



Published in final edited form as:

Nature. 2019 June ; 570(7759): 65–70. doi:10.1038/s41586-019-1235-y.

## Dissociable dopamine dynamics for learning and motivation.

Ali Mohebi<sup>1,\*</sup>, Jeffrey R. Pettibone<sup>1,\*</sup>, Arif A. Hamid<sup>2</sup>, Jenny-Marie Wong<sup>3</sup>, Leah T. Vinson<sup>4</sup>, Tommaso Patriarchi<sup>5</sup>, Lin Tian<sup>5</sup>, Robert T. Kennedy<sup>3</sup>, and Joshua D. Berke<sup>1,4,6,#</sup>

<sup>1</sup>Department of Neurology, University of California, San Francisco

<sup>2</sup>Department of Neuroscience, Brown University

<sup>3</sup>Department of Chemistry, University of Michigan, Ann Arbor.

<sup>4</sup>Neuroscience Graduate Program, University of California, San Francisco

<sup>5</sup>Department of Biochemistry and Molecular Medicine, School of Medicine, University of California, Davis.

<sup>6</sup>Weill Institute for Neurosciences and Kavli Institute for Fundamental Neuroscience, University of California, San Francisco.

### Abstract

The dopamine projection from ventral tegmental area (VTA) to nucleus accumbens (NAc) is critical for motivation to work for rewards, and reward-driven learning. How dopamine supports both functions is unclear. Dopamine spiking can encode prediction errors, vital learning signals in computational theories of adaptive behavior. By contrast, dopamine release ramps up as animals approach rewards, mirroring reward expectation. This mismatch might reflect differences in behavioral tasks, slower changes in dopamine cell spiking, or spike-independent modulation of dopamine release. Here we compare spiking of identified VTA dopamine cells with NAc dopamine release in the same decision-making task. Cues indicating upcoming reward increased both spiking and release. Yet NAc core dopamine release also covaried with dynamically-evolving reward expectations, without corresponding changes in VTA dopamine cell spiking. Our results suggest a fundamental difference in how dopamine release is regulated to achieve distinct functions: broadcast burst signals promote learning, while local control drives motivation.

---

Dopamine is famously related to “reward” – but how exactly? One function involves learning from unexpected rewards. Brief increases in dopamine cell firing encode reward prediction errors (RPEs<sup>1–3</sup>) - learning signals for optimizing future motivated behavior. Dopamine manipulations can affect learning as if altering RPEs<sup>4–6</sup>. But they also affect

---

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](http://www.nature.com/authors/editorial_policies/license.html#terms)

# Joshua.Berke@ucsf.edu.

**Contributions.** A.M. performed and analyzed the electrophysiology and photometry, and applied the computational model. J.P. performed and analyzed the microdialysis, with assistance from J.W. and supervision by R.K. A.H. developed the behavioral task and initial photometry setup, and performed the voltammetry. L.V. performed retrograde tracing and analysis. T.P. and L.T. developed the dLight sensor and shared expertise. J.D.B. designed and supervised the study, and wrote the manuscript.

\*equal contributions.

**Competing Interests.** The Authors declare no competing interests.

motivated behaviors immediately, as if dopamine signals reward expectation (value)<sup>5</sup>. Furthermore, NAc dopamine escalates during motivated approach, consistent with encoding value<sup>7-11</sup>.

With few exceptions<sup>2,13,14</sup>, midbrain dopamine firing has been examined during classical conditioning in head-fixed animals<sup>3,12</sup>, unlike forebrain dopamine release. We therefore compared firing with release under the same conditions. We identified VTA dopamine neurons using optogenetic tagging<sup>3,14</sup>. To measure NAc dopamine release we used three independent methods – microdialysis, voltammetry, and the optical sensor dLight<sup>15</sup> – with convergent results. Our primary conclusion is that although RPE-scaled VTA dopamine spike bursts provide abrupt changes in dopamine release appropriate for learning, separate NAc dopamine fluctuations associated with motivation arise independently from VTA dopamine cell firing.

## Dopamine tracks motivation in key loci

We trained rats in an operant “bandit” task<sup>5</sup> (Fig. 1a,b). On each trial illumination of a nose poke port (*Light-On*) prompted approach and entry (*Center-In*). After a variable hold period (0.5–1.5s), white noise (*Go Cue*) led the rat to withdraw (*Center-Out*) and poke an adjacent port (*Side-In*). On rewarded trials this *Side-In* event was accompanied by a food hopper click, prompting the rat to approach a food port (*Food-Port-In*) to collect a sugar pellet. Leftward, rightward choices were each rewarded with independent probabilities, which occasionally changed without warning. When rats were more likely to receive rewards, they were more motivated to perform the task. This was apparent in their “latency” – the time between *Light-On* and *Center-In* - which was sensitive to the outcome of the preceding few trials (Extended Data Fig. 1) and thereby scaled inversely with reward rate (Fig. 1b).

We previously reported<sup>5</sup> a correlation between NAc dopamine release and reward rate, consistent with the motivational role of mesolimbic dopamine<sup>16</sup>. Here, we first wished to know whether this is observed throughout forebrain targets, consistent with globally “broadcast” dopamine signaling<sup>17</sup> or is restricted to specific subregions. We further hypothesized that these dopamine dynamics would differ between striatum and cortex, since these structures have distinct dopamine uptake/degradation kinetics<sup>18</sup> and may use dopamine for distinct functions<sup>19,20</sup>.

Using microdialysis with liquid chromatography–mass spectrometry we surveyed medial frontal cortex and striatum (Fig. 1c; Extended Data Fig. 1). We simultaneously assayed 21 neurotransmitters and metabolites with 1 min time resolution, and used regression to compare chemical time series with behavioral variables (Extended Data Fig. 2).

We replicated the correlation between reward rate and NAc dopamine – unlike other neurotransmitters (Fig. 1c,d). However, this relationship was localized to NAc core, not NAc shell or dorsal-medial striatum. Contrary to our hypothesis, we observed a similar spatial pattern in frontal cortex: dopamine release correlated with reward rate in ventral prelimbic cortex, but not more dorsal or ventral subregions (Fig. 1c,e). Though unexpected, these twin “hotspots” of value-related dopamine release have an intriguing parallel in human

neuroimaging: BOLD signal correlates with subjective value specifically in NAc and ventral-medial prefrontal cortex<sup>21</sup>.

## VTA firing is unrelated to motivation

We next addressed whether this motivation-related forebrain dopamine arises from variable firing of midbrain dopamine cells. NAc core receives dopamine input from lateral portions of VTA (VTA-l;<sup>6,22,23</sup>). In head-fixed mice, VTA-l dopamine neurons reportedly have uniform, RPE-like responses to conditioned stimuli<sup>3</sup>. To record VTA-l dopamine cells we infected the VTA with virus for Cre-dependent expression of channelrhodopsin (AAV-DIO-ChR2), in rats that express Cre recombinase under a tyrosine hydroxylase promoter<sup>3,14</sup> (see Methods). Optrodes (Fig.2a,b) recorded single-unit responses to brief blue laser pulses (Fig. 2c; Extended Data Figs.3,4; Supplementary Figure). We found 27 well-isolated VTA-l cells with reliable short-latency spikes, and considered these identified dopamine neurons.

All dopamine neurons were tonically-active, with relatively low firing rates (mean 7.7Hz; range 3.7–12.9Hz; compared to all VTA-l neurons recorded together with dopamine cells,  $p < 0.001$  one-tailed Mann–Whitney). They also had longer-duration spike waveforms ( $p < 5 \times 10^{-6}$ , one-tailed Mann–Whitney), although there were exceptions (Fig. 2d), confirming that waveform duration is an insufficient marker of dopamine cells *in vivo*<sup>3,24</sup>. A distinct cluster of VTA-l neurons ( $n=38$ , from the same sessions) had brief waveforms, higher firing rates ( $>20$ Hz; mean 41.3Hz, range 20.1–97.1Hz), and included no tagged dopamine cells. We presume these cells are GABAergic and/or glutamatergic<sup>3,25</sup>, and refer to them as “non-dopamine” below.

We recorded the same dopamine cells across multiple behavioral tasks. VTA-l dopamine cells responded strongly to randomly-timed food hopper clicks, and progressively less strongly when these clicks were made more predictable by preceding cues (Extended Data Fig.5). This is consistent with canonical RPE-like coding by dopamine cells in Pavlovian tasks<sup>2,3,26</sup>,

Based on evidence from anesthetized animals, it has been argued that altered dopamine levels measured with microdialysis arise from changes in the tonic firing rate of dopamine cells<sup>27</sup>, and/or the proportion of active versus inactive dopamine neurons<sup>28</sup>. However, in the bandit task tonic dopamine cell firing in each block of trials was strikingly indifferent to reward rate (Fig.2e,g). There was no significant change in the firing rates of individual dopamine cells – or any other VTA-l neurons - between higher- and lower-reward blocks (Fig.2f,h; see also<sup>29</sup> for concordant results in head-fixed mice). There was also no overall change in the rate at which dopamine cells fire bursts of spikes (Fig.2i). Furthermore, we never observed any dopamine cells switching between active and inactive states. The proportion of time dopamine cells spent inactive (long inter-spike-intervals) was very low, and did not change between higher- and lower-reward blocks (Fig.2i).

The anatomy of the VTA-NAc dopamine projection has been intensively investigated<sup>6,22,23</sup>, but given this apparent functional mismatch we reconfirmed that we were recording from the correct portion of VTA. Small injections of the retrograde tracer cholera toxin B (CTb) into

NAc core resulted in dense labeling of TH<sup>+</sup> neurons within the same VTA-l area as our optrode recordings (Extended Data Fig.3). Within the approximate recording zone 21% of TH<sup>+</sup> cells were also CTb<sup>+</sup>, and this is likely an underestimate of the fraction of NAc core-projecting VTA-l dopamine cells as our tracer injections did not completely fill NAc core. Hence our sample of n=27 tagged VTA dopamine cells (plus many more untagged cells) almost certainly includes NAc core-projecting neurons. Finally, in an additional rat we recorded two tagged VTA-l dopamine cells after infusing virus selectively into NAc core (Extended Data Fig.3). Both retrogradely-infected cells had firing patterns that closely resembled the other tagged dopamine cells in all respects, including lack of tonic firing changes with varying reward rate (Supplementary Figure). We conclude that changes in tonic VTA-l dopamine cell firing are not responsible for motivation-related changes in forebrain dopamine release.

### Tracking release on multiple timescales

Does NAc dopamine release track reward rate per se, as suggested in some theories<sup>30</sup>, or is this correlation driven by dynamic fluctuations in dopamine release that are too fast to resolve with microdialysis? We argued for the latter possibility based on voltammetry data<sup>5</sup>, but sought confirmation using an independent measure of dopamine release that can span different timescales. The dLight1 suite of genetically-encoded optical dopamine indicators was engineered by inserting circularly-permuted GFP into dopamine D1 receptors<sup>15</sup>. Binding of dopamine causes a highly specific increase in fluorescence (Fig.3a). We infused viruses into NAc to express either dLight1.1 (4 verified NAc placements from 3 rats) or the brighter variant dLight1.3b (6 verified NAc placements from 4 rats), and monitored fluorescence by fiber photometry. We observed clear NAc dopamine responses to Pavlovian reward-predictive cues, similarly to VTA dopamine cell firing (Extended Data Fig.5).

For the bandit task we first examined the dLight signal in 1 min bins (Fig.3b) for comparison to microdialysis. We again saw a clear relationship between NAc dopamine release and reward rate, in both cross-correlation and analysis of block transitions (Fig.3c,d). We then looked more closely at how this relationship arises. Rather than slowly-varying on a timescale of minutes, the dLight signal showed highly dynamic fluctuations within and between each trial (Fig.3e). We compared these fluctuations to instantaneous state values, and RPEs, estimated from a reinforcement learning model (a Semi-Markov Decision Process<sup>5</sup>). As we previously reported using voltammetry<sup>5</sup>, moment-by-moment NAc dopamine showed a strong correlation with state values (Fig.3f), visible as ramping up within trials when rewards were expected (Fig.3e). We also saw transient increases with less-expected reward deliveries, consistent with RPE (examined below). In every individual dLight session dopamine showed a stronger correlation with values than either RPEs or reward rate (Fig.3h and Extended Data Fig.6). Correlations with both state values and RPE were maximal to the dLight signal ~0.3s later, consistent with a brief lag caused by neural processing of cues and sensor response time (Fig.3g; with voltammetry we reported a ~0.4–0.5s lag<sup>5</sup>).

## Dopamine firing does not explain release

We next compared dopamine cell firing and release around bandit task events. External stimuli at *Light-On*, *Go cue*, and rewarded *Side-In* (food hopper click) each evoked a rapid firing increase (Fig.4a). These responses were observed in the great majority of dopamine cells (Fig.4c), although the magnitude of responses to different cues varied from cell to cell (Supplementary Figure). The NAc dLight signal also responded rapidly and reliably to each of these salient cues (Fig.4b,c), consistent with burst firing of dopamine cells driving dopamine release.

We also saw clear increases in NAc dopamine release as rats approached the start port (just before *Center-In*) and the food port (just before *Food-Port-In*). This fits well with the extensive voltammetry literature showing that motivated approach behaviors are accompanied by rapid increases in NAc core dopamine<sup>5,7-11</sup>. However, the VTA-l dopamine cell population did not show a corresponding increase in firing at these times (Fig.4a; see Extended Data Fig.7 for additional comparisons, including non-dopamine cells).

To better dissociate cue-evoked, and approach-related, dopamine activity we separated trials by short (<1s) and long (>2s) latencies (Fig.4d,e). Increases in dopamine cell firing were consistently locked to the cue onset at *Light-On*, preferentially for short-latency trials. All 25 dopamine cells with significant firing rate increases after *Light-On* were better aligned to *Light-On* than *Center-In* (Fig.4e). By contrast, increases in NAc dopamine release before *Center-In* were distinct from cue-evoked dopamine release (Fig.4d,e). dLight signals consistently increased before *Center-In* on long-latency trials (10/10 sessions), and also before *Food-Port-In* (9/10 sessions), without corresponding increases in dopamine firing (Fig.4f).

Finally we considered how event-related dopamine signals depend upon recent reward history. During the early part of each trial dopamine cell firing was not dependent on reward rate (Fig.5a) - despite the influence of reward rate on motivation (Fig.5b). Subsequently, the phasic response to the reward cue was reliably stronger when reward rate was lower (Fig. 5a), consistent with positive RPE encoding. When this reward cue was omitted dopamine cells paused firing, though encoding of negative RPEs was much weaker or absent, whether examined at the population level (Fig.5a,b) or individual cells (Extended Data Fig.8). It has been proposed that negative RPEs are encoded in the duration of dopamine pauses<sup>31</sup>, but this was observed in just 2/29 individual neurons. Similar results were obtained if reward expectation was estimated in other ways, including trial-based reinforcement learning models (actor-critic, Q-learning) or simply counting recent rewards (Extended Data Fig.8).

Dopamine release at *Side-In* also showed a clear, transient encoding of positive, but not negative, RPEs (Fig.5c,d). This dLight response was slightly delayed and prolonged compared to firing, consistent with time taken for release and reuptake<sup>32</sup>, but remained a subsecond phenomenon. Unlike firing, however, dLight signals early in each trial were greater when recent trials had been rewarded (Fig.5c), consistent with value coding. We observed this dependence on reward history even while the rat was not actively moving, but was maintaining a nosepoke in the center port waiting for the *Go* cue (Fig.5d). Overall, we

conclude that NAc dopamine release reflects both cue-evoked responses and reward expectation, and that only the former can be well accounted for by VTA-l dopamine cell firing.

## Discussion.

VTA-l provides the predominant source of dopamine to NAc core<sup>6,23,24</sup>. VTA-l dopamine cells consistently display RPE-like bursts in both Pavlovian<sup>3</sup> and operant<sup>13</sup> tasks, including VTA-l dopamine cells that project to NAc core. VTA bursts are thought to be particularly important for driving NAc dopamine<sup>32</sup>, and indeed we found that cue-evoked VTA bursts were matched by NAc release. However, we additionally found value-related patterns of NAc dopamine release that were not generated by firing of VTA-l dopamine cells, either on long (“tonic”) or short (“phasic”) timescales. Other dopamine subpopulations may carry distinct signals<sup>14,33,34</sup>, and we cannot rule out the possibility that firing of dopamine cell subpopulations not recorded from here produces value-related dopamine in NAc core. However, value-related firing has never been reported for any dopamine cells, across a wide range of studies. Our results suggest that NAc dopamine dynamics are controlled in different ways, at different times, for different functions, and that recording dopamine cells is important but not sufficient for understanding dopamine signals<sup>35</sup>.

Release from dopamine terminals is powerfully controlled by local, non-spiking mechanisms<sup>36–40</sup>. For example, NAc dopamine release is modulated by the basolateral amygdala even when VTA spiking is pharmacologically suppressed<sup>41,42</sup>. It has been noted for decades that local control of dopamine release might achieve distinct functions to dopamine cell spiking<sup>36,43</sup>, but this has not been incorporated into theoretical views of dopamine. Distinct striatal subregions contribute to different types of decision, and may influence their own dopamine release according to need<sup>44</sup>. It remains to be determined just how local this local control of dopamine release can be. One limitation shared by the three ways that we measured dopamine release is that they all sample on a 100 $\mu\text{m}+$  spatial scale. There is evidence from *in vivo* microscopy that dopamine release may be heterogeneous at considerably smaller scales<sup>15</sup>.

Our results do not support the existence of any separate “tonic” dopamine signal that could mediate motivational effects of dopamine. Instead, dopamine shifts that appear slow if measured slowly (with microdialysis) resolve into rapid fluctuations if measured rapidly (with voltammetry or dLight). Furthermore, recordings of identified VTA dopamine cells by ourselves and others<sup>30</sup> provide strong evidence against the idea<sup>29</sup> that changes in tonic dopamine cell firing drive tonic changes in dopamine release. While tonic firing can be altered by lesions or drug manipulations<sup>28</sup>, we are not aware of sustained changes in firing rate in any behavioral task. Firing can ramp downwards on a ~1s timescale during anticipation of motivationally-relevant events<sup>45,46</sup>. However, this decline is the opposite of what would be required to boost dopamine release with reward expectation, and instead seems more akin to a sequence of transient negative prediction errors<sup>47</sup>. Although sustained signals encoding ongoing reward rate could be computationally useful<sup>30</sup>, dopamine instead provides rapidly-fluctuating error and value signals. It remains possible that sustained



signals are computed at a subsequent step, by intracellular signaling pathways downstream of dopamine receptors.

Many groups have observed ramping dopamine release as rats approach rewards<sup>5,7-11</sup>, consistent with encoding escalating reward expectations. Some have argued that these dopamine ramps simply reflect RPEs, by supposing that rats either rapidly forget values<sup>48</sup>, or that they have a warped set of state representations<sup>49</sup>. This latter idea is not supported by our observation that ramping is rapidly modulated from trial to trial based on updated reward expectations - becoming stronger within a short sequence of successive rewards while RPE-like responses to cues become weaker (Fig. 3e). More generally, any theory in which dopamine solely conveys RPEs (learning signals) cannot account for the very well-established connection between ongoing mesolimbic dopamine and motivation<sup>16</sup>. The NAc core is not needed for highly-trained responses to conditioned stimuli, but is particularly important when deciding to perform time-consuming work to obtain rewards<sup>50</sup>. NAc core dopamine appears to provide an essential dynamic signal of how worthwhile it is to allocate time and effort to work<sup>5,44</sup>, even though this signal is not present in VTA dopamine cell firing.

## Methods.

### Animals.

All animal procedures were approved by the University of Michigan or University of California San Francisco Institutional Committees on Use and Care of Animals. Male rats (300–500g, either wild-type Long-Evans or TH-Cre+ with a Long-Evans background<sup>51</sup>) were maintained on a reverse 12:12 light:dark cycle and tested during the dark phase. Rats were mildly food deprived, receiving 15 g of standard laboratory rat chow daily in addition to food rewards earned during task performance. No sample size precalculation was performed.

### Behavior.

Pretraining and testing were performed in computer-controlled Med Associates operant chambers (25 cm × 30 cm at widest point) each with a five-hole nose-poke wall, as previously described<sup>5</sup>. Bandit task sessions used the following parameters: block lengths were 35–45 trials, randomly selected for each block; hold period before Go cue was 500–1500 ms (uniform distribution); left/right reward probabilities were 10,50,90% (for electrophysiology, photometry, voltammetry, and previously reported microdialysis rats<sup>5</sup>, or 20,50,80% (newly reported microdialysis rats).

Current reward rate was estimated using a simple, time-based leaky-integrator<sup>52</sup>. Reward rate was incremented each time a reward was received, and decayed exponentially at a rate set by parameter  $\tau$  (the time in s for the reward rate to decrease by ~63%,  $1-1/e$ ). For all analyses,  $\tau$  was selected based on the rat's behavior, maximizing the (negative) correlation between reward rate and log(latency) in each session. The correlations between forebrain dopamine and reward rate were not highly sensitive to this choice of  $\tau$  (Extended Data Fig. 1).

To classify block transitions as “increasing” or “decreasing” in reward rate, we compared the average leaky-integrator reward rate in the last 5 min of a block to the average reward rate in the first 8 min of the subsequent block.

Rats used for electrophysiology and photometry also performed a Pavlovian approach task, in the same operant chamber with the houselight on throughout the session. Three auditory cues (2kHz, 5kHz, 9kHz) were associated with different probabilities of food delivery (counterbalanced across rats). Cues were played as a train of tone pips (100ms on / 50ms off) for a total duration of 2.6s followed by a delay period of 500ms. Cues, and unpredicted reward deliveries, were delivered in pseudorandom order with a variable inter-trial interval (15–30s, uniform distribution).

### Microdialysis.

**Surgery.**—Rats were implanted bilaterally with guide cannulae (CMA, #830 9024) in cortex and striatum. One group (n=8) received one guide cannula targeting prelimbic and infralimbic cortex (AP +3.2 mm, ML 0.6 mm relative to bregma; DV 1.4 mm below brain surface) and another targeting dorsomedial striatum and nucleus accumbens in the opposite hemisphere (AP +1.3, ML 1.9, DV 3.4). Both implants were angled 5 degrees away from each other along the rostral-caudal plane. A second group (n=4) received one guide cannula targeting anterior cingulate cortex (AP +1.6, ML 0.8, DV 0.8) and another targeting accumbens (core/shell in the opposite hemisphere at AP +1.6, ML 1.4, DV 5.5 (n=2) or AP +1.6, ML 1.9, DV 5.7 (n=2)). Implant sides were counterbalanced across rats. Animals were allowed to recover for 1 week prior to retraining.

**Chemicals.**—Water, methanol, and acetonitrile for mobile phases were Burdick & Jackson HPLC grade, purchased from VWR (Radnor, PA). All other chemicals were purchased from Sigma Aldrich (St. Louis, MO) unless otherwise noted. Artificial cerebrospinal fluid (aCSF) was comprised of 145 mM NaCl, 2.68 mM KCl, 1.40 mM CaCl<sub>2</sub>, 1.01 mM MgSO<sub>4</sub>, 1.55 mM Na<sub>2</sub>HPO<sub>4</sub>, and 0.45 mM NaH<sub>2</sub>PO<sub>4</sub>, adjusted pH to 7.4 with NaOH. Ascorbic acid (250 nM final concentration) was added to reduce oxidation of analytes.

**Sample Collection and HPLC-MS.**—On testing day, animals were placed in the operant chamber with the houselight on. Custom-made concentric polyacrylonitrile membrane microdialysis probes (1 mm dialyzing AN69 membrane; Hospal, Bologna, Italy) were inserted bilaterally into guide cannula and perfused continuously (Chemyx Inc., Fusion 400) with aCSF at 2 µL/min for 90 min to allow equilibration. After 5 min baseline collection the houselight was extinguished, cueing the animal to bandit task availability. Sample collection continued at 1 min intervals and samples were immediately derivatized<sup>53</sup> with 1.5 µL sodium carbonate, 100 mM; 1.5 µL BzCl, 2% (v/v) BzCl in acetonitrile; and 1.5 µL isotopically labeled internal standard mixture diluted in 50% (v/v) acetonitrile containing 1% (v/v) sulfuric acid, and spiked with deuterated ACh and choline (C/D/N isotopes, Pointe-Claire, Canada) to a final concentration of 20 nM. Sample series collection alternated between the two probes at 30-second intervals in each of 26 sessions, except for one session in which a broken membrane resulted in just one series (51 sample series total). Samples were analyzed using Thermo Scientific UHPLC systems (Accela, or Vanquish Horizon



interfaced to a Quantum Ultra triple quadrupole mass spectrometer fitted with a HESI II ESI probe), operating in multiple reaction monitoring. Five  $\mu\text{L}$  samples were injected onto a Phenomenex core-shell biphenyl Kinetex HPLC column (2.1 mm  $\times$  100 mm). Mobile phase A was 10 mM ammonium formate with 0.15% formic acid, and mobile phase B was acetonitrile. The mobile phase was delivered an elution gradient at 450  $\mu\text{L}/\text{min}$  as follows: initial, 0% B; 0.01 min, 19% B; 1 min, 26% B; 1.5 min, 75% B; 2.5 min, 100% B; 3 min, 100% B; 3.1 min, 5% B; and 3.5 min, 5% B. Thermo Xcalibur QuanBrowser (Thermo Fisher Scientific) was used to automatically process and integrate peaks. Each of the >100,000 peaks were visually inspected individually to ensure proper integration.

**Analysis.**—All neurochemical concentration data were smoothed with a 3-point moving average ( $y' = [0.25*(y-1) + 0.5(y) + 0.25*(y+1)]$ ) and z-score normalized within each session to facilitate between-session comparisons. For each target region, a cross-correlogram was generated for each session and the average of the sessions was plotted. 1% confidence boundaries were generated for each subplot by shuffling one time series 100,000 times and generating a distribution of correlation coefficients for each session. Multiple regression models were generated using the *regress* function in MATLAB, with the neurochemical as the outcome variable and behavioral metrics as predictors. Regression coefficients were determined significant at three alpha levels (0.05, 0.0005, 0.000005), after Bonferroni-correction for multiple comparisons ( $\alpha / (21 \text{ chemicals} * 7 \text{ regions} * 9 \text{ behavioral regressors})$ ). For analysis of block transitions data were binned into 3 min epochs, discarding the sample that included the transition time.

### Electrophysiology.

Rats ( $n=25$ ) were implanted with custom designed drivable optrodes, each consisting of 16 tetrodes (constructed from 12.5 $\mu\text{m}$  nichrome wire, Sandvik, Palm Coast, FL) glued onto the side of a 200 $\mu\text{m}$  optic fiber and extending up to 500 $\mu\text{m}$  below the fiber tip. During the same surgery, we injected 1 $\mu\text{l}$  of AAV2/5-EF1a-DIO-ChR2(H134R)-EYFP into the lateral VTA (AP 5.6, ML 0.8, DV 7.5) or NAc core (AP 1.6, ML 1.6, DV 6.4). Wideband (1–9000Hz) brain signals were sampled (30,000 samples/s) using Intan digital headstages. Optrodes were lowered at least 80 $\mu\text{m}$  at the end of each recording session. Individual units were isolated offline using a MATLAB implementation of MountainSort<sup>54</sup> followed by careful manual inspection.

**Classification.**—To identify whether an isolated VTA-I unit was dopaminergic (TH+), we used the stimulus-associated latency test<sup>55</sup>. Briefly, at the end of each experimental session, we connected the optrode to a laser diode and delivered light pulse trains of different widths and frequencies. For a unit to be identified as light responsive it needed to reach the significance level of  $p < 0.001$  for 5ms and 10ms pulse trains. We also compared the light evoked waveforms (within 10ms of laser pulse onset) to session-wide averages; all light-evoked units had a Pearson correlation coefficient of  $> 0.9$ . Dopamine neurons were successfully recorded from four rats with VTA-I virus infusions (IM657, 1 unit; IM1002, 3 units; IM1003, 15 units; IM1037, 9 units) and one rat with NAc core virus (IM-1078, 2 units). Peak width was defined as the full-width-at-half-maximum of the most prominent negative component of the aligned, averaged spike waveform. Non-tagged VTA neurons

with session-wide firing rate > 20Hz and peak width < 200 $\mu$ s were classified as non-dopamine cells. To ensure that we were comparing dopamine and non-dopamine cells within the same subregions, we only analyzed non-dopamine cells recorded during sessions with at least one optically-tagged dopamine cell.

**Analysis.**—Spike bursts were detected by the conventional “80/160 template” approach<sup>56</sup>: each time an inter-spike-interval of 80 ms or less occurs, these and subsequent spikes are considered part of a burst until there is an interval of 160 ms or more. For comparison of “tonic” firing to reward rate, dopamine spikes were counted in 1 min bins. To examine faster changes, spike density functions were constructed by convolving spike trains with a Gaussian kernel with variance 20ms. To determine how quickly a neuron responded to a given cue, we used 40ms bins (sliding in steps of 20ms) and used a shuffle test (10,000 shuffles) for each time bin comparing the firing rate after cue onset to firing rate in the 250 ms immediately preceding the cue. The first bin at which the post cue firing rate was significantly ( $p < 0.01$ , correcting for multiple comparisons) greater than baseline firing was considered the time to cue response.

Peak firing rate was calculated as the maximum (Gaussian-smoothed) firing rate of each trial in a 250 ms window after *Side-In* for rewarded trials, and the valley was calculated as the minimum firing rate in a 2 s window, starting one second after *Side-In* for unrewarded trials.

To calculate a ramp angle during approach behaviors we smoothed mean firing rates with a 50ms Gaussian kernel, detected the maximum/minimum of the resulting signal in a 0.5s window prior to each event (*Center-In* or *Food-Port-In*) and measured the signed angle connecting the two extrema. To compare firing rates in “high” and “low” reward blocks, for each session we performed a median split of average leaky-integrator reward rate in each block.

### **Voltammetry and computational model.**

Fast-scan cyclic voltammetry results shown here reanalyze data previously presented in detail<sup>5</sup>. Within-trial estimates of state value and reward prediction errors were calculated using a semi-Markov decision process reinforcement learning model, exactly as previously described<sup>5</sup>.

### **Photometry.**

We used a viral approach to express the genetically encoded optical dopamine sensor dLight<sup>15</sup>. Under isoflurane anesthesia, 1 $\mu$ L of AAV9-CAG-dLight ( $1 \times 10^{12}$  vg/mL - UC Davis vector core) was slowly (100nL/min) injected (Nanject III, Drummond, Broomall, PA) through a 30 $\mu$ m glass micropipette in ventral striatum bilaterally (AP:1.7mm, ML: 1.7mm, DV: -7.0mm). During the same surgery optical fibers (400 $\mu$ m core, 430 $\mu$ m total diameter) attached to a metal ferrule (Doric) were inserted (target depth 200 $\mu$ m higher than virus) and cemented in place. Data were collected >3 weeks later, to allow for dLight expression.

For dLight excitation blue (470nm) and violet (405nm; control) LEDs were sinusoidally modulated at distinct frequencies (211Hz, 531Hz respectively<sup>57</sup>). Both excitation and

emission signals passed through minicube filters (Doric) and bulk fluorescence was measured with a femtowatt detector (Newport, Model 2151) sampling at 10KHz. Demodulation produced separate 470nm (dopamine) and 405nm (control) signals, which were then rescaled to each other via a least-square fit<sup>57</sup>. Fractional fluorescence signal ( $dF/F$ ) was then defined as  $(470-405\_fit)/405\_fit$ . For all analyses this signal was downsampled to 50Hz and smoothed with a 5-point median filter. For presentation of 470nm and 405nm signals separately, see Extended Data Fig.7.

Data from an optic fiber placement were included in analyses if the fiber tip was in NAc, and the fluorescence response of one to at least one task cue had a Z-score of  $>1$ . These criteria excluded one rat, and yielded three rats/four placements (IM1065-left, IM1066-bilateral, IM1089-right) for dLight1.1, and four rats/six placements (IM1088-bilateral, IM1105-right, IM1106-bilateral, IM1107-right) for dLight1.3b. Similar results were obtained for dLight1.1 and dLight1.3 (Extended Data Fig.7) so data were combined.

To calculate a ramp angle during approach behaviors we detected the maximum/minimum of the resulting signal in a 0.5s window prior to each event (*Center-In* or *Food-Port-In*) and measured the signed angle connecting the two extrema.

### Affinity and molecular specificity of dLight1.3b.

*In vitro* measurements were performed as previously described<sup>15</sup>. Briefly, HEK293T (ATCC CRL#1573) cells were cultured and transfected with plasmids encoding dlight1.3b driven by a CMV promoter, and washed with HBSS (Life Technologies) supplemented with  $Ca^{2+}$  (4mM) and  $Mg^{2+}$  (2 mM) before imaging. Imaging was performed using a 40X oil-based objective on an inverted Zeiss Observer LSN710 confocal microscope with 488nm/513nm (excitation/emission) wavelengths. For testing the sensor's fluorescence responses, neurotransmitters were directly applied to the bath during time-lapse imaging, in at least two independent experiments. Titrations of dopamine and norepinephrine were obtained by performing 10-fold serial dilutions to achieve 8 different concentrations. All other neurotransmitters were tested at three sequential concentrations (100nM, 1  $\mu$ M, 10  $\mu$ M). All neurotransmitter concentrations were obtained by dilution from a 1 mM stock concentration in HBSS, prepared fresh. Raw fluorescence intensities from time lapse imaging were quantified on Fiji; each region of interest (ROI) was manually drawn on the membrane of individual cells. Fluorescent fold change ( $F/F$ ) was calculated as  $F_{peak}$  (averaged fluorescence intensity of 4 frames) -  $F_{basal}$  (averaged fluorescence intensity of 4 frames before addition of ligands) /  $F_{basal}$ . Graphs and statistical analysis were performed using GraphPad Prism 6. Data points were analyzed with a one-site specific binding curve fit to obtain  $K_d$  values. In box-and-whisker plots, the box covers the 25% to 75% range and whiskers extend from minimum to maximum values.

### Availability of Reagents, Code, and Data.

The AAV.Synapsin.dLight1.3b virus used in this study has been deposited with Addgene ([www.addgene.org](http://www.addgene.org)). Custom MATLAB code is available upon request to J.D.B. All data will be available through the Collaborative Research in Computational Neuroscience (CRCNS.org) data sharing website.

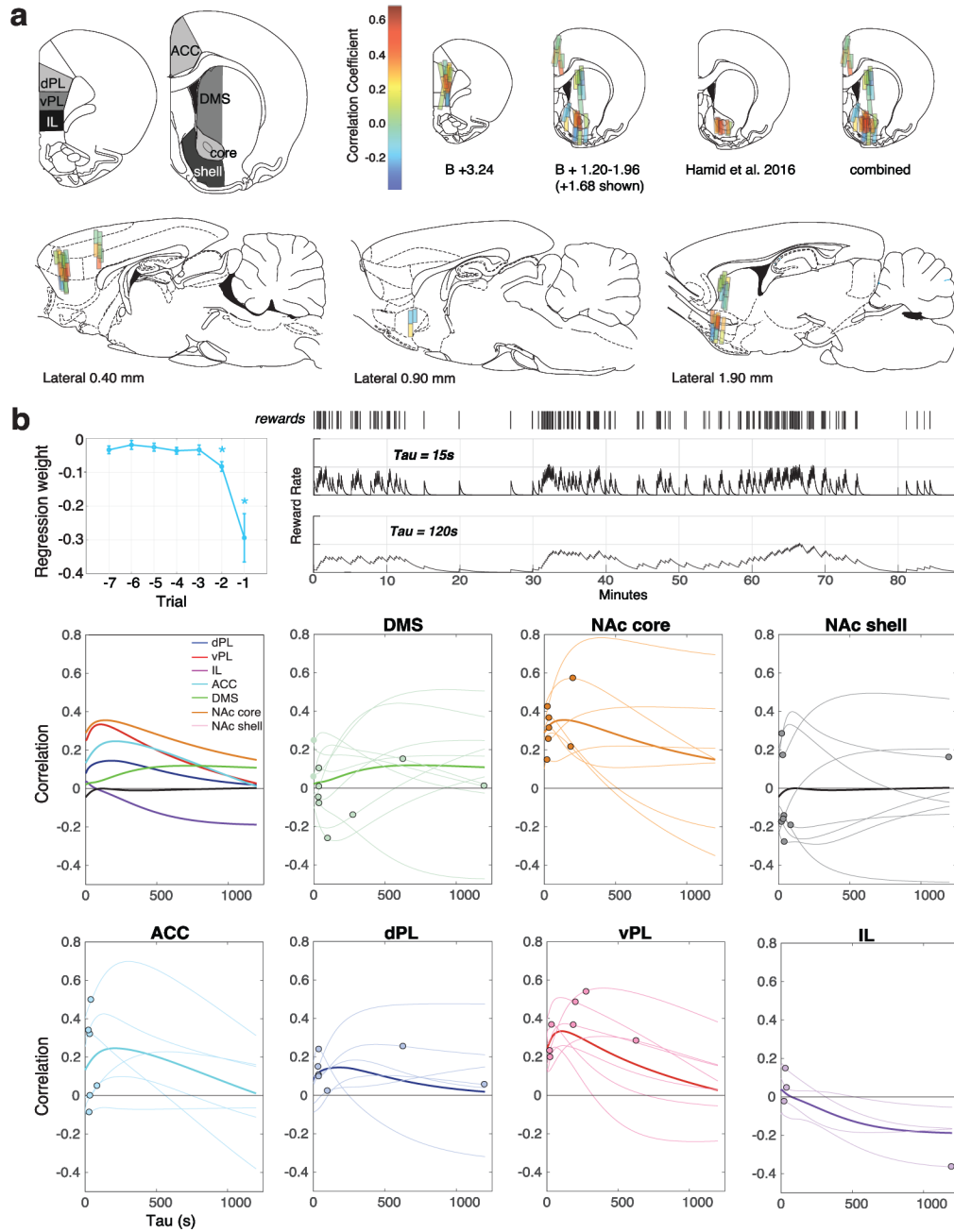
## Extended Data

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Extended Data Figure 1.**

**a**, Top left, anatomical definitions of the subregions examined with microdialysis. Atlas section schematics are from ref. 58. Other panels map the correlation between dopamine release and reward rate at individual probe placements in coronal (mm from bregma, B) and sagittal (mm from midline) planes. Color bar shows strength of correlation. **b**, Top left, Regression analysis showing dependency of (log-) latency on the outcome of recent trials, during microdialysis sessions (n=26 sessions, 7113 trials, from 12 rats; error bars show SEM). Asterisks indicate average regression weights significantly different from zero (t-test,  $p < 0.05$ ). Top right, Illustration of how the reward rate definition depends on the time

Author Manuscript

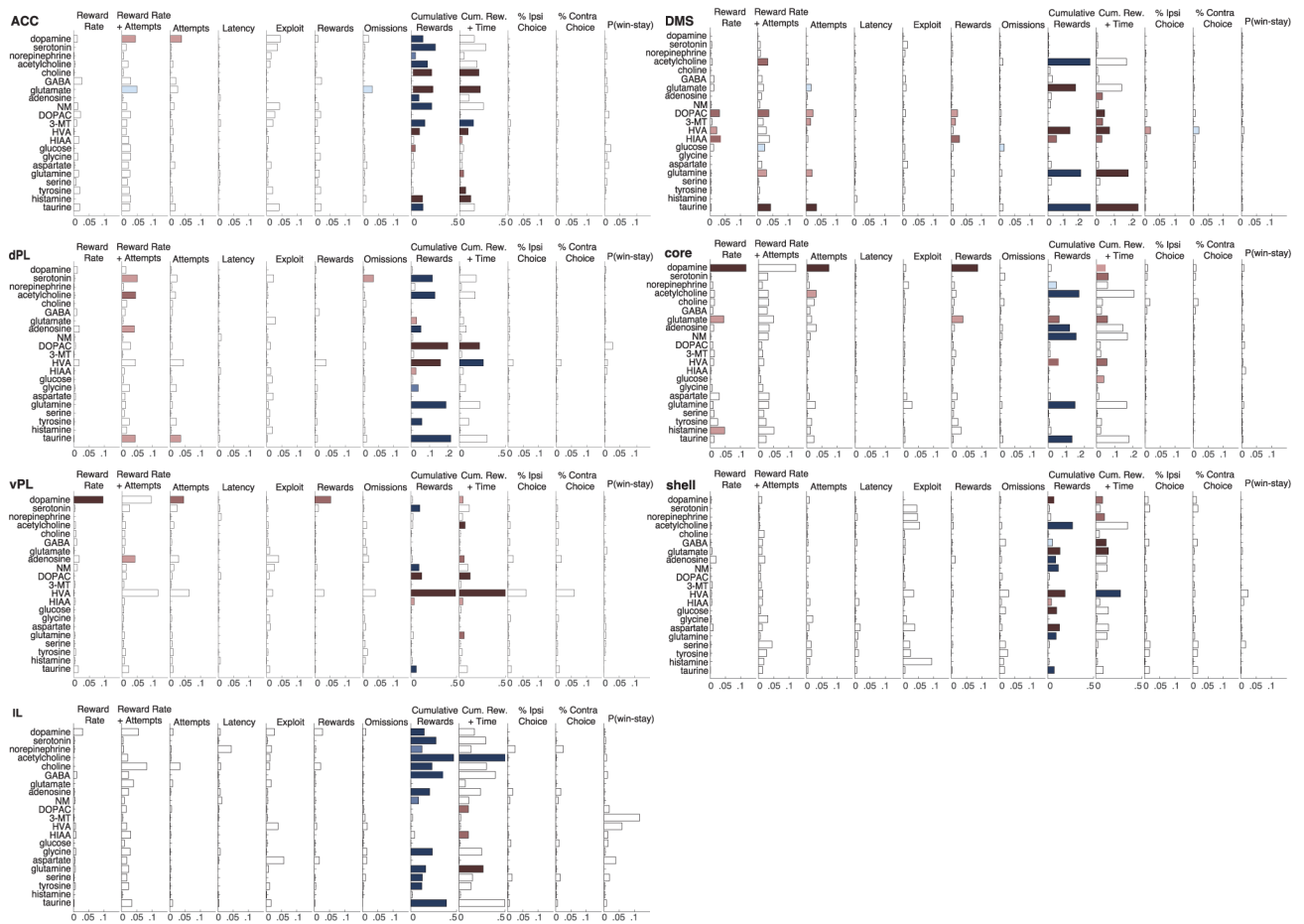
Author Manuscript

Author Manuscript

Author Manuscript

constant ( $\tau$ ) of the leaky integrator. Below, dopamine : reward rate correlations as a function of  $\tau$ . In the main figures  $\tau$  was chosen (from a range of 1-1200s) to maximize the (negative) correlation between reward rate and (log) latency in each session. Thin lines represent individual sessions, with the best fit  $\tau$  used in regression analyses indicated by a dot. Thick lines indicate the average of all dopamine : reward rate correlations for a given  $\tau$  within each subregion. Overall behavioral metrics were similar between sessions sampling from each of the seven subregions (mean rewards/min: range 1.42-1.77, ANOVA  $F(6,44)=0.58$ ,  $p=0.746$ ; mean attempts/min: range 3.32-3.97,  $F(6,44)=0.40$ ,  $p=0.872$ ; mean latency: range 5.99-8.02,  $F(6,44)=0.27$ ,  $p=0.948$ ).





**Extended Data Figure 2. Correlations between all neurochemicals and a range of behavioral factors.**

Bars represent  $R^2$  values for linear tests between each analyte (rows) and behavioral covariates (columns). In models with more than one covariate, bar length indicates the  $R^2$  for the full model. Negative relationships are reported in blue and positive relationships are in red. P-values are reported at three alpha levels (0.05, 0.0005, 0.000005) after Bonferroni correction for multiple comparisons (7 subregions  $\times$  21 analytes  $\times$  12 measures). To calculate reward rate, we averaged the leaky-integrator-estimated reward rate in 1 min bins defined by the start and end of each dialysis sample. ‘Attempts’ is the number of initiated trials (including trials that resulted in an error) in each dialysis minute. Attempts and reward rate and an interaction term were combined in a single model (column 2) to examine whether adding attempts could explain additional variance in the analyte signal that could not be explained by reward rate alone. ‘Latency’ is the average of the (log)-latency in each minute. ‘Exploit’ is the proportion of choices of the higher reward probability option, in the last half of blocks for which the two ports had different probabilities. ‘Rewards’ and ‘Omissions’ were defined as the number of rewarded and unrewarded trials in each min, respectively. ‘Cumulative Rewards’ and ‘Time’ were included in the same regression model to estimate progressive factors such as satiety, and possible slow timescale increases or decreases in analyte concentration across the session. Cumulative Rewards represents the

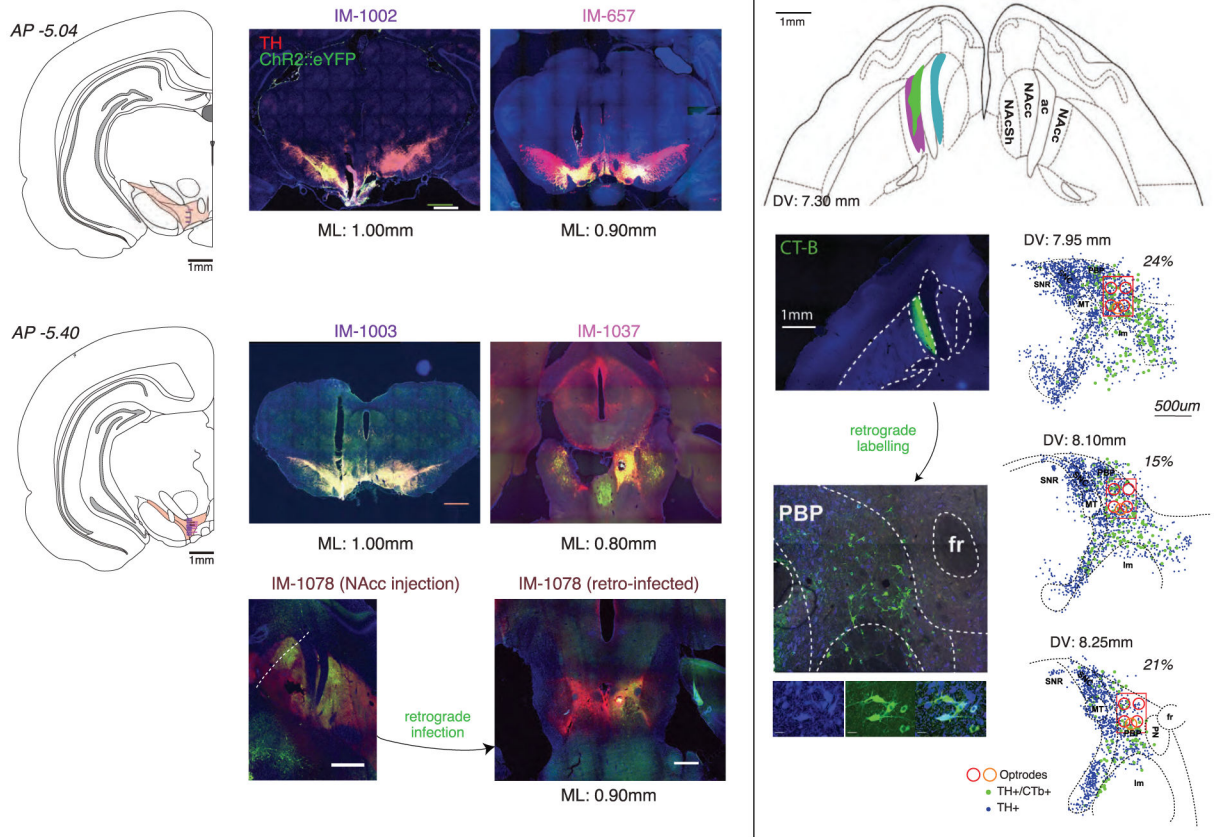
total number of rewards received by the end of the current dialysis minute, and Time was simply the number of min elapsed since the session began. Bars in this column show color when only the coefficient for the cumulative reward variable was significant. %Ipsi and %Contra represent the fraction of choices to ipsi- or contra-versive ports (relative to probe location in the brain) in each minute, independent of block probability. P(win-stay) is the probability of repeating the previous choice, given the previous choice was rewarded.

Author Manuscript

Author Manuscript

Author Manuscript

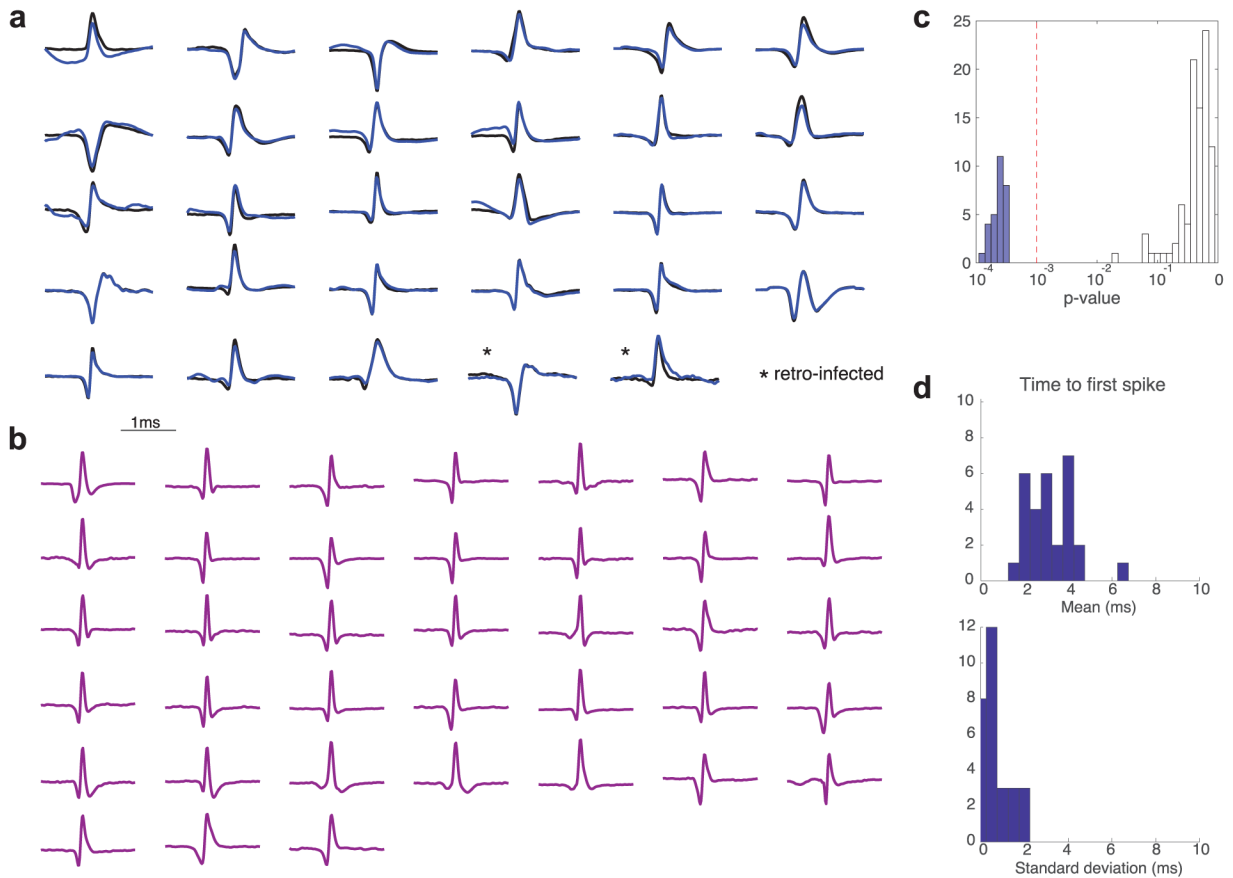
Author Manuscript



**Extended Data Figure 3. Histological analysis of electrophysiological recording locations.**

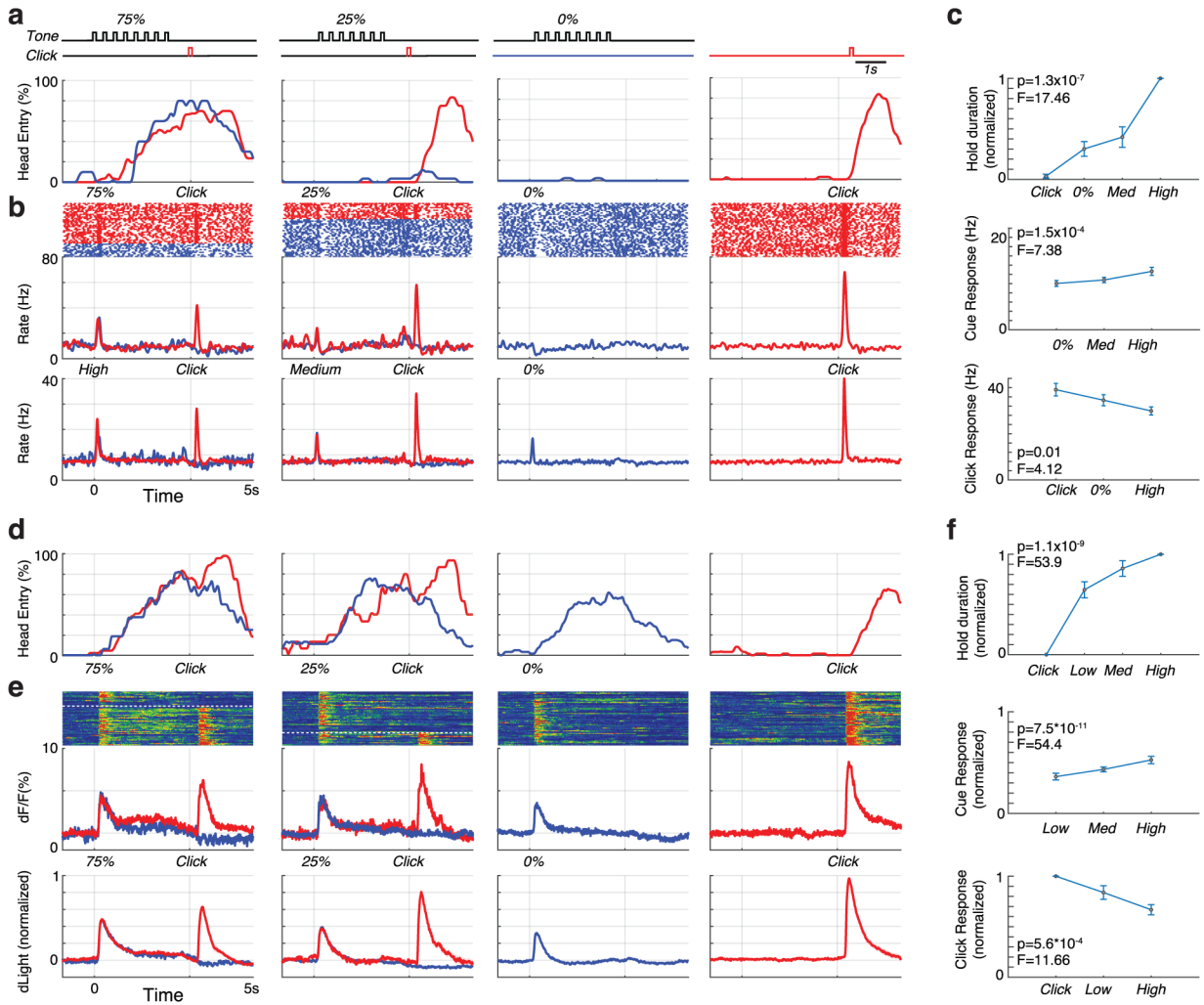
*Left*, Atlas locations (schematics from ref. 58) and histology photomicrographs for each rat (IM-657, IM-1002, IM-1003, IM-1037, IM-1078) from which opto-tagged dopamine cells were obtained. Red: TH-staining; green: ChR2::eYFP; blue: DAPI. Scale bars: 1mm. IM-1037 and IM-1078 brains were sliced horizontally, so fiber tracks appear as a circle. Font colors for rat ID# correspond to colors of tick marks in coronal atlas sections, indicating estimated recording locations for opto-tagged dopamine cells. For IM-1078 virus was injected into NAc core, and retrogradely-infected dopamine neurons were recorded in VTA.

*Right*, Retrograde tracing of CTb from NAc core (top) to VTA-l (bottom). Top panel shows approximate extent of NAc labeling in each of the 3 rats (each rat indicated by a different color). Bottom left panels show close-ups of TH labeling (blue), CTb (green) and merged image. Bottom right panels show reconstructed locations of TH+ and double-labeled TH+/CTb+ midbrain neurons, on horizontal atlas sections. Estimated optrode locations are shown by red circles (or orange circle, in the case of the retrograde tagging rat IM-1078). Labelled neurons were counted within the red rectangles that span the AP and ML extent of estimated recording locations. Percentages shown are the fraction of TH+ neurons that are also CTb+.



**Extended Data Figure 4. Identification of light-responsive cells.**

**a.** Average waveforms of optogenetically-identified dopamine neurons. Average light-evoked waveforms are shown in blue and session-wide average waveforms are in black. All spikes within 10ms of laser onset were used to construct light-evoked waveform average. Averaged waveforms are normalized to have similar total peak-valley voltages (see Extended Data Fig. 5 for individual voltage ranges). **b.** Session-wide average waveform for non-dopamine cells. **c.** Opto-tagging p-value for all units plotted in log-scale, showing a strong bimodal distribution. To classify cells as light-responsive we used a threshold of  $p < 0.001$ . **d.** Times to first spike after laser onset, showing mean for each identified dopamine neuron, and standard deviation (jitter).



### Extended Data Figure 5. Dopaminergic responses to Pavlovian cues.

**a**, Tone pips were followed by reward delivery (“Click”) with different probabilities (zero, medium, high) depending on the tone pitch. During prior training (average, 15.6 sessions; range 2-26) rats had learned about these different probabilities, as indicated by their corresponding scaled likelihood of entering the food port during cue presentation. “Head entry %” indicates proportion of trials for which the rat was at the food port at each moment in time, for one example session. Red, blue indicate rewarded, unrewarded trials. This rat was more likely to go to the food port during the cue that was highly (75%) predictive of rewards compared to the other cues (25% and 0%; one-way ANOVA,  $F=11.1$ ,  $p<1.2\times 10^{-6}$ ). Unpredictable reward delivery (right) prompts rapid approach. Bottom, raster plots and perievent time histograms from an identified dopamine neuron during that same session. **b**, Averaged firing for identified dopamine cells ( $n=27$ ) in this task. “High”/“Medium” tones were either 75%/25% predictive of reward ( $n=9$  cells), or 100%/50% ( $n=18$ ) respectively. Data on each individual dopamine neuron is presented in Extended Data Fig. 5. **c**, Behavior (top), cue response (middle), and click response (bottom) for all Pavlovian sessions with opto-tagged dopamine cells. Statistical comparisons were all one-way ANOVA, using Food Port head entry during 0.3s-3s epoch relative to cue onset, and peak firing rate during 0.5s

duration epochs after cue onset or food hopper clicks. dLight **d-f**, same as above except for dLight measurements (n=10 sessions total). All dLight sessions used tones with 75, 25, and 0% reward probability, and ANOVA tests examined peak signal within 1s of cue onset or food hopper clicks.

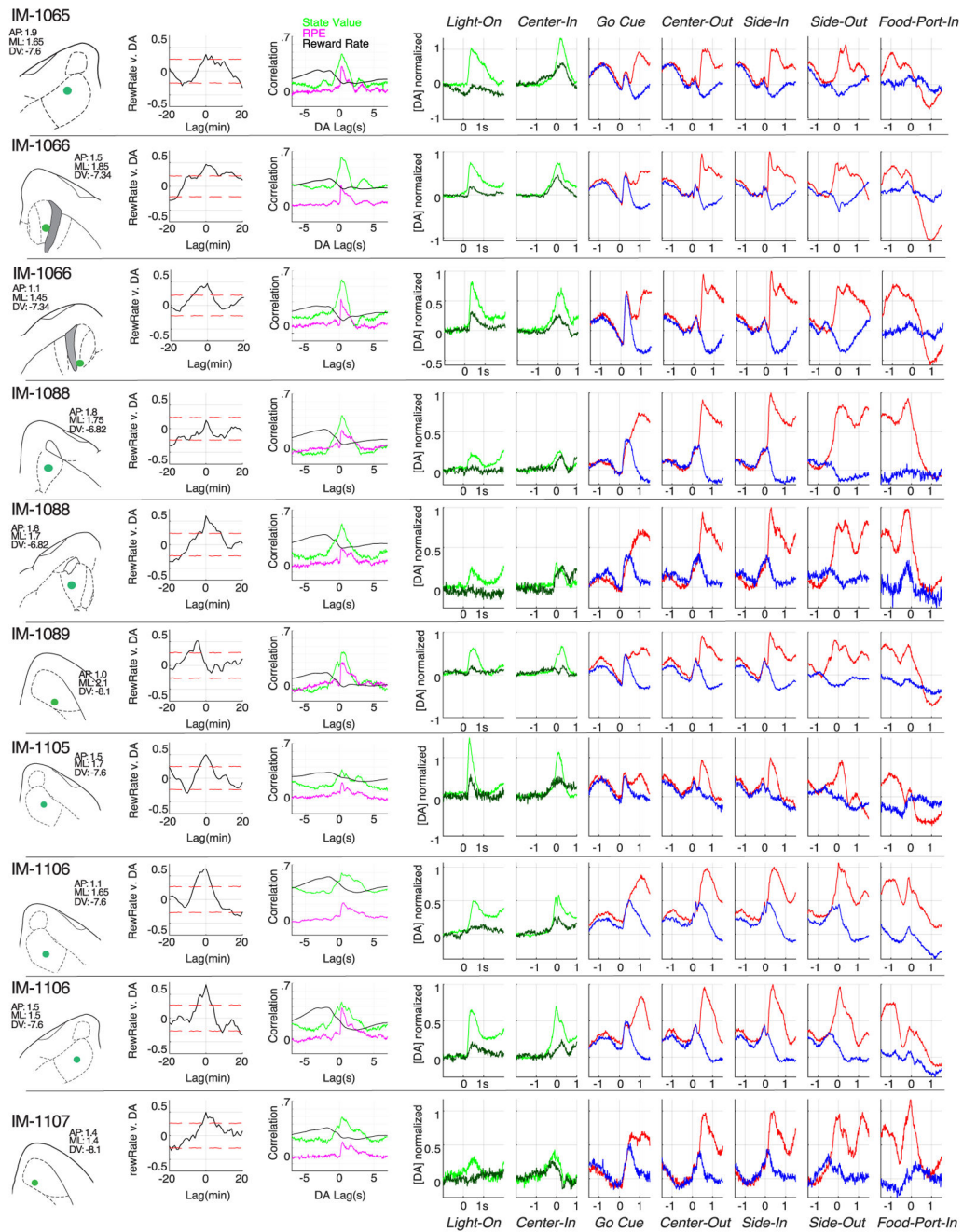
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Extended Data Figure 6. Results from each dLight recording session.**

Each row shows a distinct optic fiber placement, and the corresponding recording session that was included in data analyses. For two rats (IM-1066, IM-1088) we obtained bilateral NAc dLight recordings. From left to right, panels show histologically-determined NAc location of fiber tip (within horizontal brain atlas section, including atlas coordinates<sup>58</sup>), long timescale cross-correlation with reward rate (as in Fig. 3c), short timescale cross-correlation with reward rate (black), SMDP state value (green) and RPE (magenta; as in Fig 3f); event-aligned averages (as in Fig. 4b, but including more events). For *Light-On* and

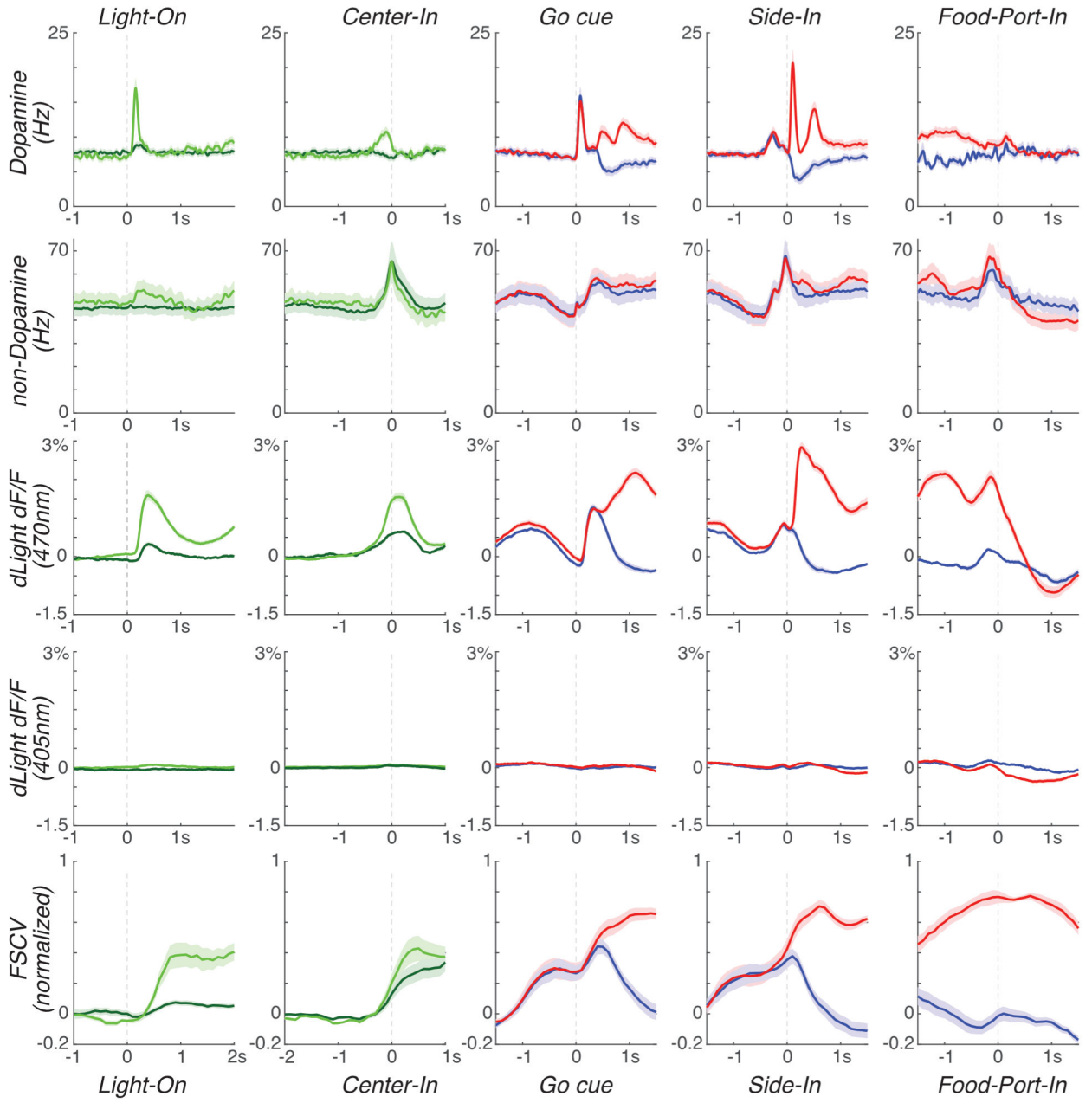
*Center-In* alignments data are split by latencies <1s (light green) or >2s (dark green; as in Fig. 4d), for other alignments data are split by rewarded (red) and unrewarded (blue) trials.

Author Manuscript

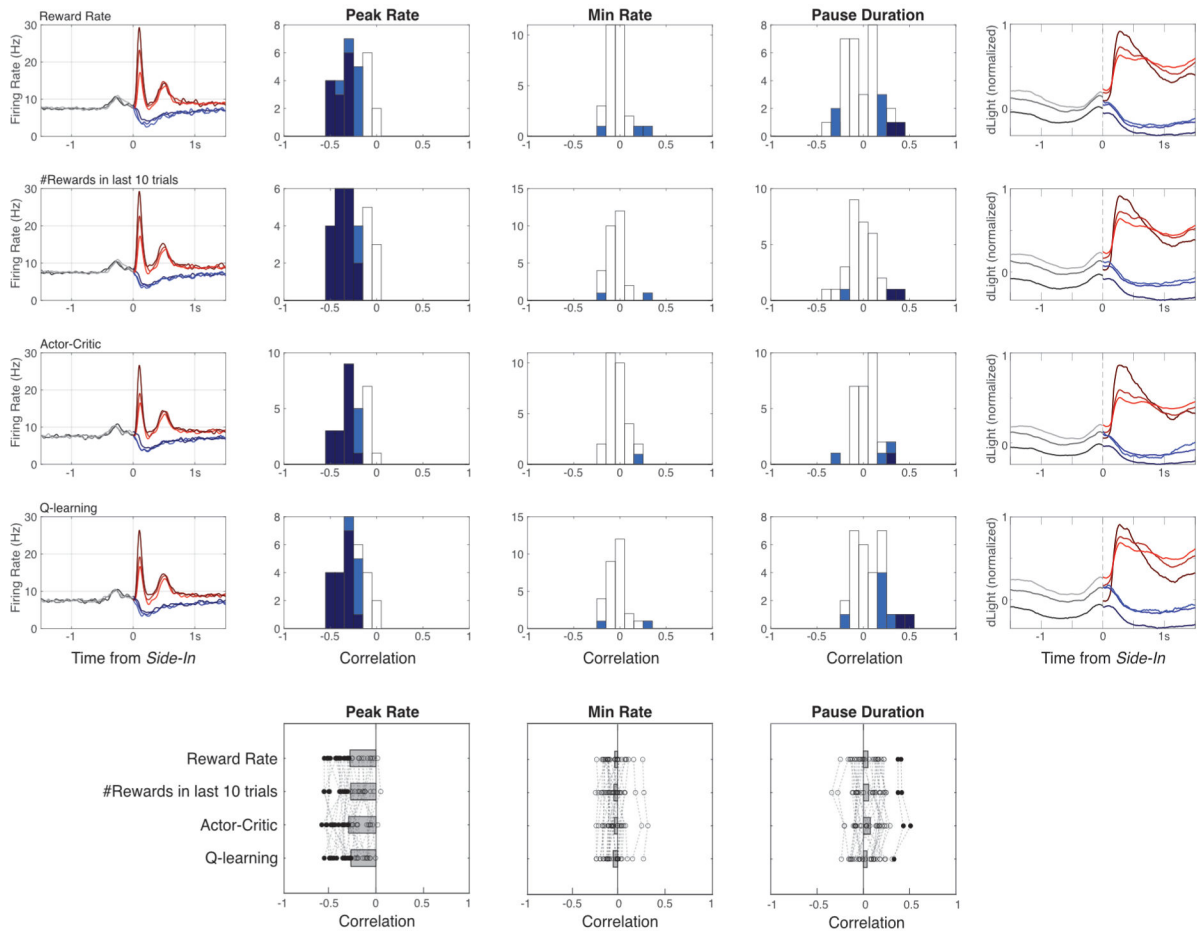
Author Manuscript

Author Manuscript

Author Manuscript



**Extended Data Figure 7. Comparing event-aligned activity between different signals.** Format is as Fig. 4. dLight fluorescence is here shown separately for 470nm (control) excitation. To note: 1) Rapid, behavior-linked dLight fluorescence changes occur at 470nm, as expected, not in the control 405nm band. 2) Distinct timing of spiking, dLight, and voltammetry (FSCV) responses to cue onsets; 3) Non-dopamine cell firing is much more variable (wider error bands), but on average shows activity during movements: starting just before *Center-In* (irrespective of latency), just before *Side-In*, and just before *Food-Port-In*.



**Extended Data Figure 8. Different methods for calculating reward expectation produce similar results.**

Left column, average firing rate of dopamine cells around *Side-In*, broken down by terciles of reward expectation, based either on recent reward rate (top; same as Fig. 5a), # of rewards in previous 10 trials, state value ( $V$ ) of an actor-critic model, or state value ( $Q_{\text{left}} + Q_{\text{right}}$ ) of a Q-learning model. The actor-critic and Q-learning models were both trial-based, rather than evolving continuously in time. The actor-critic model estimated the overall probability of receiving a reward on each trial,  $V$ , using the update rule  $V' = V + \alpha (\text{RPE})$ , where  $\text{RPE} = \text{actual reward} [1 \text{ or } 0] - V$ . The Q-learning model kept separate estimates of the probabilities of receiving rewards for left and right choices ( $Q_{\text{left}}$ ,  $Q_{\text{right}}$ ) and updated  $Q$  for the chosen action (only) using  $Q' = Q + \alpha (\text{RPE})$ , where  $\text{RPE} = \text{actual reward} [1 \text{ or } 0] - Q$ . The learning parameter  $\alpha$  was determined for each session by best fit to latencies, for  $V$  or ( $Q_{\text{left}} + Q_{\text{right}}$ ) respectively. Next columns show correlations between reward expectation and dopamine cell firing after *Side-In*, measuring either peak firing rate (within 250ms after rewarded *Side-In*), minimum firing rate (middle; within 2s after unrewarded *Side-In*), and pause duration (bottom; maximum inter-spike-interval within 2s after unrewarded *Side-In*). For all histograms, light blue indicates cells with significant correlations ( $p < 0.01$ ) before multiple comparisons correction, dark blue indicates cells that remained significant after correction. Positive RPE coding is strong and consistent, negative RPE coding less so.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments.

We thank Peter Dayan, Howard Fields, Loren Frank, Chris Donaghue, and Thomas Faust for their comments on an early version of the manuscript, and Vaughn Hetrick, Rahim Hashim, Tom Davidson for technical assistance and advice. This work was supported by the National Institute on Drug Abuse, the National Institute of Mental Health, the National Institute on Neurological Disorders and Stroke, the University of Michigan, Ann Arbor, and the University of California, San Francisco.

## References.

1. Schultz W, Dayan P & Montague PR A neural substrate of prediction and reward. *Science* 275, 1593–1599 (1997). [PubMed: 9054347]
2. Pan WX, Schmidt R, Wickens JR & Hyland BI Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J Neurosci* 25, 6235–6242 (2005). [PubMed: 15987953]
3. 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]
4. Steinberg EE, Keiflin R, Boivin JR, Witten IB, et al. A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* (2013).
5. Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, et al. Mesolimbic dopamine signals the value of work. *Nat Neurosci* 19, 117–126 (2016). [PubMed: 26595651]
6. Saunders BT, Richard JM, Margolis EB & Janak PH Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nature Neuroscience* 21, 1072 (2018). [PubMed: 30038277]
7. Phillips PE, Stuber GD, Heien ML, Wightman RM & Carelli RM Subsecond dopamine release promotes cocaine seeking. *Nature* 422, 614–618 (2003). [PubMed: 12687000]
8. Roitman MF, Stuber GD, Phillips PE, Wightman RM & Carelli RM Dopamine operates as a subsecond modulator of food seeking. *J Neurosci* 24, 1265–1271 (2004). [PubMed: 14960596]
9. Wassum KM, Ostlund SB & Maidment NT Phasic mesolimbic dopamine signaling precedes and predicts performance of a self-initiated action sequence task. *Biol Psychiatry* 71, 846–854 (2012). [PubMed: 22305286]
10. Howe MW, Tierney PL, Sandberg SG, Phillips PE & Graybiel AM Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* 500, 575–579 (2013). [PubMed: 23913271]
11. Syed EC, Grima LL, Magill PJ, Bogacz R, et al. Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat Neurosci* (2015).
12. Fiorillo CD, Tobler PN & Schultz W Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299, 1898–1902 (2003). [PubMed: 12649484]
13. Morris G, Nevet A, Arkadir D, Vaadia E & Bergman H Midbrain dopamine neurons encode decisions for future action. *Nat Neurosci* 9, 1057–1063 (2006). [PubMed: 16862149]
14. Silva JAD, Tecuapetla F, Paixão V & Costa RM Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* 554, 244 (2018). [PubMed: 29420469]
15. Patriarchi T, Cho JR, Merten K, Howe MW, et al. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science* 360, eaat4422 (2018). [PubMed: 29853555]
16. Salamone JD & Correa M The mysterious motivational functions of mesolimbic dopamine. *Neuron* 76, 470–485 (2012). [PubMed: 23141060]
17. Schultz W Predictive reward signal of dopamine neurons. *J Neurophysiol* 80, 1–27 (1998). [PubMed: 9658025]

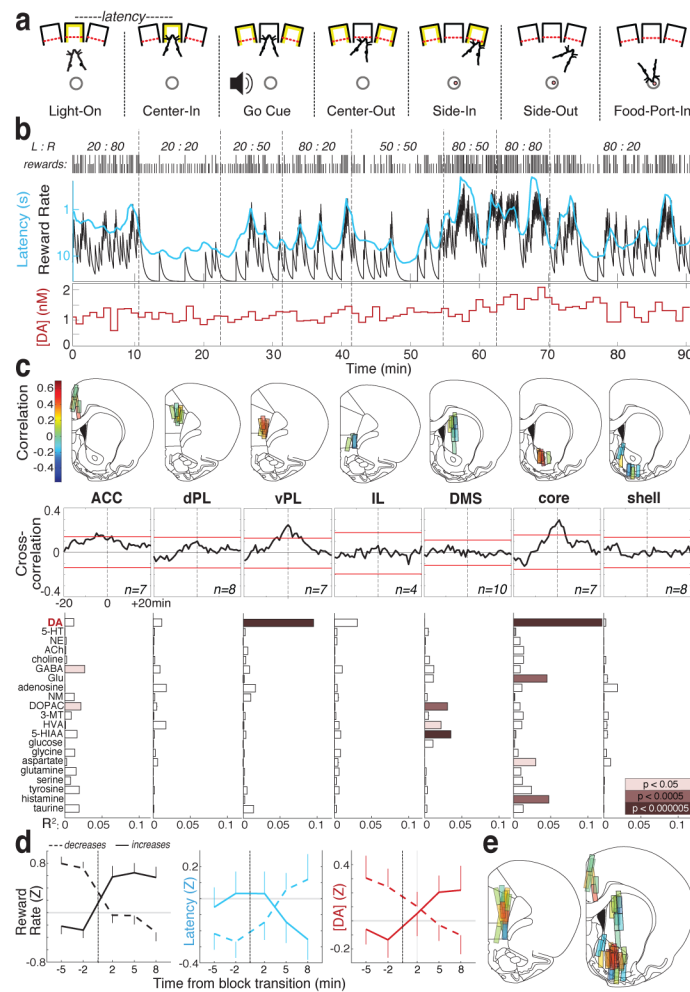
18. Garris PA & Wightman RM Different kinetics govern dopaminergic transmission in the amygdala, prefrontal cortex, and striatum: an in vivo voltammetric study. *J Neurosci* 14, 442–50. (1994). [PubMed: 8283249]
19. Frank MJ, Doll BB, Oas-Terpstra J & Moreno F Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat Neurosci* 12, 1062–1068 (2009). [PubMed: 19620978]
20. St Onge JR, Ahn S, Phillips AG & Floresco SB Dynamic fluctuations in dopamine efflux in the prefrontal cortex and nucleus accumbens during risk-based decision making. *J Neurosci* 32, 16880–16891 (2012). [PubMed: 23175840]
21. Bartra O, McGuire JT & Kable JW The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* 76, 412–427 (2013). [PubMed: 23507394]
22. Ikemoto S Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Res Rev* 56, 27–78 (2007). [PubMed: 17574681]
23. Breton JM, Charbit AR, Snyder BJ, Fong PTK, et al. Relative contributions and mapping of ventral tegmental area dopamine and GABA neurons by projection target in the rat. *J Comp Neurol* (2018).
24. Ungless MA, Magill PJ & Bolam JP Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science* 303, 2040–2042 (2004). [PubMed: 15044807]
25. Morales M & Margolis EB Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nat Rev Neurosci* 18, 73–85 (2017). [PubMed: 28053327]
26. Morris G, Arkadir D, Nevet A, Vaadia E & Bergman H Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* 43, 133–143 (2004). [PubMed: 15233923]
27. Floresco SB, West AR, Ash B, Moore H & Grace AA Afferent modulation of dopamine neuron firing differentially regulates tonic and phasic dopamine transmission. *Nat Neurosci* 6, 968–973 (2003). [PubMed: 12897785]
28. Grace AA Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nature Reviews Neuroscience* 17, 524 (2016). [PubMed: 27256556]
29. Cohen JY, Amoros MW & Uchida N Serotonergic neurons signal reward and punishment on multiple timescales. *Elife* 4, (2015).
30. Niv Y, Daw N & Dayan P How fast to work: Response vigor, motivation and tonic dopamine. *Advances in neural information processing systems* 18, 1019 (2006).
31. Bayer HM, Lau B & Glimcher PW Statistics of midbrain dopamine neuron spike trains in the awake primate. *J Neurophysiol* 98, 1428–1439 (2007). [PubMed: 17615124]
32. Chergui K, Suaud-Chagny MF & Gonon F Nonlinear relationship between impulse flow, dopamine release and dopamine elimination in the rat brain in vivo. *Neuroscience* 62, 641–65. (1994). [PubMed: 7870295]
33. Parker NF, Cameron CM, Taliaferro JP, Lee J, et al. Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat Neurosci* (2016).
34. 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, (2017).
35. Trulson ME Simultaneous recording of substantia nigra neurons and voltammetric release of dopamine in the caudate of behaving cats. *Brain Res Bull* 15, 221–223 (1985). [PubMed: 4041929]
36. Glowinski J, Chéramy A, Romo R & Barbeito L Presynaptic regulation of dopaminergic transmission in the striatum. *Cellular and Molecular Neurobiology* 8, 7–17 (1988). [PubMed: 2900072]
37. Zhou FM, Liang Y & Dani JA Endogenous nicotinic cholinergic activity regulates dopamine release in the striatum. *Nat Neurosci* 4, 1224–1229 (2001). [PubMed: 11713470]
38. Threlfell S, Lalic T, Platt NJ, Jennings KA, et al. Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron* 75, 58–64 (2012). [PubMed: 22794260]



39. Cacho R, Mateo Y, Mathur BN, Irving J, et al. Selective activation of cholinergic interneurons enhances accumbal phasic dopamine release: setting the tone for reward processing. *Cell Rep* 2, 33–41 (2012). [PubMed: 22840394]
40. Sulzer D, Cragg SJ, Rice ME Striatal dopamine neurotransmission: Regulation of release and uptake. *Basal Ganglia* 6, 123–148 (2016). [PubMed: 27141430]
41. Floresco SB, Yang CR, Phillips AG & Blaha CD Basolateral amygdala stimulation evokes glutamate receptor-dependent dopamine efflux in the nucleus accumbens of the anaesthetized rat. *Eur J Neurosci* 10, 1241–1251 (1998). [PubMed: 9749778]
42. Jones JL, Day JJ, Aragona BJ, Wheeler RA, et al. Basolateral amygdala modulates terminal dopamine release in the nucleus accumbens and conditioned responding. *Biol Psychiatry* 67, 737–744 (2010). [PubMed: 20044074]
43. Schultz W Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. *Journal of neurophysiology* 56, 1439–1461 (1986). [PubMed: 3794777]
44. Berke JD What does dopamine mean? *Nature Neuroscience* (2018).
45. Bromberg-Martin ES, Matsumoto M & Hikosaka O Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. *Neuron* 67, 144–155 (2010). [PubMed: 20624598]
46. Pasquereau B & Turner RS Dopamine neurons encode errors in predicting movement trigger occurrence. *Journal of Neurophysiology* 113, 1110–1123 (2014). [PubMed: 25411459]
47. Fiorillo CD, Newsome WT & Schultz W The temporal precision of reward prediction in dopamine neurons. *Nat Neurosci* (2008).
48. Morita K & Kato A Striatal dopamine ramping may indicate flexible reinforcement learning with forgetting in the cortico-basal ganglia circuits. *Front Neural Circuits* 8, 36 (2014). [PubMed: 24782717]
49. Gershman SJ Dopamine ramps are a consequence of reward prediction errors. *Neural Comput* 26, 467–471 (2014). [PubMed: 24320851]
50. Nicola SM The flexible approach hypothesis: unification of effort and cue-responding hypotheses for the role of nucleus accumbens dopamine in the activation of reward-seeking behavior. *J Neurosci* 30, 16585–16600 (2010). [PubMed: 21147998]

## Methods References.

51. Witten IB, Steinberg EE, Lee SY, Davidson TJ, et al. Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* 72, 721–733 (2011). [PubMed: 22153370]
52. Sugrue LP, Corrado GS & Newsome WT Matching behavior and the representation of value in the parietal cortex. *Science* 304, 1782–1787 (2004). [PubMed: 15205529]
53. Wong JM, Malec PA, Mabrouk OS, Ro J, et al. Benzoyl chloride derivatization with liquid chromatography-mass spectrometry for targeted metabolomics of neurochemicals in biological samples. *J Chromatogr A* 1446, 78–90 (2016). [PubMed: 27083258]
54. Chung JE, Magland JF, Barnett AH, Tolosa VM, et al. A Fully Automated Approach to Spike Sorting. *Neuron* 95, 1381–1394.e6 (2017). [PubMed: 28910621]
55. Kvitsiani D, Ranade S, Hangya B, Taniguchi H, et al. Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* 498, 363–366 (2013). [PubMed: 23708967]
56. Grace AA & Bunney BS The control of firing pattern in nigral dopamine neurons: burst firing. *J Neurosci* 4, 2877–2890 (1984). [PubMed: 6150071]
57. Lerner TN, Shilyansky C, Davidson TJ, Evans KE, et al. Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell* 162, 635–647 (2015). [PubMed: 26232229]
58. Paxinos G & Watson C *The rat brain in stereotaxic coordinates* (5th edition) (Elsevier Academic Press, 2005).



**Fig. 1: Dopamine release covaries with reward rate specifically in NAc core and ventral prefrontal cortex.**

**a**, Bandit task events. **b**, Example session. Top row, reward probabilities in each block (left:right); Next row, ticks indicate outcome of each trial (tall, rewarded; short, unrewarded). Next row, leaky-integrator estimate of reward rate (black) and running-average of latency (cyan; inverted log scale). Bottom, NAc core dopamine in the same session (1 min samples). **c**, Top, microdialysis locations in medial frontal cortex and striatum (see also Extended Data Fig.1).  $n=51$  probe locations, from 12 rats, each with two microdialysis probes that were lowered between sessions. Bar color indicates correlation between dopamine and reward rate. ACC, anterior cingulate cortex; dPL, dorsal prefrontal cortex; vPL, ventral prefrontal cortex; IL, infralimbic cortex; DMS, dorsal-medial striatum. Middle, averaged cross-correlograms between dopamine, reward rate. Red bars indicate 99% confidence interval from shuffled time series. Bottom, relationships between neurochemicals and reward rate (multiple regression). **d**, Effect of block transitions on reward rate (left), latency (middle) and NAc core dopamine (right). Transitions were classified by whether the experienced reward rate increased ( $n=25$ ) or decreased ( $n=33$ ). Data are from all 14 sessions in which NAc core dopamine was measured (one per rat, combining data from new and previously reported<sup>5</sup> animals), and plotted as mean  $\pm$  SEM. **e**, Composite maps of

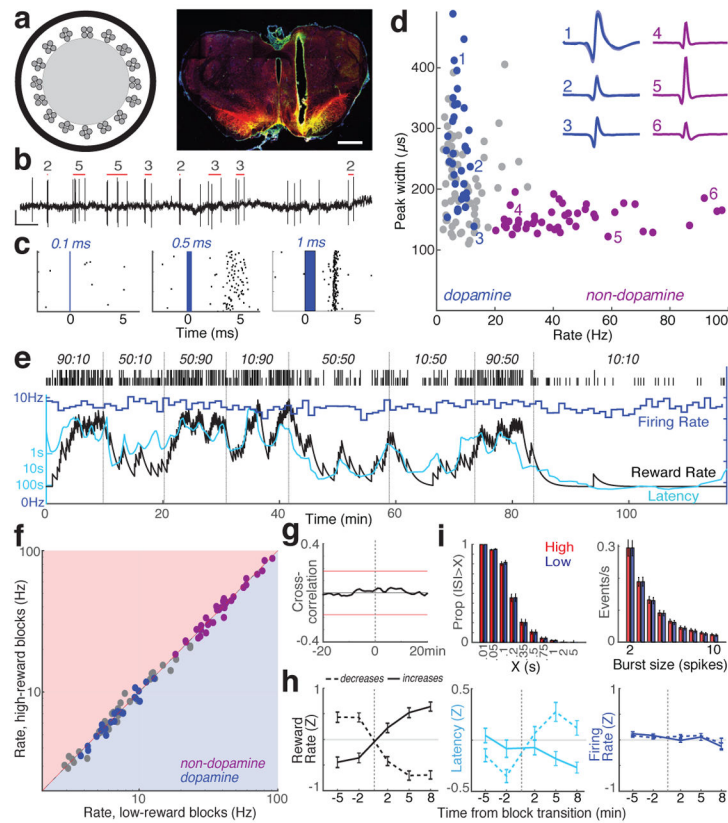
correlations between dopamine and reward rate (n=19 rats, 33 sessions, 58 probe placements).

Author Manuscript

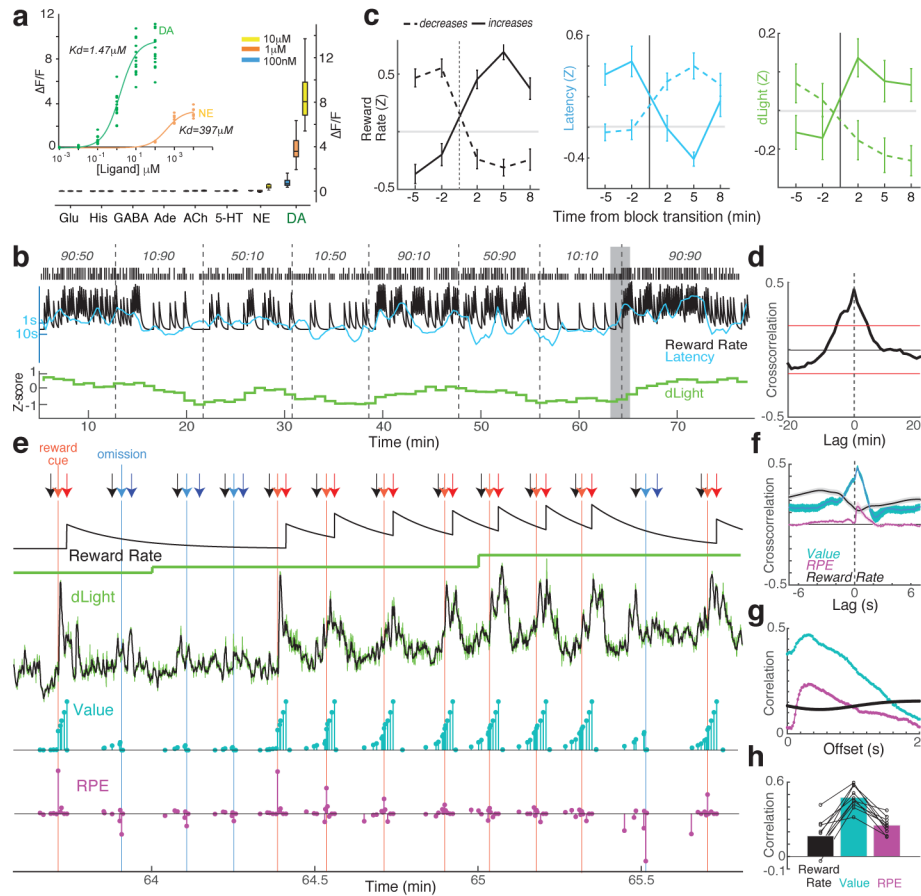
Author Manuscript

Author Manuscript

Author Manuscript

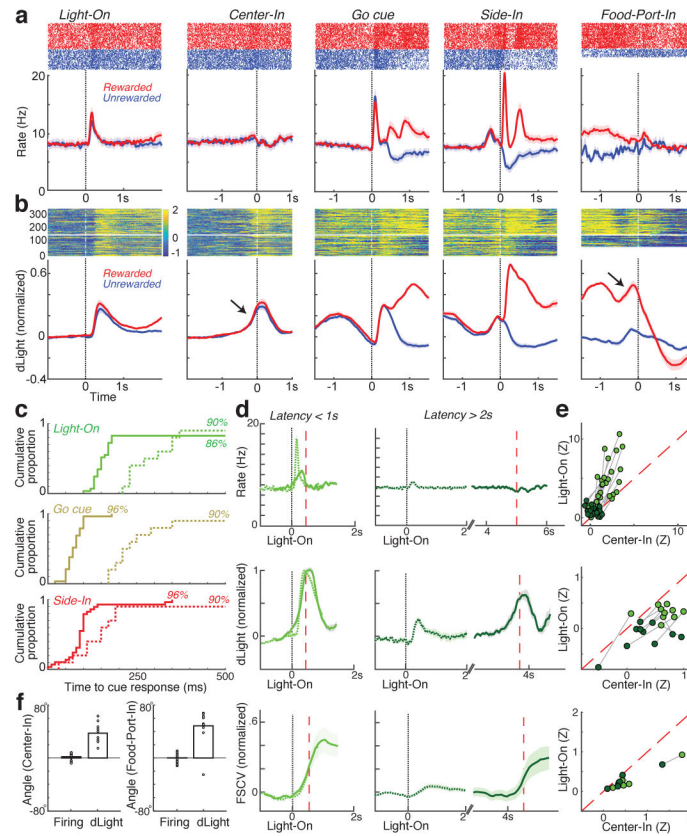


**Fig. 2: Activity of identified VTA dopamine neurons does not change with reward rate.**  
**a**, Left, optrode schematic with 16 tetrodes around 200 $\mu$ m-diameter optic fiber. Right, example of optrode placement within lateral VTA. Scale bar = 1mm. Red = dopamine cell marker tyrosine hydroxylase; green = ChR2-EYFP; yellow = overlap. For all placements, see Extended Data Fig.3. **b**, VTA dopamine cell spikes. Red bars indicate detected bursts, numbers of spikes in those bursts (see Methods). Scale, 0.5s, 0.5mV. **c**, Example neuron response to laser pulses of increasing duration. **d**, Session-wide firing rate versus spike width (at half-maximum) for each VTA cell. Blue, tagged dopamine cells; purple, a distinct cluster of presumed non-dopamine neurons. Insets, examples of average waveforms. **e**, Firing rate (blue; 1min bins) of a VTA dopamine neuron during bandit task. Latency (cyan) covaries with reward rate, but firing rate does not. **f**, Firing rate for all VTA neurons (blue, dopamine; purple, non-dopamine; grey, unclassified) in low vs. high reward rate blocks. None showed significant differences (Wilcoxon signed rank test using 1-min bins, all  $p > 0.05$  after correcting for multiple comparisons). **g**, Average cross-correlation between dopamine cell firing and reward rate shows no significant relationship. **h**, Analysis of dopamine firing rate at block transitions (same format as Fig.1d).  $n=95$  reward increases, 76 decreases. **i**, Distributions of inter-spike-intervals (ISIs, left) and spike bursts (right) are unchanged between higher- and lower reward rate blocks (Kolmogorov-Smirnov statistics: ISIs, 0.138,  $p=0.92$ , bursts, 0.165,  $p=0.63$ ).



**Fig. 3: Bridging timescales of dopamine measurement.**

**a**, Fluorescence response of dLight1.3b. Inset, titrations of dopamine (DA;  $n=15$  ROIs) and norepinephrine (NE;  $n=9$ ). Main figure, bath-applied neurotransmitters (all  $n=12$  ROIs). Glu, glutamate; His, histamine; ACh, acetylcholine. **b**, Sample bandit session including normalized NAc dLight1.3b signal (1 min bins). **c**, dLight signal changes with block transitions.  $n=35$  reward rate increases, 45 decreases. **d**, Cross-correlation between dLight and reward rate. **e**, Closer view of the shaded portion of **b**. Arrows: black, *Center-Nose-In*; light red, *Side-In* (rewarded); light blue, *Side-In* (unrewarded); dark red, *Food-Port-In* (rewarded); dark blue, *Food-Port-In* (unrewarded). Next rows: leaky-integrator estimate of reward rate, dLight at low resolution (1 min), high resolution (50Hz, green; 5-point median-filtered, black); model state values (cyan) and RPEs (magenta). After several unrewarded trials state values early in the trial are low, then reward delivery evokes a positive RPE and accompanying sharp increase in dopamine. Successive rewarded trials diminish RPEs, but increase state values, accompanied by ramping dopamine. **f**, Short timescale crosscorrelations show close relationship between dLight and value, and smaller relationship to RPE. **g**, Within-trial correlations between model variables and dLight with different lags; correlation to both value and RPE is strongest to dLight  $\sim 0.3$ s later. **h**, In all sessions maximum correlation was greater for value than RPE or reward rate.

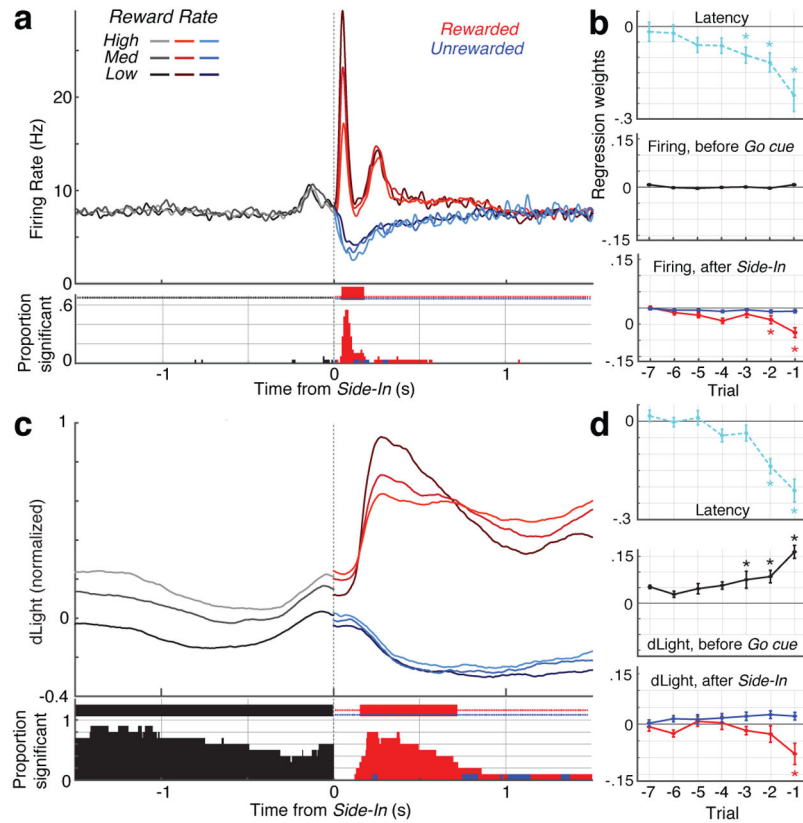


**Fig. 4: Phasic VTA dopamine firing does not account for NAc dopamine dynamics.**

**a**, Event-aligned activity of VTA-l dopamine cells. Top, spike rasters for one representative cell; bottom, average ( $n=29$ ). **b**, Event-aligned NAc dLight. Top, representative session; bottom, average ( $n=10$ ), normalized to peak rewarded *Side-In* response. Throughout this figure, dLight signals are shown relative to a 2s “baseline” epoch ending 1s before *Center-In*. Note increases (arrows) shortly before *Center-In*, *Food-Port-In*. **c**, Cumulative distributions of time for dopamine cells (solid;  $n=29$ ), dLight (dashed;  $n=10$ ), to increase following cue onsets (shuffle test compared to baseline, 10,000 shuffles,  $p<0.01$ , multiple comparisons corrected). For *Light-On*, only latencies  $<1s$  included; for *Side-In* only rewarded trials. Median latencies (from sigmoid fit): *Light-On*, firing 152ms, dLight 266ms; *Go cue*, firing 67ms, dLight 212ms; *Side-In*, firing 85ms, dLight 129ms. Non-dopamine cells were typically indifferent to cue onsets (Extended Data Fig. 8). **d**, Distinct cue-evoked, approach-related dopamine release. Top, average dopamine cell firing ( $n=29$ ); middle, bottom, average dLight ( $n=10$ ), voltammetry ( $n=6$ ), normalized to peak short-latency *Light-On* response. Left panels, latencies  $<1s$ , right, latencies  $>2s$ . Data are aligned on *Light-On* (solid) or *Center-In* (dotted); red dashed line, median latency. For longer latencies there is no increase in firing near *Center-In*, but dLight and voltammetry show a marked increase. **e**, Scatter plot comparing peak signals aligned on *Light-On* (y-axis) or *Center-In* (x-axis). For each cell, session connected lines indicate data for distinct latency ranges ( $<1s$ ,  $>2s$ ). Dopamine firing (top) consistently shows *Light-On* response for short-latency trials (2-way ANOVA, Alignment  $\times$  Latency interaction  $F=7.47$ ,  $p=0.0008$ ). dLight (middle), voltammetry (bottom) signals are consistently better aligned to *Center-In* (2-way ANOVA

for dLight: Alignment  $\times$  Latency interaction,  $F=9.28$ .  $p=0.0043$ ). **f**, Dopamine increases during approach, quantified as ramp angle (see Methods). Circles indicate individual dopamine cells ( $n=29$ ), dLight sessions ( $n=10$ ).





**Fig. 5. Reward history affects VTA dopamine cell firing and NAc dopamine release differently.** **a**, *Top*, averaged firing rates of dopamine cells ( $n=29$ ) aligned to *Side-In*, broken down by reward rate (terciles, calculated separately for each cell). Before *Side-In*, activity does not depend on reward expectation. After *Side-In* rewarded (red), unrewarded (blue) trials are shown separately. Food click response is stronger when reward rate is low, consistent with encoding of positive RPEs. *Bottom*, fraction of individual dopamine cells whose firing rate significantly varies with reward rate at each moment (shuffle test,  $p<0.01$ , multiple comparisons corrected). Tick marks at top indicate times when this fraction was significantly higher than chance (binomial,  $p<0.01$ ). After *Side-In*, only negative correlations are tested, i.e. potential RPE-coding. **b**, Regression plots for sessions with recorded dopamine cells, showing the impact of recent reward history on (log-) latency (top) and dopamine spiking. Asterisks indicate significant regression weights (t-test,  $p<0.05$ ). During the 0.5s before *Go* cue (while rat must maintain steady nosepoke for trial to proceed) dopamine spiking is unaffected by reward history (middle). This changes once the outcome is revealed (bottom; assessing peak or trough of activity in the 0.5s after *Side-In*), but only for rewarded trials. **c,d**, same as above, except for dLight (normalized to peak *Side-In* response). Dopamine release reliably scales with reward rate even before *Side-In*.