot and perievent time histogram (PETH) for example SNc (left, dark blue) and VTA (right, cyan) DA neurons, (middle) Heat maps of z-scored PETHs for the two populations, (lower) and average PETHs recorded after acclimation to the recording rig but before animal ever received sweetened water rewards. PETHs are aligned to the beginning of self-initiated movements. Shaded areas indicate SEM.

- b)** During intertrial intervals after training had begun, single example mDA neurons from SNc with clear negative (left) and positive (right) modulation of firing around self-initiated movement onset. Shaded areas indicate SEM.
- c)** Heat maps and PETH for population data as in (a), but for all neurons recorded over the course of associative trace conditioning. Shaded areas indicate SEM.

Fig. 3. Initial mDA reward responses encode reward-related movements

a) Raw cell-attached recording from a mDA^{SNC} neuron aligned to water delivery during the first 30 training trials experienced by the mouse. Trials are sorted according to the latency for 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 sec surrounding reward delivery ($n = 22$ neurons recorded early in learning, mean $\pm$ SEM). Points of max correlation above (circles, with overlaid mean and SEM bars). Lag was -96 ± 44 ms. Significantly different from zero; $p = 0.04$ two-sided t-test.

c) Mean firing rate (top) and lick (bottom) rates aligned to the first contact of the tongue to the lick port after water delivery for 10 (out of 22) mDA neurons with significant responses to water delivery (as determined by one-tailed sign rank test $p < 0.05$) early in training (sessions 1–3). Shaded areas indicate SEM.

d) (left) Average 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) vs trials 301–900 (black points) in the time windows indicated by pink (1–150 ms post reward, two-sided t-test $p = 0.0002$) and gray (151–400 ms post reward, two-sided t-test $p = 0.6$) bars drawn below traces at left. Black bars indicate mean $\pm$ SEM.

e) Data from 2 additional mice trained with an inaudible solenoid (see Supplementary Fig. 3) controlling water delivery. Mean firing rate for 5 (out of 10) mDA neurons with significant responses to water delivery (as determined by one-tailed sign rank test $p < 0.05$) recorded early in training (session 1–3). Shaded areas indicate SEM.

Fig. 4. mDA neuron responses to predictive cue and reward stimuli evolve independently during acquisition learning

a) Behavioral (upper row) and mean mDA firing rate aligned to the opening of the water valve and divided in to early (sessions 1–3, left column) middle (sessions 4–8, middle) and late (sessions >8, right) training sessions. Conditioned responding was assessed by measuring relative basket position (basket mvmt) and lick rate. Smoothed peri-event time histograms (PETHs) for identified mDA neuron populations recorded in SNc (dark blue, middle row) and VTA (cyan, lower row). Shaded area indicates the standard error of the mean (s.e.m.); number of cells composing the mean 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 in CR, 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 correlation 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 running average overlaid (blue line), $y=0$ highlighted as thin dashed lines. Colors reflect SNc (dark blue) and VTA (cyan) populations. Results of Pearson's correlation 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 darker.

Fig. 5. Peri-movement activity reflects reward expectation and sums with cue responses

a (left) PETHs aligned to initiation time of movements from the intertrial intervals (self-initiated) or following reward-repdctive cue (cued) from the combined population of SNc and VTA mDA neurons. mDA neurons were sorted into populations with significant excitation (left) or inhibition (right) around time of movement initiation (thin vertical lines). Each set of PETHs is separated into self-initiated movements absent licking (dashed light green, Lick-), self-initiated movements accompanied by licking (dark green, Lick+). (right) 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 (mvmt) are shown in lower rows.

b Quantification of individual mDA neurons for populations represented in (a).

c Mean modulation of mDA neuron activity during movement initiation (Δ DA at movement) is not correlated to training trials experienced (Pearson's, $n = 96$).

d Δ DA at movement is correlated to mean mDA response to the auditory cue (Δ DA at cue, Pearson's, $n = 96$). SNc (dark blue) and VTA (cyan) mDA neurons identified by color. Dotted line represents best-fit trend.

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). Briefly, AAVs carrying a cre-dependent transgene for a red genetically encoded calcium indicator (jRCaMp1a) were injected into either VTA or SNc. Fibers were implanted in either SNc and dorsal striatum (dSTR; upper) or VTA and ventral striatum (vSTR; lower) to allow simultaneous optogenetic stimulation and fiberometry of mDA axons.

b) Comparison of reward responses recorded in mDA^{SNc>dSTR} projections to responses to 150 and 500 ms-duration optogenetic burst stimulations at low (1 mW peak, dark green) and high (7 mW peak, light green) laser power.

c) For an example mouse (1 of 4) average basket movement is plotted relative to time of stimulation onset. Stimulus regimes are indicated by color and correspond to data in **b**. Right panels show average modulation of movement for all mice (N=4) and stimulus conditions (N=4). Asterisk indicates a significant difference, post-hoc Bonferroni's test $p < 0.0001$ following 1-way ANOVA.

d-e) Same as **b)** and **c)** but for mDA^{VTA>vSTR}. 1-way ANOVA for (e) failed to find significance, $F = 0.3$, $p = 0.8$.

f) Example conditioned place preference session where mouse received 150 ms, low power burst (1 mW 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 color coded corresponding to enclosure edges.

g) Mean time spent in stimulated vs non-stimulated quadrants for the final 15 minutes of the session ($n = 4$, $p = 0.03$ mDA^{SNc>dSTR}, $n = 4$, $p = 0.02$ mDA^{MTA>vSTR})

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) Average mDA neuron PETHs from early (n=12), middle (n=29), and late (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 grey), or 'omitted' (cue followed by no reward, blue) water delivery (dashed blue line) for early (left, n=12), middle (center, n=30), and late (right, 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. Upper panels are trials with cued water delivery aligned to the time of water delivery (dashed blue line). Lower panels are trials with uncued water delivery aligned to the time of water delivery and sorted by the latency to the first lick of a consumptive bout (cyan squares), distinguished from within-bout licks (dark blue circles). (Left) an 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 grey) and cued (red) water delivery. Lower row shows latency to the first rewarded lick for each population (bars represent mean $\pm$ SEM), asterisk indicates two-tailed t-test $p = 0.02$. Insets show mean reward responses, with significant effects of prediction for both response types (two sided t-tests, unimodal: 7.6 ± 0.5 vs 9.4 ± 0.7 , $p=0.03$, $n= 7$; bimodal: 9.2 ± 0.7 vs 11.6 ± 1.1 , $p = 0.006$, $n = 11$).

Fig. 8. mDA neuron responses are consistent with temporal summation of sensory cue and action initiation components

Overview of simulation data from conceptual model (upper panels), through simulation (lower panels). (Left box) 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 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 “naïve” condition a predominant inhibition is observed around movement initiation and reflected in the increased weight of movement-related inhibition.

(Lower box) Afferent activation of pathways at time t are indicated by the saturation of pathway outline.

input strength panels:

Changes in the strength of movement-related and cue-related inputs are schematized in top row for 3 main learning stages described in text. Darker shading represents a greater input strength.

input timing panels:

Changes in the relative timing of cue-related and movement-related activation for tone cue - > water delivery (predicted) trials (upper) and for unpredicted water delivery (lower).

Thickness of line indicates weight of input.

predicted mDA activity panels:

Impulse response functions were derived from data using either a single excitatory Gaussian function for cue-related activity or a sequence of excitatory and delayed inhibitory Gaussian functions for movement-related activity (Fig. 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 fit to Fig. 7). This yielded two components of mDA activity (cue-related in cyan and movement-related in magenta). Simulated mDA responses are the sum of these components for predicted (red) and unpredicted (black) water delivery. Lower box compares simulated mDA responses and to mDA neuron recording data for comparison. See Matlab code in Supplemental Methods to recreate figure.