Dissociable dopamine dynamics for learning and motivation

Ali Mohebi[1,7], Jeffrey R. Pettibone[1,7], Arif A. Hamid[2], Jenny-Marie T. Wong[3], Leah T. Vinson[4], Tommaso Patriarchi[5], Lin Tian[5], Robert T. Kennedy[3] & Joshua D. Berke[1,4,6]*

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 cell spiking can encode prediction errors, which are vital learning signals in computational theories of adaptive behaviour. By contrast, dopamine release ramps up as animals approach rewards, mirroring reward expectation. This mismatch might reflect differences in behavioural 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 that indicate an upcoming reward increased both spiking and release. However, 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, whereas 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)[1–3]—learning signals for optimizing future motivated behaviour. Dopamine manipulations can affect learning as if they are altering RPEs[4–6], but they also affect motivated behaviours immediately, as if dopamine signals reward expectation (value)[5]. Furthermore, NAc dopamine escalates during motivated approach, consistent with dopamine encoding value[7–11].

With few exceptions[2,12,13], midbrain dopamine firing has been examined during classical conditioning in head-fixed animals[3,14], unlike forebrain dopamine release. We therefore compared firing with release under the same conditions. We identified VTA dopamine neurons using optogenetic tagging[3,13]. To measure NAc dopamine release, we used three independent methods—microdialysis, voltammetry and the optical sensor dLight[15]—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[5] (Fig. 1a, b). On each trial, illumination of a nose-poke port ('Light-on') prompted approach and entry ('Centre-in'). After a variable hold period (0.5–1.5 s), white noise ('Go cue') led the rat to withdraw ('Centre-out') and poke an adjacent port ('Side-in'). On rewarded trials, this Side-in event was accompanied by a food-hopper click that prompted the rat to approach a food port ('Food-port-in') to collect a sugar pellet. Leftward and 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 Centre-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[5] a correlation between NAc dopamine release and reward rate, consistent with the motivational role of mesolimbic dopamine[16]. Here, we first aimed to determine whether this relationship is observed throughout forebrain targets, consistent with 'globally broadcast' dopamine signalling[17], or is restricted to specific subregions. We further hypothesized that these dopamine dynamics would differ between striatum and cortex, as these structures have distinct dopamine uptake–degradation kinetics[18] and may use dopamine for distinct functions[19,20].

Using microdialysis with high performance liquid chromatography–mass spectrometry (HPLC–MS), 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 behavioural variables (Extended Data Fig. 2).

We replicated the correlation between reward rate and NAc dopamine—in contrast to other neurotransmitters (Fig. 1c, d). However, this relationship was localized to NAc core, and did not hold in the 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 in 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: blood oxygen level-dependent signal correlates with subjective value, specifically in NAc and ventral–medial prefrontal cortex[21].

VTA firing is unrelated to motivation

We next addressed whether this motivation-related forebrain dopamine arises from variable firing of midbrain dopamine cells. The NAc core receives dopamine input from lateral portions of VTA (VTA-l)[6,22,23]. In head-fixed mice, VTA-l dopamine neurons reportedly have uniform,
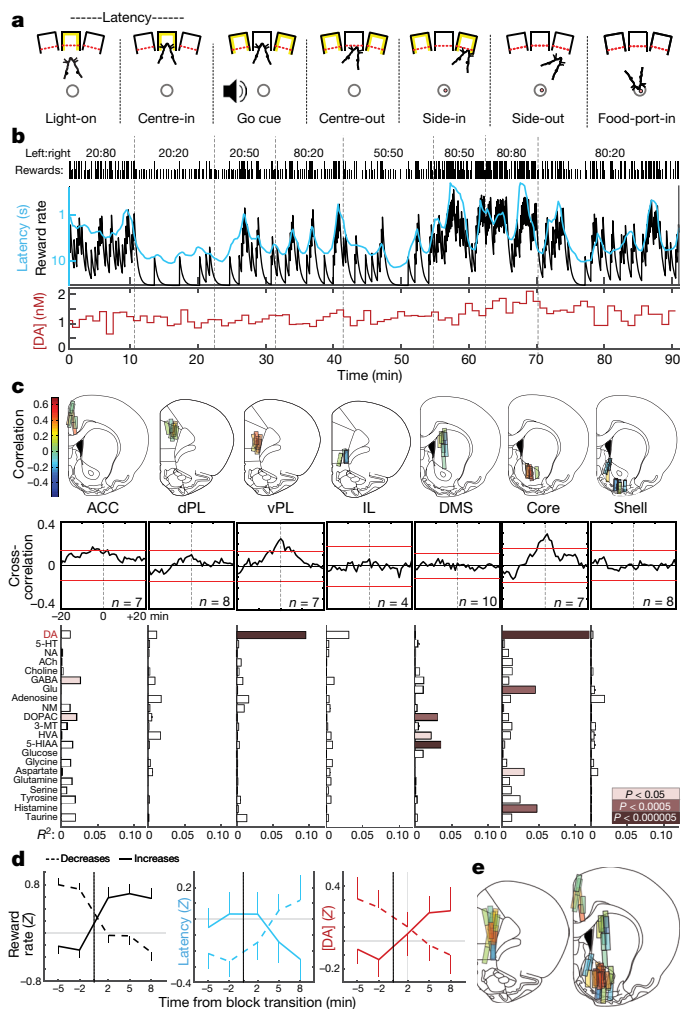
Fig. 1 | Dopamine release covaries with reward rate specifically in NAc core and ventral prelimbic cortex. a, Bandit-task events. **b**, Example session. Top row, reward probabilities in each block (left:right); row two, ticks indicate outcome of each trial (tall, rewarded; short, unrewarded); row three, leaky-integrator estimate of reward rate (black) and running average of latency (cyan; inverted log scale); bottom row, NAc core dopamine in the same session (1-min samples). DA, dopamine. **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 2 microdialysis probes that were lowered between sessions. Bar colour indicates correlation between dopamine and reward rate. ACC, anterior cingulate cortex; dPL, dorsal prelimbic cortex; vPL, ventral prelimbic cortex; IL, infralimbic cortex; DMS, dorsal-medial striatum. Middle, averaged cross-correlograms between dopamine and reward rate. Red bars indicate 99% confidence interval from shuffled time series. Bottom, relationships between neurochemicals and reward rate (multiple regression). NA, noradrenaline; 5-HT, serotonin; ACh, acetylcholine; GABA, γ-aminobutyric acid; Glu, glutamate; NM, normetadrenaline; DOPAC, 3,4-dihydroxyphenylacetic acid; 3-MT, 3-methoxytyramine; HVA, homovanillic acid; 5-HIAA, 5-hydroxyindoleacetic acid. **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[5] animals), and plotted as mean ± s.e.m. **e**, Composite maps of correlations between dopamine and reward rate ($n = 19$ rats, 33 sessions, 58 probe placements). Brain atlas outlines in this figure were reproduced with permission from Paxinos and Watson, 2005[51].

RPE-like responses to conditioned stimuli[3]. To record VTA-l dopamine cells, we infected the VTA with adeno-associated virus (AAV) for Cre-dependent expression of channelrhodopsin (AAV-DIO-ChR2) in rats that express Cre recombinase under a tyrosine hydroxylase (TH) promoter (see Methods). Optrodes (Fig. 2a, b) recorded single-unit responses to brief blue-laser pulses (Fig. 2c, Extended Data Figs. 3, 4, Supplementary Fig. 1). We found 27 well-isolated VTA-l cells with reliable short-latency spikes, and identified them as dopamine neurons.

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

We recorded the same dopamine cells across multiple behavioural 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[2,3,26],

On the basis of evidence from anaesthetized animals, it has previously been argued that altered dopamine levels measured with microdialysis arise from changes in the tonic firing rate of dopamine cells[27] and/or the proportion of active versus inactive dopamine neurons[28]. However, in the bandit task, tonic dopamine cell firing in each block of trials was indifferent to reward rate (Fig. 2e, g). There was no significant change in the firing rates of individual dopamine cells, or those of any other VTA-l neurons, between higher- and lower-reward blocks (Fig. 2f, h; see also ref. [29] 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 did not observe 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[6,22,23], but—given this apparent functional mismatch between firing and release—we reconfirmed that we were recording from the correct portion of the VTA. Small injections of the retrograde tracer cholera toxin B (CTb) into NAc core resulted in dense labelling of TH[+] neurons within the same VTA-l area as our optrode recordings (Extended Data Fig. 3). Within the approximate recording zone, 21% of TH[+] cells were also CTb[+], and this is likely to be an underestimate of the fraction of NAc core-projecting VTA-l dopamine cells, as our tracer injections did not completely fill the NAc core. Thus, 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 AAV selectively into the 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 a lack of tonic firing changes with varying reward rate (Supplementary Fig. 1). 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[30], 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 on the basis of voltammetry data[5], 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
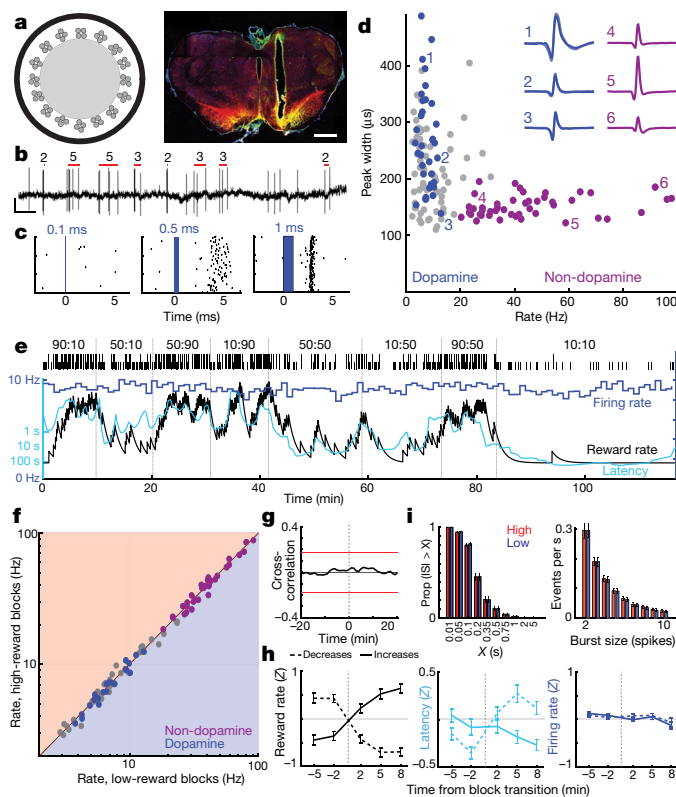
Fig. 2 | Activity of identified VTA dopamine neurons does not change with reward rate. a, Left, optrode schematic with 16 tetrodes around 200-μm-diameter optic fibre. Right, example of optrode placement within lateral VTA. Scale bar, 1 mm. Red, dopamine cell marker tyrosine hydroxylase; green, ChR2–EYFP; yellow, overlap. For all placements, see Extended Data Fig. 3. **b**, VTA dopamine cell spikes. Red bars indicate detected bursts and number of spikes in each burst (see Methods). Scale, 0.5 s, 0.5 mV. **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 (negative voltage upwards). **e**, Firing rate (blue; 1-min bins) of a VTA dopamine neuron during a 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 versus 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, $P = 0.63$).

circularly permuted GFP into dopamine D1 receptors[15]. Binding of dopamine causes a highly specific increase in fluorescence (Fig. 3a). We infused AAV into NAc to express either dLight1.1 (four verified NAc placements from three rats) or the brighter variant dLight1.3b (six verified NAc placements from four rats) and monitored fluorescence by fibre 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 next examined more closely 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[5]). As was previously reported using voltammetry[5], moment-by-moment NAc dopamine showed a strong correlation with state values (Fig. 3f), visible as ramping up within trials when rewards were expected. We also saw transient increases with less-expected reward deliveries, consistent with RPE (examined below). In every dLight session, dopamine showed a stronger correlation with values than either RPEs or reward rate (Fig. 3h, Extended Data Fig. 6). Correlations with both state values and RPE were maximal with respect to the dLight signal ~0.3 s later, consistent with a brief lag caused by neural processing of cues and sensor-response time (Fig. 3g; with voltammetry, we reported a lag of 0.4–0.5 s)[5].

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 relative magnitude of responses to different cues varied from cell to cell (Supplementary Fig. 1). 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 Centre-in) and the food port (just before Food-port-in). This fits well with the extensive voltammetry literature showing that motivated approach behaviours are accompanied by rapid increases in NAc core dopamine[5,7–11]. 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 to non-dopamine cells).

To better dissociate cue-evoked, and approach-related, dopamine activity, we separated trials by short (<1 s) and long (>2 s) 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 Centre-in (Fig. 4e). By contrast, increases in NAc dopamine release before Centre-in were distinct from cue-evoked dopamine release (Fig. 4d, e). dLight signals consistently increased before Centre-in on long-latency trials (ten out of ten sessions) and before food-port-in (nine out of ten sessions), without corresponding increases in dopamine firing (Fig. 4f).

Finally we considered how event-related dopamine signals depend on 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 at Side-in was reliably stronger when the 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 as individual cells (Extended Data Fig. 8). It has previously been proposed that negative RPEs are encoded in the duration of dopamine pauses[31], but this was observed in just 2 out of 29 individual neurons. Similar results were obtained if reward expectation was estimated in other ways, including trial-based reinforcement learning models (actor-critic and Q-learning) or simply by counting recent rewards (Extended Data Fig. 8).

Dopamine release at Side-in also showed a clear, transient encoding of positive RPEs, but not of negative RPEs (Fig. 5c, d). This dLight response was slightly delayed and prolonged compared to firing, consistent with time taken for release and reuptake[32], but remained a sub-second 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 when the rat was not actively moving, but was maintaining
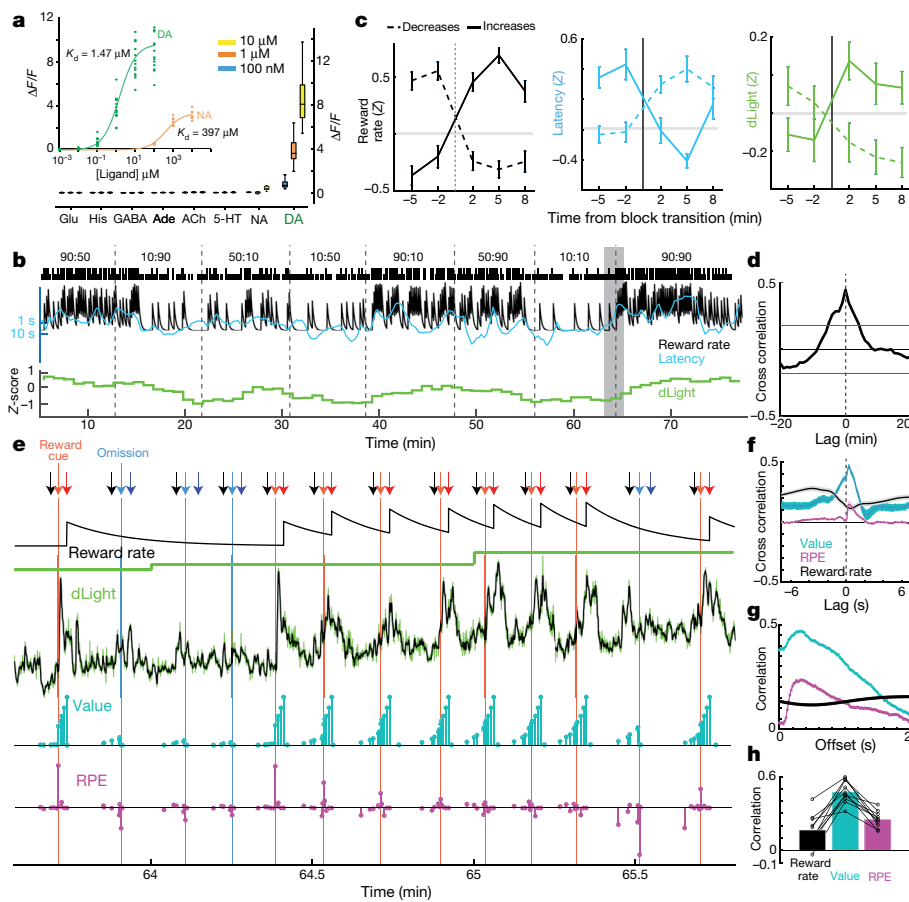
Fig. 3 | Bridging timescales of dopamine measurement. a, Fluorescence response of dLight1.3b. Inset, titrations of dopamine ($n = 15$ regions of interest (ROIs)) and noradrenaline ($n = 9$). Main figure, bath-applied neurotransmitters (all $n = 12$ ROIs). His, histamine. **b**, Sample bandit session 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, Centre-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); dLight at high resolution (50 Hz, green; five-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 about 0.3 s later. **h**, In all sessions, maximum correlation was greater for value than for RPE or reward rate.

a nose poke in the centre port while 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 the NAc core[6,23,24]. VTA-l dopamine cells, including those that project to the NAc core, consistently display RPE-encoding bursts[3,12]. VTA bursts are thought to be particularly important for driving NAc dopamine[32], 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[13,33,34], 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 and for different functions, and that recording dopamine cells is important but not sufficient for understanding dopamine signals[35].

Release from dopamine terminals is potently influenced by local, non-spiking mechanisms[36–40]. For example, NAc dopamine release is modulated by the basolateral amygdala even when VTA spiking is pharmacologically suppressed[41,42]. It has been noted for decades that local control of dopamine release might achieve functions distinct from those of dopamine cell spiking[36,43], but this has not been incorporated into theoretical views of dopamine. Distinct striatal subregions contribute to different types of decisions, and may influence their own dopamine release according to need[44]. It remains to be determined just how localized this control of dopamine release can be. One limitation

shared by the 3 ways that we measured dopamine release is that they all sample on a spatial scale of at least 100 μm, whereas in vivo microscopy suggests that dopamine release may be heterogeneous at considerably smaller scales[15].

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[30] provide strong evidence against the idea[29] that changes in tonic dopamine cell firing drive tonic changes in dopamine release. Although tonic firing can be altered by lesions or drug manipulations[28], we are not aware of sustained changes in firing rate in any behavioural task. Firing can ramp downwards on a timescale of about 1 s during anticipation of motivationally relevant events[45,46]. However, this decline is the opposite of what would be required to boost dopamine release with reward expectation, and instead bears more resemblance to a sequence of transient negative prediction errors[47]. Although sustained signals encoding ongoing reward rate could be computationally useful[30], dopamine instead provides rapidly fluctuating error and value signals. It remains possible that sustained signals are computed at a subsequent step, by intracellular signalling pathways downstream of dopamine receptors.

Many groups have observed ramping dopamine release as rats approach rewards[5,7–11], consistent with encoding escalating reward expectations. Some have argued that these dopamine ramps simply reflect RPEs, by supposing that rats either rapidly forget values[48] or that they have a warped set of state representations[49]. This latter idea is not supported by our observation that ramping is rapidly modulated from trial to trial on the basis of 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
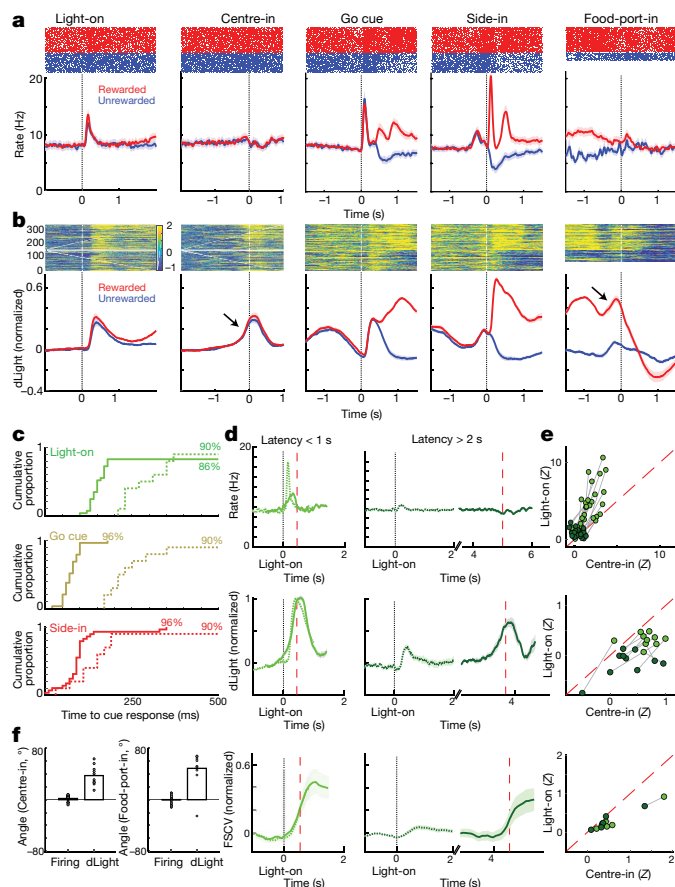
Fig. 4 | Phasic VTA dopamine firing does not account for NAc dopamine dynamics. a, Event-aligned activity of VTA-l dopamine cells. Top, spike rasters for one representative cell; bottom, average spike rate ($n = 29$). In all panels, error bands indicate ± s.e.m. **b**, Event-aligned NAc dLight. Top, representative session; bottom, average ($n = 10$), normalized to peak rewarded Side-in response. Throughout this figure, dLight signals are shown relative to a 2-s 'baseline' epoch ending 1 s before Centre-in. Note increases (arrows) shortly before Centre-in and Food-port-in. **c**, Cumulative distributions of time taken 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 <1 s included; for Side-in only rewarded trials. Median latencies (from sigmoid fit): Light-on, firing 152 ms, dLight 266 ms; Go cue, firing 67 ms, dLight 212 ms; Side-in, firing 85 ms, dLight 129 ms. 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, average dLight ($n = 10$); bottom, voltammetry ($n = 6$), normalized to peak short-latency Light-on response. Left panels, latencies <1 s, right, latencies >2 s. Data are aligned on Light-on (solid) or Centre-in (dotted); red dashed line, median latency. For longer latencies there is no increase in firing near Centre-in, but dLight and voltammetry show a marked increase. **e**, Scatter plot comparing peak signals aligned on Light-on (*y* axis) or Centre-in (*x* axis). For each cell, session connected lines indicate data for distinct latency ranges (<1 s, >2 s). Dopamine firing (top) consistently shows Light-on response for short-latency trials (two-way analysis of variance (ANOVA), alignment × latency interaction, $F = 7.47$, $P = 0.0008$). dLight (middle), voltammetry (bottom) signals are consistently better aligned to Centre-in (two-way ANOVA for dLight: alignment × 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 ($n = 10$).

account for the very well-established connection between ongoing mesolimbic dopamine and motivation[16]. 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[50]. NAc core dopamine appears to provide an essential dynamic
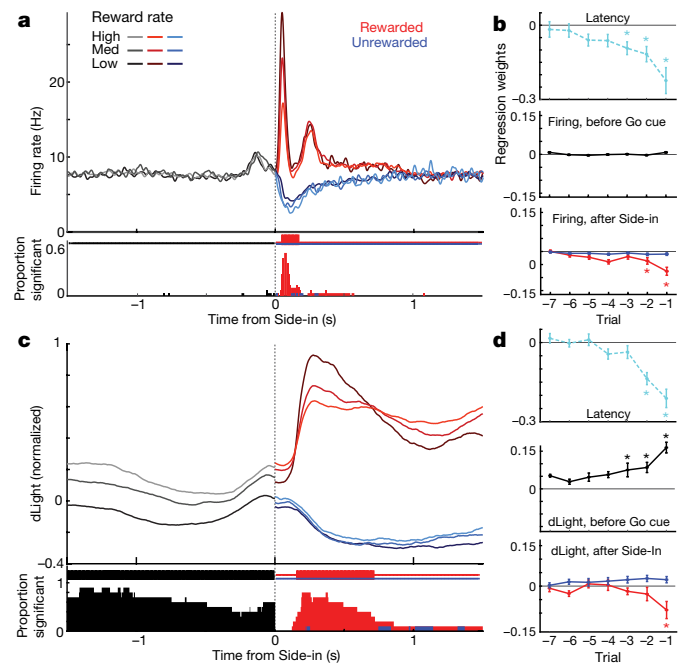

Fig. 5 | Reward history affects VTA dopamine cell firing and NAc dopamine release differently. a, Top, averaged firing rates of dopamine cells ($n = 29$) aligned to Side-in, broken down by reward rate (terciles, calculated separately for each cell). Before Side-in, activity does not depend on reward expectation. After Side-in rewarded (red) and 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 with a firing rate that 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—that is, potential RPE coding. **b**, Regression plots for sessions with recorded dopamine cells, showing the effect of recent reward history on (log-) latency (top) and dopamine spiking. Asterisks indicate significant regression weights (*t*-test, $P < 0.05$). During the 0.5 s before Go cue (while rat must maintain a steady nose poke 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 0.5 s 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.

signal of how worthwhile it is to allocate time and effort to work[5,44], even though this signal is not present in VTA dopamine cell firing.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1235-y.

1. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
2. Pan, W. X., Schmidt, R., Wickens, J. R. & Hyland, B. I. Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J. Neurosci.* **25**, 6235–6242 (2005).
3. Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B. & Uchida, N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
4. Steinberg, E. E. et al. A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* **16**, 966–973 (2013).
5. Hamid, A. A. et al. Mesolimbic dopamine signals the value of work. *Nat. Neurosci.* **19**, 117–126 (2016).
6. Saunders, B. T., Richard, J. M., Margolis, E. B. & Janak, P. H. Dopamine neurons create Pavlovian conditioned stimuli with circuit-defined motivational properties. *Nat. Neurosci.* **21**, 1072–1083 (2018).

7. Phillips, P. E., Stuber, G. D., Heien, M. L., Wightman, R. M. & Carelli, R. M. Subsecond dopamine release promotes cocaine seeking. *Nature* **422**, 614–618 (2003).

8. Roitman, M. F., Stuber, G. D., Phillips, P. E., Wightman, R. M. & Carelli, R. M. Dopamine operates as a subsecond modulator of food seeking. *J. Neurosci.* **24**, 1265–1271 (2004).

9. Wassum, K. M., Ostlund, S. B. & Maidment, N. T. Phasic mesolimbic dopamine signaling precedes and predicts performance of a self-initiated action sequence task. *Biol. Psychiatry* **71**, 846–854 (2012).

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

11. Syed, E. C. et al. Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat. Neurosci.* **19**, 34–36 (2016).

12. Morris, G., Nevet, A., Arkadir, D., Vaadia, E. & Bergman, H. Midbrain dopamine neurons encode decisions for future action. *Nat. Neurosci.* **9**, 1057–1063 (2006).

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

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

15. Patriarchi, T., Cho, J. R., Merten, K., Howe, M. W., et al. Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. *Science* **360**, eaat4422 (2018).

16. Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**, 470–485 (2012).

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

18. Garris, P. A. & Wightman, R. M. Different kinetics govern dopaminergic transmission in the amygdala, prefrontal cortex, and striatum: an in vivo voltammetric study. *J. Neurosci.* **14**, 442–450 (1994).

19. Frank, M. J., Doll, B. B., Oas-Terpstra, J. & Moreno, F. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat. Neurosci.* **12**, 1062–1068 (2009).

20. St Onge, J. R., Ahn, S., Phillips, A. G. & Floresco, S. B. Dynamic fluctuations in dopamine efflux in the prefrontal cortex and nucleus accumbens during risk-based decision making. *J. Neurosci.* **32**, 16880–16891 (2012).

21. Bartra, O., McGuire, J. T. & Kable, J. W. The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *Neuroimage* **76**, 412–427 (2013).

22. Ikemoto, S. Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. *Brain Res. Brain Res. Rev.* **56**, 27–78 (2007).

23. Breton, J. M. 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, M. A., Magill, P. J. & Bolam, J. P. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science* **303**, 2040–2042 (2004).

25. Morales, M. & Margolis, E. B. Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nat. Rev. Neurosci.* **18**, 73–85 (2017).

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).

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

28. Grace, A. A. Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nat. Rev. Neurosci.* **17**, 524–532 (2016).

29. Cohen, J. Y., Amoroso, M. W. & Uchida, N. Serotonergic neurons signal reward and punishment on multiple timescales. *eLife* **4**, e06346 (2015).

30. Niv, Y., Daw, N. & Dayan, P. How fast to work: response vigor, motivation and tonic dopamine. *Adv. Neural Inf. Process. Syst.* **18**, 1019 (2006).

31. Bayer, H. M., Lau, B. & Glimcher, P. W. Statistics of midbrain dopamine neuron spike trains in the awake primate. *J. Neurophysiol.* **98**, 1428–1439 (2007).

32. Chergui, K., Suaud-Chagny, M. F. & Gonon, F. Nonlinear relationship between impulse flow, dopamine release and dopamine elimination in the rat brain in vivo. *Neuroscience* **62**, 641–645 (1994).

33. Parker, N. F. et al. Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat. Neurosci.* **19**, 845–854 (2016).

34. Menegas, W., Babayan, B. M., Uchida, N. & Watabe-Uchida, M. Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. *eLife* **6**, e21886 (2017).

35. Trulson, M. E. Simultaneous recording of substantia nigra neurons and voltammetric release of dopamine in the caudate of behaving cats. *Brain Res. Bull.* **15**, 221–223 (1985).

36. Glowinski, J., Chéramy, A., Romo, R. & Barbeito, L. Presynaptic regulation of dopaminergic transmission in the striatum. *Cell. Mol. Neurobiol.* **8**, 7–17 (1988).

37. Zhou, F. M., Liang, Y. & Dani, J. A. Endogenous nicotinic cholinergic activity regulates dopamine release in the striatum. *Nat. Neurosci.* **4**, 1224–1229 (2001).

38. Threlfell, S. et al. Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron* **75**, 58–64 (2012).

39. Cachope, R. et al. Selective activation of cholinergic interneurons enhances accumbal phasic dopamine release: setting the tone for reward processing. *Cell Reports* **2**, 33–41 (2012).

40. Sulzer, D., Cragg, S. J. & Rice, M. E. Striatal dopamine neurotransmission: regulation of release and uptake. *Basal Ganglia* **6**, 123–148 (2016).

41. Floresco, S. B., Yang, C. R., Phillips, A. G. & Blaha, C. D. Basolateral amygdala stimulation evokes glutamate receptor-dependent dopamine efflux in the nucleus accumbens of the anaesthetized rat. *Eur. J. Neurosci.* **10**, 1241–1251 (1998).

42. Jones, J. L. et al. Basolateral amygdala modulates terminal dopamine release in the nucleus accumbens and conditioned responding. *Biol. Psychiatry* **67**, 737–744 (2010).

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

44. Berke, J. D. What does dopamine mean? *Nat. Neurosci.* **21**, 787–793 (2018).

45. Bromberg-Martin, E. S., Matsumoto, M. & Hikosaka, O. Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. *Neuron* **67**, 144–155 (2010).

46. Pasquereau, B. & Turner, R. S. Dopamine neurons encode errors in predicting movement trigger occurrence. *J. Neurophysiol.* **113**, 1110–1123 (2015).

47. Fiorillo, C. D., Newsome, W. T. & Schultz, W. The temporal precision of reward prediction in dopamine neurons. *Nat. Neurosci.* **11**, 966–973 (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).

49. Gershman, S. J. Dopamine ramps are a consequence of reward prediction errors. *Neural Comput.* **26**, 467–471 (2014).

50. Nicola, S. M. 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).

51. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* 5th edn (Elsevier Academic, 2005).

Author contributions A.M. performed and analysed the electrophysiology and photometry, and applied the computational model. J.R.P. performed and analysed the microdialysis with assistance from J.-M.T.W. and supervision by R.T.K. A.A.H. developed the behavioural task and initial photometry setup, and performed the voltammetry. L.T.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.

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 (300–500 g, either wild-type Long-Evans or TH-Cre[+] with a Long-Evans background[52]) 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. The investigators were not blinded to allocation during experiments and outcome assessment.

Behaviour. 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[5]. Bandit-task sessions used the following parameters: block lengths were 35-45 trials, randomly selected for each block; hold period before Go cue was 500–1,500 ms (uniform distribution); left–right reward probabilities were 10, 50 and 90% (for electrophysiology, photometry, voltammetry and previously reported microdialysis rats[5]) or 20, 50 and 80% (newly reported microdialysis rats).

Current reward rate was estimated using a time-based leaky-integrator[53]. Reward rate was incremented each time a reward was received, and decayed exponentially at a rate set by parameter τ (the time in s for the reward rate to decrease by ~63%, that is, $1-1/e$). For all analyses, τ was selected on the basis of the rat's behaviour, 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 τ (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 (2 kHz, 5 kHz and 9 kHz) were associated with different probabilities of food delivery (counterbalanced across rats). Cues were played as a train of tone pips (100 ms on, 50 ms off) for a total duration of 2.6 s followed by a delay period of 500 ms. Cues and unpredicted reward deliveries were delivered in pseudorandom order with a variable inter-trial interval (15–30 s, 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 (anteroposterior (AP) +3.2 mm, mediolateral (ML) 0.6 mm relative to bregma; and dorsoventral (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 and 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 and DV 0.8) and another targeting accumbens (core/shell in the opposite hemisphere at AP +1.6, ML 1.4 and DV 5.5 ($n = 2$) or AP +1.6, ML 1.9 and DV 5.7 ($n = 2$). Implant sides were counterbalanced across rats. Animals were allowed to recover for one week before retraining.

Chemicals. Water, methanol, and acetonitrile for mobile phases were Burdick & Jackson HPLC grade, purchased from VWR (Radnor). All other chemicals were purchased from Sigma Aldrich unless otherwise noted. Artificial cerebrospinal fluid (aCSF) comprised 145 mM NaCl, 2.68 mM KCl, 1.40 mM CaCl$_2$, 1.01 mM MgSO$_4$, 1.55 mM Na$_2$HPO$_4$ and 0.45 mM NaH$_2$PO$_4$, 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 dialysing AN69 membrane; Hospal) were inserted bilaterally into guide cannula and perfused continuously (Chemyx, 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[54] with 1.5 µl sodium carbonate, 100 mM; 1.5 µl benzoyl chloride (2% (v/v) benzoyl chloride in acetonitrile); and 1.5 µl isotopically labelled 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) to a final concentration of 20 nM. Sample series collection alternated between the two probes at 30-s 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 analysed 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-microlitre samples were injected onto a Phenomenex core-shell biphenyl Kinetex HPLC column (2.1 mm × 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 µl/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 three-point moving average ($y' = [0.25 \times (y{-}1) + 0.5y + 0.25 \times (y{+}1)]$) 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. One-per cent 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 behavioural metrics as predictors. Regression coefficients were determined significant at three alpha levels (0.05, 0.0005 and 0.000005), after Bonferroni-correction for multiple comparisons (alpha/(21 chemicals × 7 regions × 9 behavioural 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-µm nichrome wire, Sandvik) glued onto the side of a 200-µm optic fibre and extending up to 500 µm below the fibre tip. During the same surgery, we injected 1 µl 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 (1–9,000 Hz) brain signals were sampled (30,000 samples per s) using Intan digital headstages. Optrodes were lowered at least 80 µm at the end of each recording session. Individual units were isolated offline using a MATLAB implementation of MountainSort[55] followed by careful manual inspection.

Classification. To identify whether an isolated VTA-l unit was dopaminergic (TH[+]), we used the stimulus-associated latency test[56]. In brief, 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 5-ms and 10-ms pulse trains. We also compared the light evoked waveforms (within 10 ms 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-l AAV infusions (IM657, 1 unit; IM1002, 3 units; IM1003, 15 units; IM1037, 9 units) and one rat with NAc core AAV (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 >20 Hz and peak width <200 µs were classified as non-dopamine cells. To ensure that we were comparing dopamine and non-dopamine cells within the same subregions, we only analysed 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[57]: 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 20 ms. To determine how quickly a neuron responded to a given cue, we used 40-ms bins (sliding in steps of 20 ms) 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 behaviours, we smoothed mean firing rates with a 50-ms Gaussian kernel, detected the maximum/minimum of the resulting signal in a 0.5-s window before each event (centre-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 reanalyse data previously presented in detail[5]. 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[5].

Photometry. We used a viral approach to express the genetically encoded optical dopamine sensor dLight[15]. Under isoflurane anaesthesia, 1 µl of AAV9-CAG-dLight (1×10^{12} viral genomes per ml; UC Davis vector core) was slowly (100 nl/min) injected (Nanoject III, Drummond) through a 30-µm glass micropipette in ventral

striatum bilaterally (AP: 1.7 mm, ML: 1.7 mm, DV: −7.0 mm). During the same surgery optical fibres (400-μm core, 430-μm total diameter) attached to a metal ferrule (Doric) were inserted (target depth 200 μm higher than AAV) and cemented in place. Data were collected > three weeks later, to allow for dLight expression.

For dLight excitation blue (470 nm) and violet (405 nm; control) LEDs were sinusoidally modulated at distinct frequencies (211 Hz and 531 Hz, respectively[58]). Both excitation and emission signals passed through minicube filters (Doric) and bulk fluorescence was measured with a femtowatt detector (Newport, Model 2151) sampling at 10 kHz. Demodulation produced separate 470 nm (dopamine) and 405 nm (control) signals, which were then rescaled to each other via a least-square fit[58]. Fractional fluorescence signal (dF/F) was then defined as $(470–405\_fit)/405\_fit$. For all analyses this signal was downsampled to 50 Hz and smoothed with a five-point median filter. For presentation of 470 nm and 405 nm signals separately, see Extended Data Fig. 7.

Data from an optic fibre placement were included in analyses if the fibre tip was in NAc, and the fluorescence response 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 behaviours, we detected the maximum/minimum of the resulting signal in a 0.5-s window before each event (centre-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[15]. In brief, 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 40× oil-based objective on an inverted Zeiss Observer LSN710 confocal microscope with 488 nm/513 nm (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 noradrenaline were obtained by performing tenfold serial dilutions to achieve eight different concentrations. All other neurotransmitters were tested at three sequential concentrations (100 nM, 1 μM and 10 μM). 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 ROI was manually drawn on the membrane of individual cells. Fluorescent fold change ($\Delta F/F$) was calculated as F peak (averaged fluorescence intensity of four frames) − F basal (averaged fluorescence intensity of four frames before addition of ligands)/F basal. Graphs and statistical analysis were performed using GraphPad Prism 6. Data points were analysed 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. **Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The AAV.Synapsin.dLight1.3b virus used in this study has been deposited with Addgene (no. 125560; http://www.addgene.org). All data will be available through the Collaborative Research in Computational Neuroscience data sharing website (https://doi.org/110.6080/K0VQ30V9).

Code availability

Custom MATLAB code is available on request from J.D.B.

52. Witten, I. B. et al. Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* **72**, 721–733 (2011).
53. Sugrue, L. P., Corrado, G. S. & Newsome, W. T. Matching behavior and the representation of value in the parietal cortex. *Science* **304**, 1782–1787 (2004).
54. Wong, J. M. 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).
55. Chung, J. E. et al. A fully automated approach to spike sorting. *Neuron* **95**, 1381–1394 (2017).
56. Kvitsiani, D. et al. Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature* **498**, 363–366 (2013).
57. Grace, A. A. & Bunney, B. S. The control of firing pattern in nigral dopamine neurons: burst firing. *J. Neurosci.* **4**, 2877–2890 (1984).
58. Lerner, T. N. et al. Intact-brain analyses reveal distinct information carried by SNc dopamine subcircuits. *Cell* **162**, 635–647 (2015).
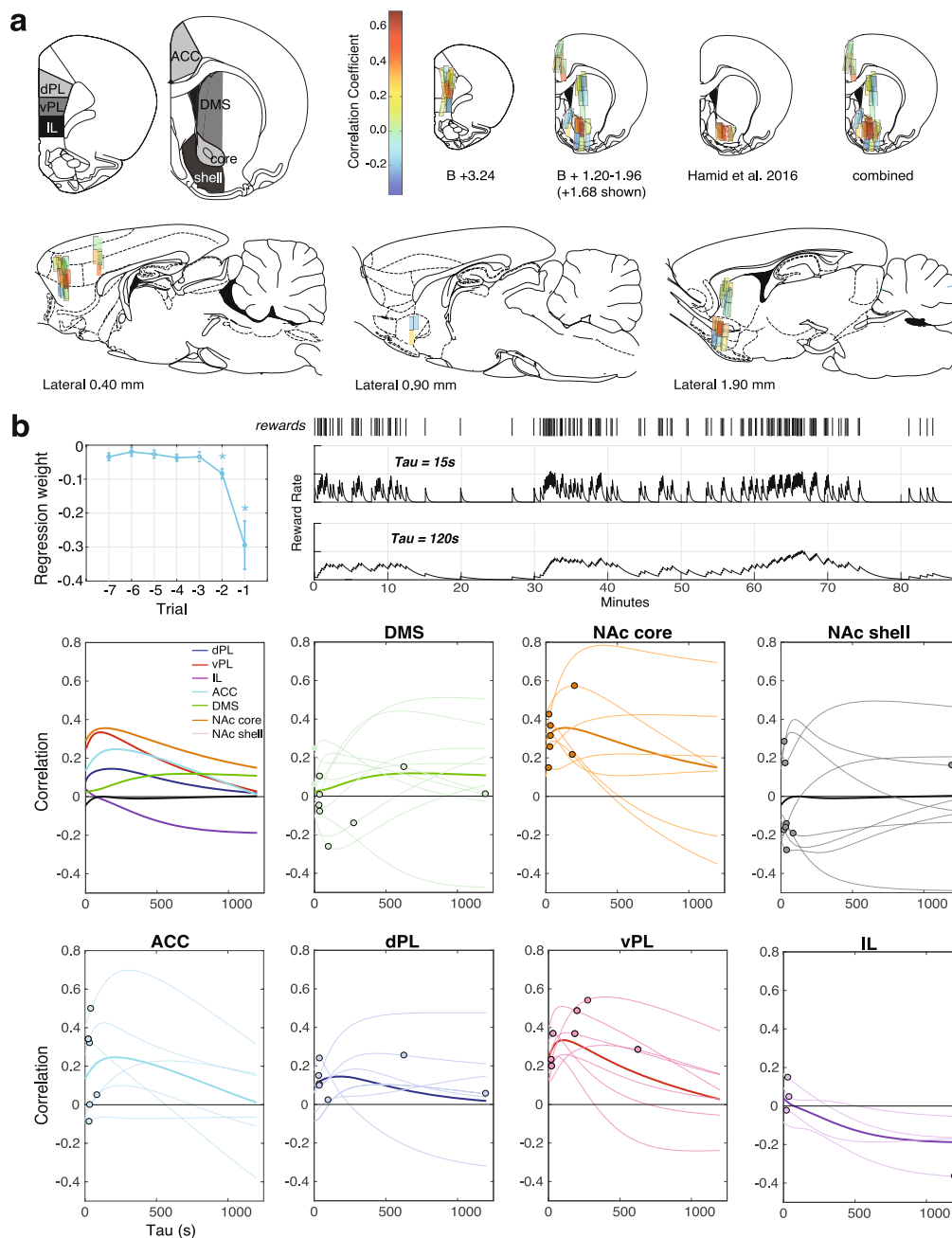
Extended Data Fig. 1 | Microdialysis subregions and the reward rate parameter. a, Top left, anatomical definitions of the subregions examined with microdialysis. Brain atlas outlines in this figure were reproduced with permission from Paxinos and Watson, 2005[51]. Other panels map the correlation between dopamine release and reward rate at individual probe placements in coronal (mm from bregma, B) and sagittal (mm from midline) planes. Colour 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, 7,113 trials, from 12 rats; error bars show s.e.m.). *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 constant (tau) of

the leaky integrator. Top middle, dopamine: reward rate correlations as a function of τ. In the main Figs., τ was chosen (from a range of 1–1,200 s) to maximize the (negative) correlation between reward rate and (log) latency in each session. Thin lines represent individual sessions, with the best fit τ 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 behavioural metrics were similar between sessions sampling from each of the seven subregions (mean rewards per min: range 1.42–1.77, ANOVA $F_{(6,44)} = 0.58$, $P = 0.746$; mean attempts per min: range 3.32–3.97, $F_{(6,44)} = 0.40$, $P = 0.872$; mean latency: range 5.99–8.02, $F_{(6,44)} = 0.27$, $P = 0.948$).
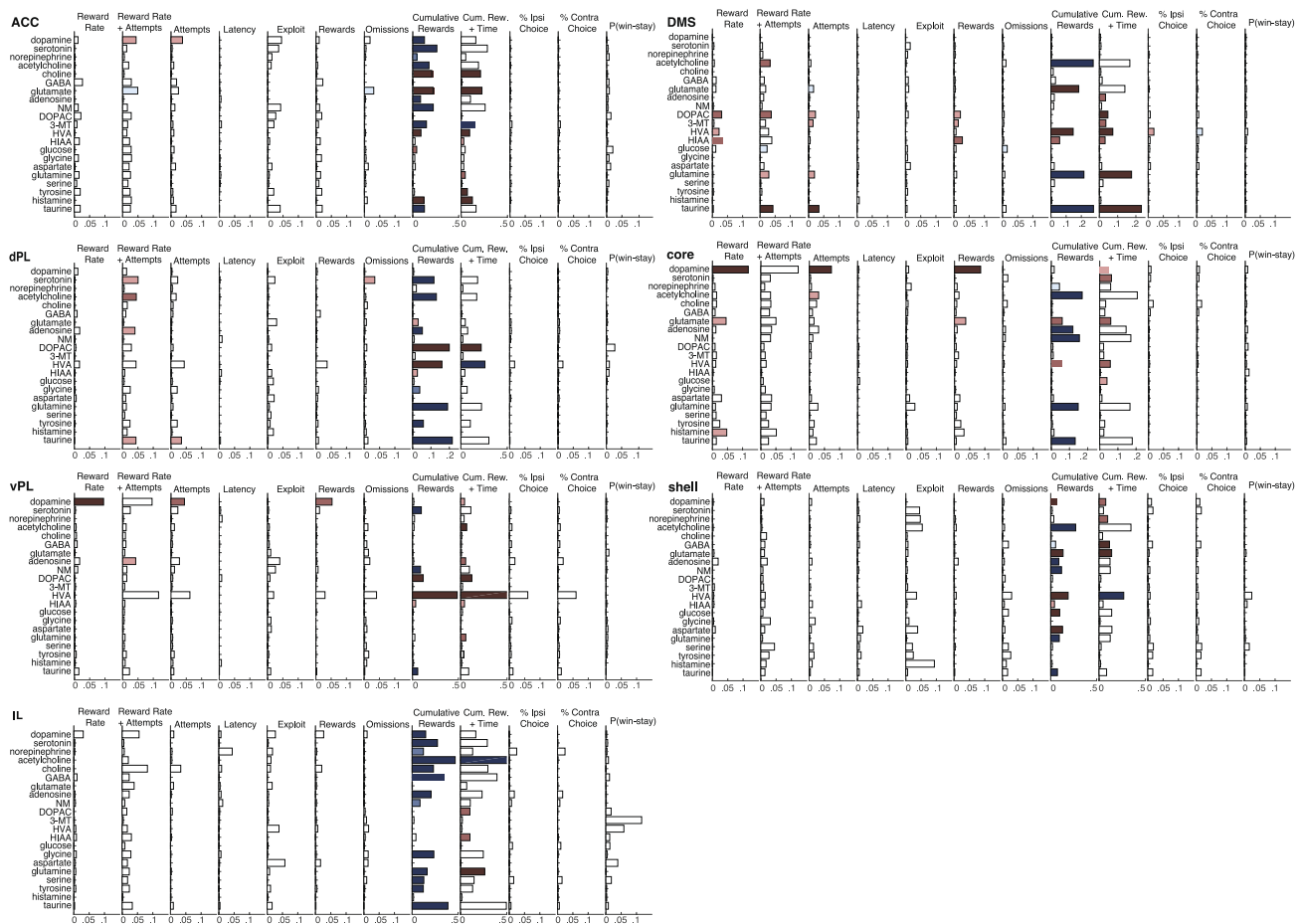
Extended Data Fig. 2 | Correlations between all neurochemicals and a range of behavioural factors. Bars represent R^2 values for linear tests between each analyte (rows) and behavioural covariates (columns). In models with more than one covariate, bar length indicates the R^2 for the full model. Negative relationships are reported in blue and positive relationships are in red. P values are reported at three alpha levels (0.05, 0.0005 and 0.000005) after Bonferroni correction for multiple comparisons (7 subregions × 21 analytes × 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 minute, 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 minutes elapsed since the session began. Bars in this column show colour 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.
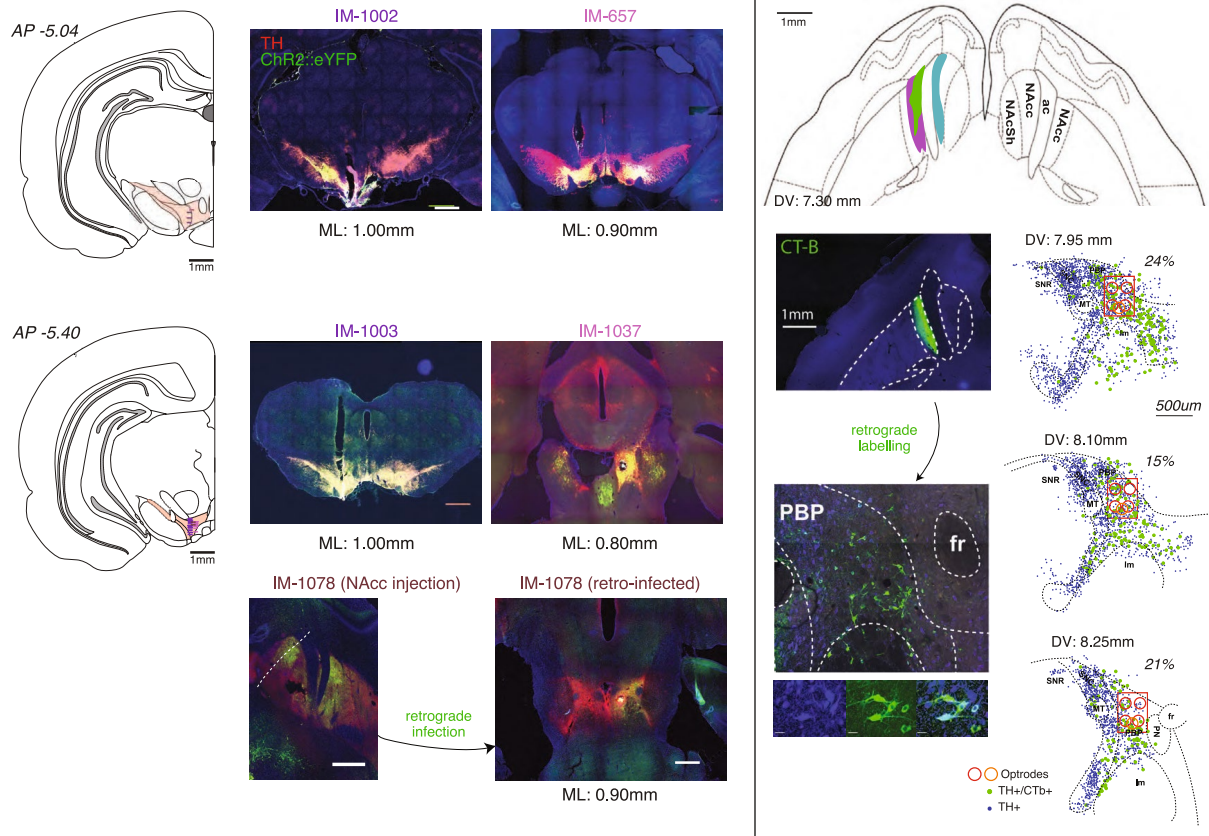
Extended Data Fig. 3 | Histological analysis of electrophysiological recording locations. Left, atlas locations and histology photomicrographs for each rat (IM-657, IM-1002, IM-1003, IM-1037 and IM-1078) from which opto-tagged dopamine cells were obtained. Red, TH-staining; green: ChR2–eYFP; blue: DAPI. Scale bars, 1 mm. IM-1037 and IM-1078 brains were sliced horizontally, so fibre tracks appear as a circle. Font colours for rat ID numbers correspond to colours 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 labelling in each of the three rats (each rat indicated by a different colour). Bottom left panels show close-ups of TH labelling (blue), CTb (green) and merged image. Bottom right panels show reconstructed locations of TH$^+$ and double-labelled 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$^+$. Brain atlas outlines in this figure were reproduced with permission from Paxinos and Watson, 2005[51].
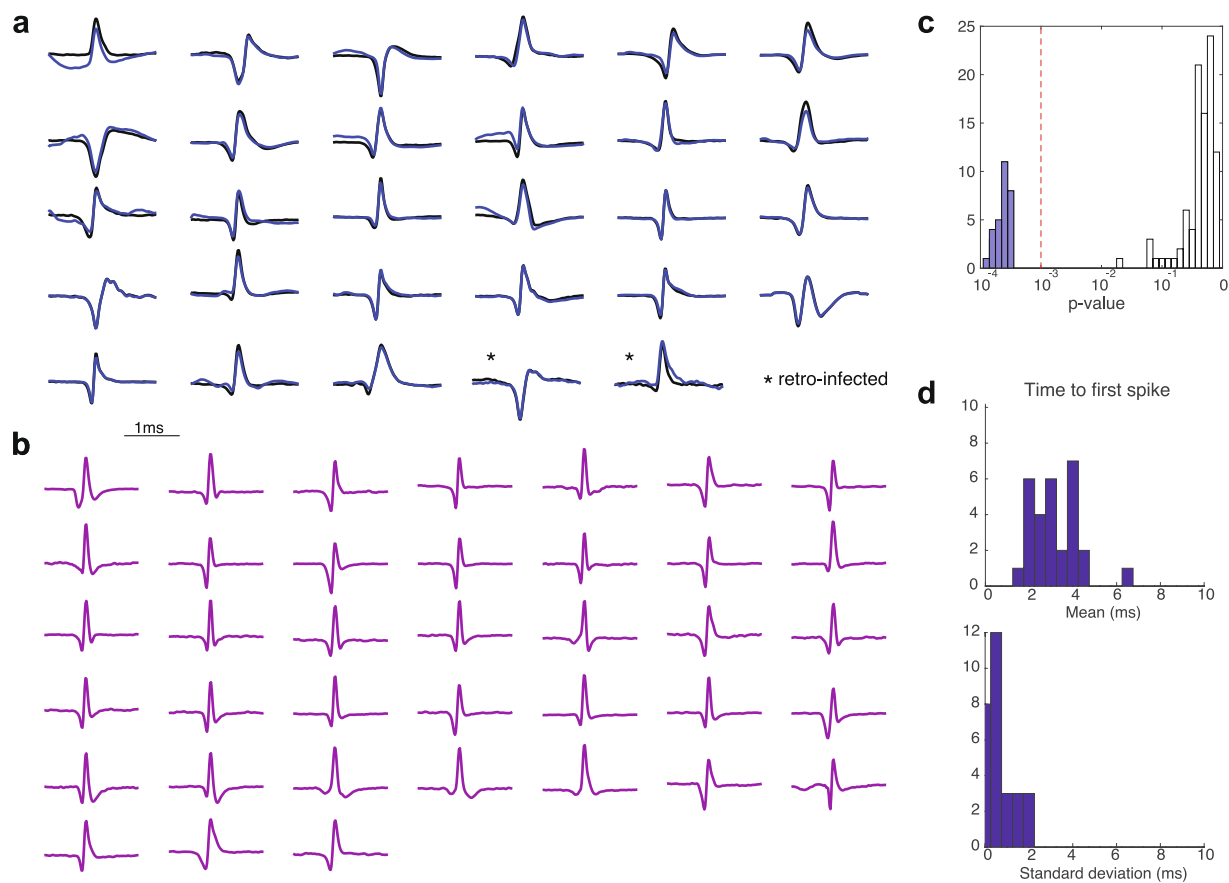
Extended Data Fig. 4 | Identification of light-responsive cells. a, Average waveforms of optogenetically identified dopamine neurons (negative voltage upwards). Average light-evoked waveforms are shown in blue and session-wide average waveforms are in black. All spikes within 10 ms of laser onset were used to construct light-evoked waveform average. Averaged waveforms are normalized to have similar total peak-valley voltages (see Supplementary Fig. 1 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).
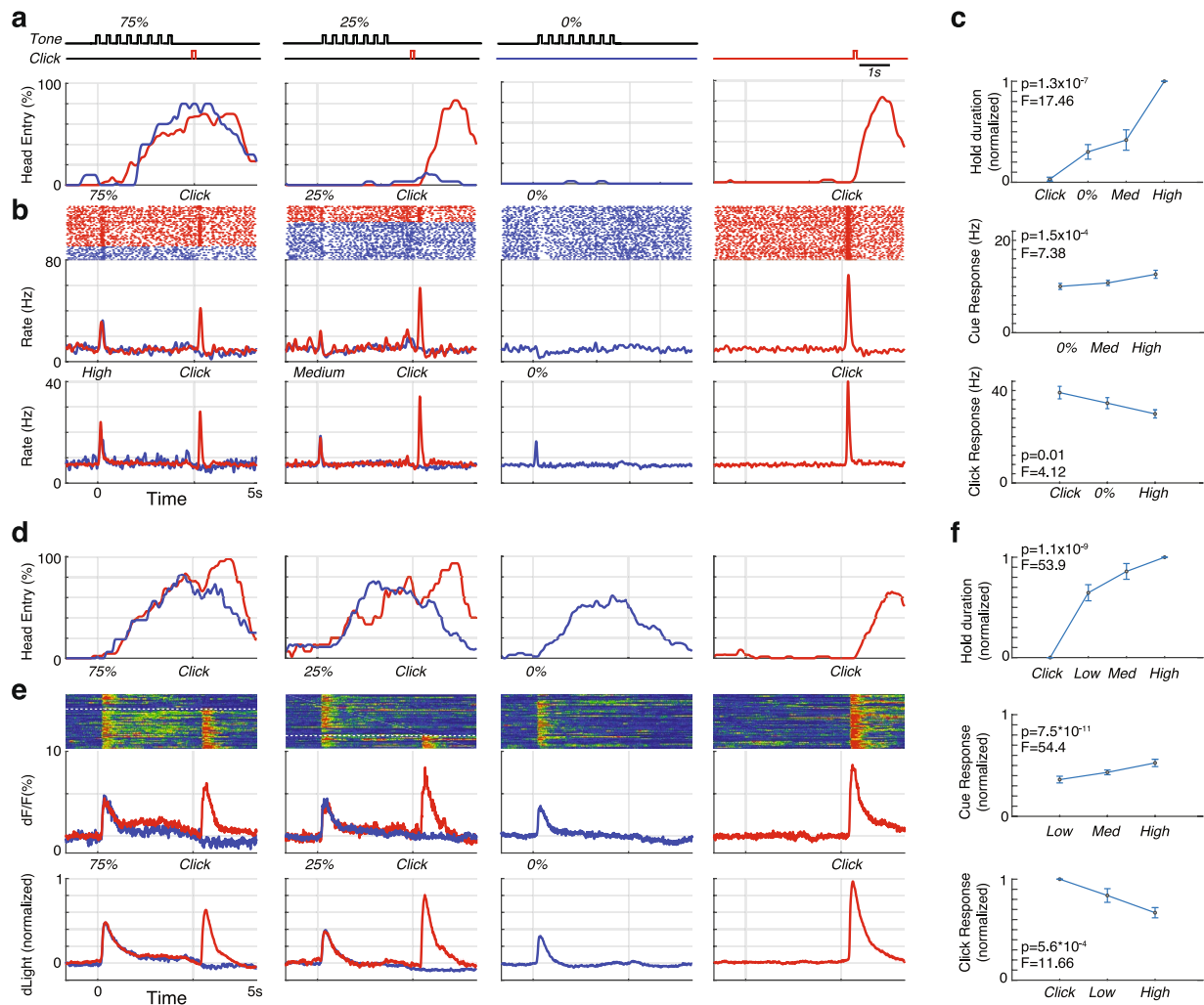
Extended Data Fig. 5 | Dopaminergic responses to Pavlovian cues.
a, Tone pips were followed by reward delivery ('click') with different probabilities (zero, medium or high) depending on the tone pitch. During prior training (average 15.6 sessions, range 2–26) rats had learned about these different probabilities, 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 and blue indicate rewarded and unrewarded trials, respectively. 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 peri-event 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 are presented in the Supplementary Fig. 1. **c**, Behaviour (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 0.3–3-s epoch relative to cue onset, and peak firing rate during 0.5-s duration epochs after cue onset or food-hopper clicks. **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 1 s of cue onset or food-hopper clicks.
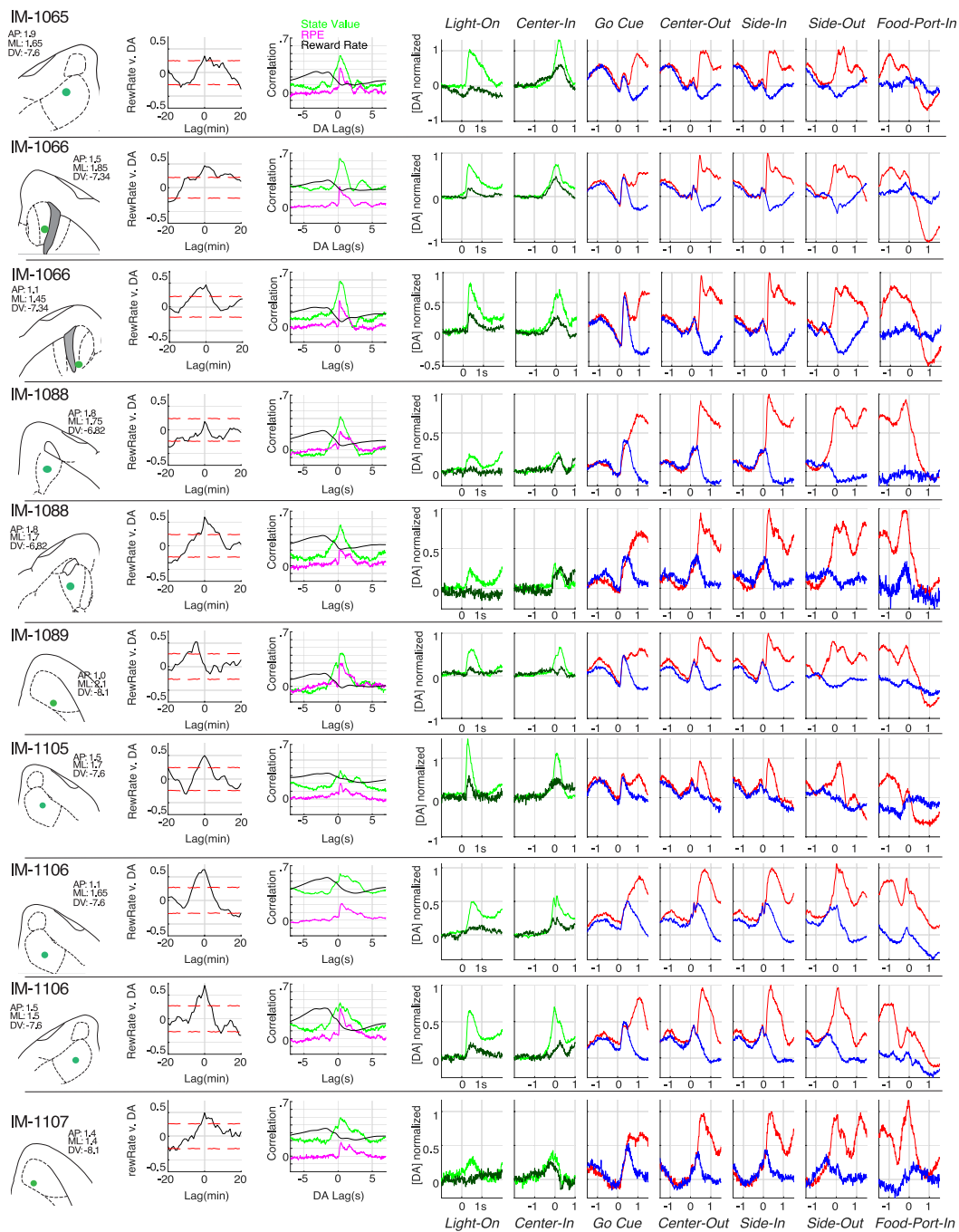
**Extended Data Fig. 6 | Results from each dLight recording session.**
Each row shows a distinct optic fibre placement, and the corresponding recording session that was included in data analyses. For two rats (IM-1066 and IM-1088) we obtained bilateral NAc dLight recordings. From left to right, panels show histologically determined NAc location of fibre tip (within horizontal brain atlas section, including atlas coordinates[51]), 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 Centre-in alignments data are split by latencies <1 s (light green) or >2 s (dark green; as in Fig. 4d); for other alignments, data are split by rewarded (red) and unrewarded (blue) trials. Brain atlas outlines in this figure were reproduced with permission from Paxinos and Watson, 2005[51].
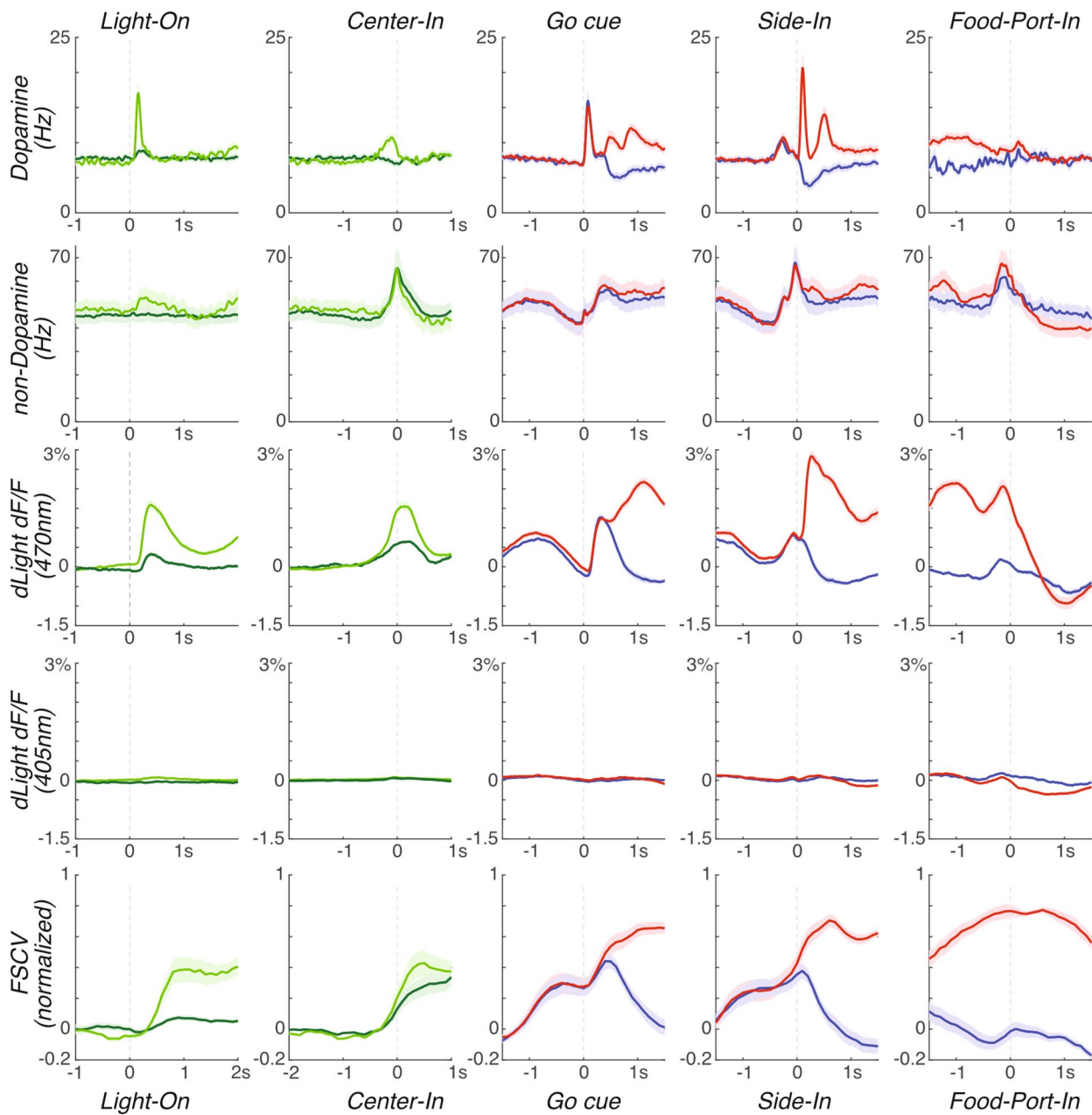
Extended Data Fig. 7 | Comparing event-aligned activity between different signals. Format is as in Fig. 4. dLight fluorescence is here shown separately for 470-nm and 405-nm (control) excitation. Of note, (1) rapid, behaviour-linked dLight fluorescence changes occur at 470 nm, as expected, not in the control 405-nm band; (2) distinct timing of spiking, dLight, and voltammetry responses to cue onsets; and (3) non-dopamine cell firing is much more variable (wider error bands) but on average shows activity during movements: starting just before Centre-in (irrespective of latency), just before Side-in, and just before Food-port-in.
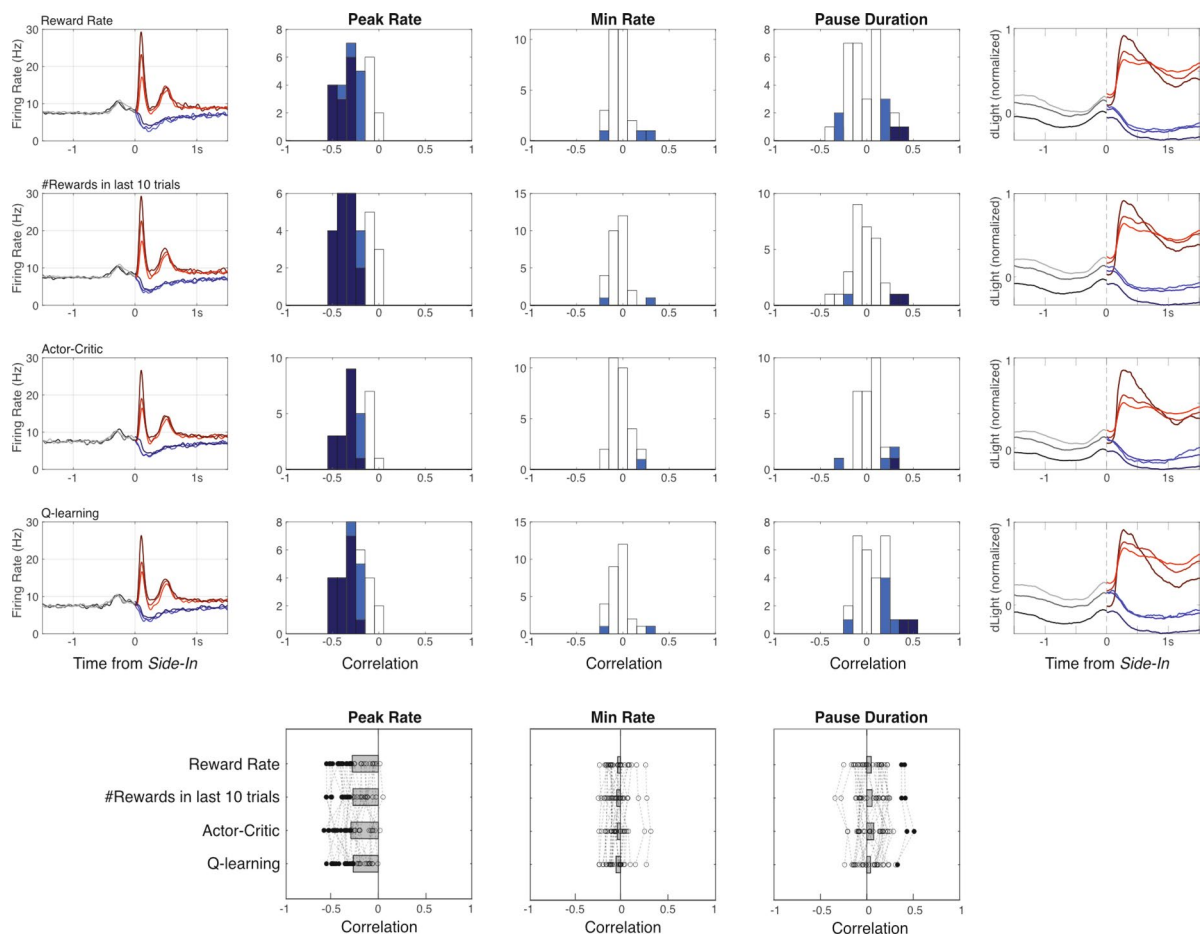
Extended Data Fig. 8 | Different methods for calculating reward expectation produce similar results. Left column, average firing rate of dopamine cells around Side-in, broken down by terciles of reward expectation, based either on recent reward rate (top; same as Fig. 5a), number of rewards in previous ten trials, state value (V) of an actor-critic model or state value ($Q_{left} + Q_{right}$) 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 $V' = V + alpha(RPE)$, in which RPE = actual reward [1 or 0] − V. The Q-learning model kept separate estimates of the probabilities of receiving rewards for left and right choices (Q_{left} and Q_{right}) and updated Q for the chosen action (only) using $Q' = Q + alpha(RPE)$, in which RPE = actual reward [1 or 0] – Q. The learning parameter alpha was determined for each session by best fit to latencies, for V or ($Q_{left} + Q_{right}$) respectively. The subsequent columns show correlations between reward expectation and dopamine cell firing after Side-in, measuring either peak firing rate (within 250 ms after rewarded Side-in), minimum firing rate (middle; within 2 s after unrewarded Side-in) and pause duration (bottom; maximum inter-spike-interval within 2 s 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 is less so.

nature research

Corresponding author(s):

Last updated by author(s): YYYY-MM-DD

Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.

Reporting Summary

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

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Details are provided in methods section. Labview (National Instruments) and Intan (Intan Technologies) were used for behavioral control and electrophysiological recordings.
Data analysis	All analyses were done in MATLAB (Mathworks).

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data is available [at time of publication] through the Collaborative Research in Computational Neuroscience (CRCNS.org) data sharing website. Custom MATLAB code is available upon request to J.D.B.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	No sample size precalculation was performed. Sample sizes are consistent with or larger than related prior publications in Nature.
Data exclusions	No microdialysis data was excluded. For electrophysiology we excluded single units that were not stable over the course of the recording session. For dLight photometry, 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.
Replication	The observation of value-coding by dopamine in the NAc core is a replication of our previous paper, Hamid et al. 2016. No other attempts at replication were made.
Randomization	We did not assign different animals to different groups.
Blinding	Blinding is not relevant because we did not assign different animals to different groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-TH (ab112, abcam)
Validation	Reactivity in rat brain tissue has been validated by abcam. http://www.abcam.com/tyrosine-hydroxylase-antibody-ab112.html

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Long-Evans rats, either wild-type or TH-Cre+, obtained through RRRC.org or bred in-house.
Wild animals	n/a
Field-collected samples	n/a
Ethics oversight	All animal procedures were approved by the University of Michigan or University of California San Francisco Institutional Committees on Use and Care of Animals.

Note that full information on the approval of the study protocol must also be provided in the manuscript.