

Prolonged dopamine signalling in striatum signals proximity and value of distant rewards

Mark W. Howe¹, Patrick L. Tierney¹, Stefan G. Sandberg², Paul E. M. Phillips² & Ann M. Graybiel¹

Predictions about future rewarding events have a powerful influence on behaviour. The phasic spike activity of dopamine-containing neurons, and corresponding dopamine transients in the striatum, are thought to underlie these predictions, encoding positive and negative reward prediction errors^{1–5}. However, many behaviours are directed towards distant goals, for which transient signals may fail to provide sustained drive. Here we report an extended mode of reward-predictive dopamine signalling in the striatum that emerged as rats moved towards distant goals. These dopamine signals, which were detected with fast-scan cyclic voltammetry (FSCV), gradually increased or—in rare instances—decreased as the animals navigated mazes to reach remote rewards, rather than having phasic or steady tonic profiles. These dopamine increases (ramps) scaled flexibly with both the distance and size of the rewards. During learning, these dopamine signals showed spatial preferences for goals in different locations and readily changed in magnitude to reflect changing values of the distant rewards. Such prolonged dopamine signalling could provide sustained motivational drive, a control mechanism that may be important for normal behaviour and that can be impaired in a range of neurologic and neuropsychiatric disorders.

The spike activity patterns of midbrain dopamine-containing neurons signal unexpected and salient cues and outcomes^{1–4,6,7}, and the dynamics of these phasic neural signals have been found to follow closely the principles of reinforcement learning theory^{3–6}. In accordance with this view, selective genetic manipulation of the phasic firing of dopamine neurons alters some forms of learning and cue-guided movements^{8,9}. Episodes of transient dopamine release in the ventral striatum have been detected with FSCV, and these also occur in response to primary rewards and, after learning, to cues predicting upcoming rewards^{10–13}. Thus, dopamine transients in the striatum share many features of the phasic spike activity of midbrain dopamine neurons.

Classic studies of such dopamine transients have focused on Pavlovian and instrumental lever-press tasks, in which rewards were within arm's reach^{1–3,10–13}. However, in many real-life situations, animals must move over large distances to reach their goals. These behaviours require ongoing motivational levels to be adjusted flexibly according to changing environmental conditions. The importance of such control of ongoing motivation is reflected in the severe impairments suffered in dopamine deficiency disorders, including Parkinson's disease. In addition, in pioneering experimental studies, dopamine signalling has been implicated in controlling levels of effort, vigour and motivation during the pursuit of goals in maze tasks^{14–17}. It has been unclear how phasic dopamine signalling alone could account for persistent motivational states¹⁸. We adapted chronic FSCV to enable prolonged measurement of real-time striatal dopamine release as animals learned to navigate towards spatially distant rewards.

We measured dopamine levels in the dorsolateral striatum (DLS) and ventromedial striatum (VMS) (Extended Data Figs 1 and 2, and Methods) as rats navigated mazes of different sizes and shapes to retrieve rewards (Figs 1–4, and Methods). The rats were trained first on an associative T-maze task to run and to turn right or left as instructed

by tones to receive a chocolate milk reward at the indicated end-arms¹⁹ ($n = 9$, Figs 1, 2 and 4). Unexpectedly, instead of mainly finding isolated dopamine transients at the initial cue or at goal-reaching, we primarily found gradual increases in the dopamine signals that began at the onset of the trial and ended after goal-reaching (Fig. 1a, b). These 'ramping' dopamine responses, identified in session averages by linear regression (Pearson's $R > 0.5$, $P < 0.01$), were most common in the VMS (75% of sessions) but were also present at DLS recording sites (42% of sessions). They were evident both in single trials (Fig. 1a–c) and in population averages (Fig. 1e, f, and Extended Data Figs 2g, h and 3), bore no clear relationship to run speed within or across trials (Fig. 1d), and matched, in electrochemical profile, dopamine release evoked by tonic electrical stimulation *in vivo* (Extended Data Fig. 2i, j). Before goal-reaching, the ramps had similar amplitudes in correct (65% overall) and incorrect trials (Fig. 1e, f). After goal-reaching, the signals were significantly larger in correct trials, particularly in the VMS (paired t -test, $P = 0.01$, Fig. 1e, f). Notably, a subset of the session-averaged signals in the DLS (22%, 58 out of 262; 7 probes in 5 rats) showed sustained inhibition up to goal-reaching (Extended Data Fig. 3). Such negative signals were rare in the VMS (5%, 15 out of 300 recordings), suggesting that ramping dopamine signals in the DLS, but not VMS, exhibit heterogeneity in polarity.

We identified isolated phasic transients at warning click indicating trial start and after goal-reaching. These were clearly distinct from the slower ramping responses in approximately 10% of single trials (Extended Data Fig. 4) but were often superimposed on the ramping signals, indicating that the signals recorded could include combinations of transient increases after warning click, slower ramps to goal-reaching, and transient increases after goal-reaching (Extended Data Fig. 4d). The peak magnitudes of the dopamine ramps were comparable to, or slightly smaller than, those of isolated phasic dopamine signals recorded here (Extended Data Fig. 4) and in other studies^{11,12}, and they were correlated with the peak magnitudes of free-reward evoked dopamine measured on the same probes (Pearson's $R = 0.45$, $P < 0.001$, Extended Data Fig. 5), indicating that the ramping signals could be subject to similar regulatory mechanisms and display similar anatomic heterogeneity as classical phasic reward-evoked dopamine signals.

We took advantage of the trial-to-trial variability in the rats' run times (Fig. 2a) to determine whether ramping dopamine release reflected elapsed time or reward proximity, or whether the ramps reflected sums of multiple, accumulated transients to fixed maze cues²⁰. If the dopamine ramps tracked elapsed time, peak dopamine values should have scaled directly with trial time (same slope, different peak height; Fig. 2b, f). If the ramping reflected distance or spatial location relative to goal-reaching (proximity), peak dopamine levels should have been equivalent for shorter and longer trials (different slope, same peak height; Fig. 2c, f). If the ramps were generated by summation of multiple transients, then for characteristic transient dynamics, the signals should have tended to peak at lower values for longer runs than for shorter runs (different slope, different peak height; Extended Data Fig. 6a, b, and Supplementary Discussion). The measured peak dopamine values at goal-reaching were nearly equivalent for short and long

¹McGovern Institute for Brain Research and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Department of Psychiatry and Behavioral Sciences, and Department of Pharmacology, University of Washington, Seattle, Washington 98195, USA.

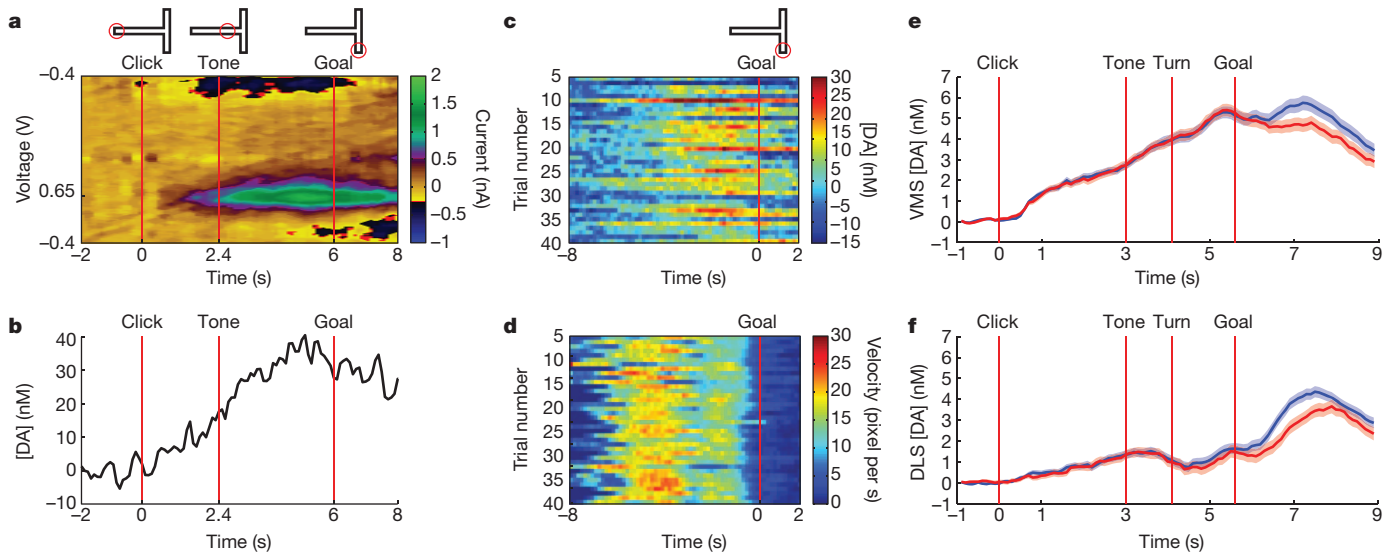


Figure 1 | Ramping striatal dopamine signals occur during maze runs. a, b, Baseline subtracted current (**a**) and dopamine concentration ([DA], **b**) measured by FSCV in VMS during a single T-maze trial. **c, d,** Trial-by-trial changes in dopamine concentration (**c**) and velocity (**d**) relative to

goal-reaching. **e, f,** Dopamine concentration (mean \pm s.e.m.) for VMS (**e**, $n = 300$ session-averaged recordings from 18 probes across 214 sessions) and for DLS (**f**, $n = 262$, 13 probes) for correct (blue) and incorrect (red) trials, averaged over all 40 trial sessions.

trials (Fig. 2e), and were not correlated with trial length (Fig. 2d–f) or with run velocity or acceleration (Extended Data Fig. 6e, f). Moreover, on trials in which rats paused mid-run, the signals remained sustained (or dipped slightly) and resembled the actual proximity to reward (Extended Data Fig. 7). These observations indicated that the ramping signals could represent a novel form of dopamine signalling that provides a continuous estimate of the animal's spatial proximity to distant rewards (Fig. 2, Extended Data Fig. 6, and Supplementary Discussion).

Given that phasic responses of dopamine-containing neurons can reflect the relative value of stimuli²¹, we asked, in a subset of rats, whether the ramping dopamine signals could also be modulated by the size of the delivered rewards (Methods). We used mazes with T, M or S configurations and different total lengths (Fig. 3, and Extended Data Fig. 8). We required the animals to run towards one or the other end of the maze and varied the rewards available at the alternate goal regions. With all three mazes, dopamine ramping became strongly biased towards the goal with the larger reward (Fig. 3, and Extended Data Fig. 8). Run speed was slightly higher for the high-reward maze arms (Fig. 3i, k), but these small differences were unlikely to account

fully for the large differences in the dopamine signals recorded. When we then reversed the locations of the small and large rewards, the ramping signals also shifted, across sessions or just a few trials, to favour the new high-value maze arm (Fig. 3, and Extended Data Fig. 8). These bias effects were statistically significant for each experimental paradigm (Extended Data Fig. 8h–j, Mann–Whitney U-test, $P < 0.05$) and across all rats (Fig. 3d, $n = 4$, Mann–Whitney U-test, $P = 0.02$).

In the M-maze, the ramps became extended to cover the longer end-arm distances to goal-reaching, and critically, peaked at nearly the same level before goal-reaching as did the ramping signals recorded in the T-maze, despite the longer distance travelled (Fig. 3e). This result suggested that the ramping dopamine signals do not signal reward proximity in absolute terms but, instead, scale with the path distance to a fixed level that depends on the relative reward value.

To determine whether such value-related differences in the ramping dopamine signals would occur when the actions to reach the distant goal sites were equivalent, we used the S-shaped maze. The ramping signals were larger for the run trajectories leading to the larger rewards (Fig. 3c, j, and Extended Data Fig. 9), despite the fact that the sequence

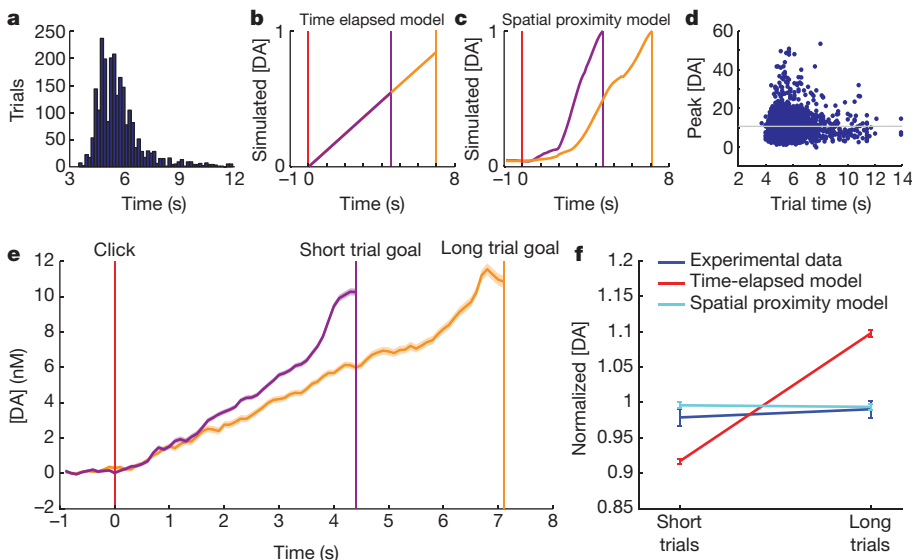


Figure 2 | Ramping dopamine signals proximity to distant rewards. a, Distribution of trial times (from warning click to goal-reaching, $n = 3,933$ trials). **b, c,** Dopamine release modelled as a function of time elapsed since maze-running onset (**b**) and as a function of spatial proximity to visited goal (**c**) for short (purple) and long (orange) trials (see Methods). Vertical lines indicate trial start (red) and end (purple and orange) times. **d,** Peak dopamine concentration versus trial time for all ramping trials ($n = 2,273$, Pearson's $R = 0.0004$, $P = 0.98$). **e,** Experimentally recorded dopamine release (mean \pm s.e.m.) in short ($n = 327$, purple) and long ($n = 423$, orange) trials. Dopamine peaks at equivalent levels, as in the proximity model in **c**. **f,** Normalized peak dopamine levels (mean \pm s.e.m.) predicted by time-elapsed (red) and proximity (light blue) models, and measured experimental data (dark blue).

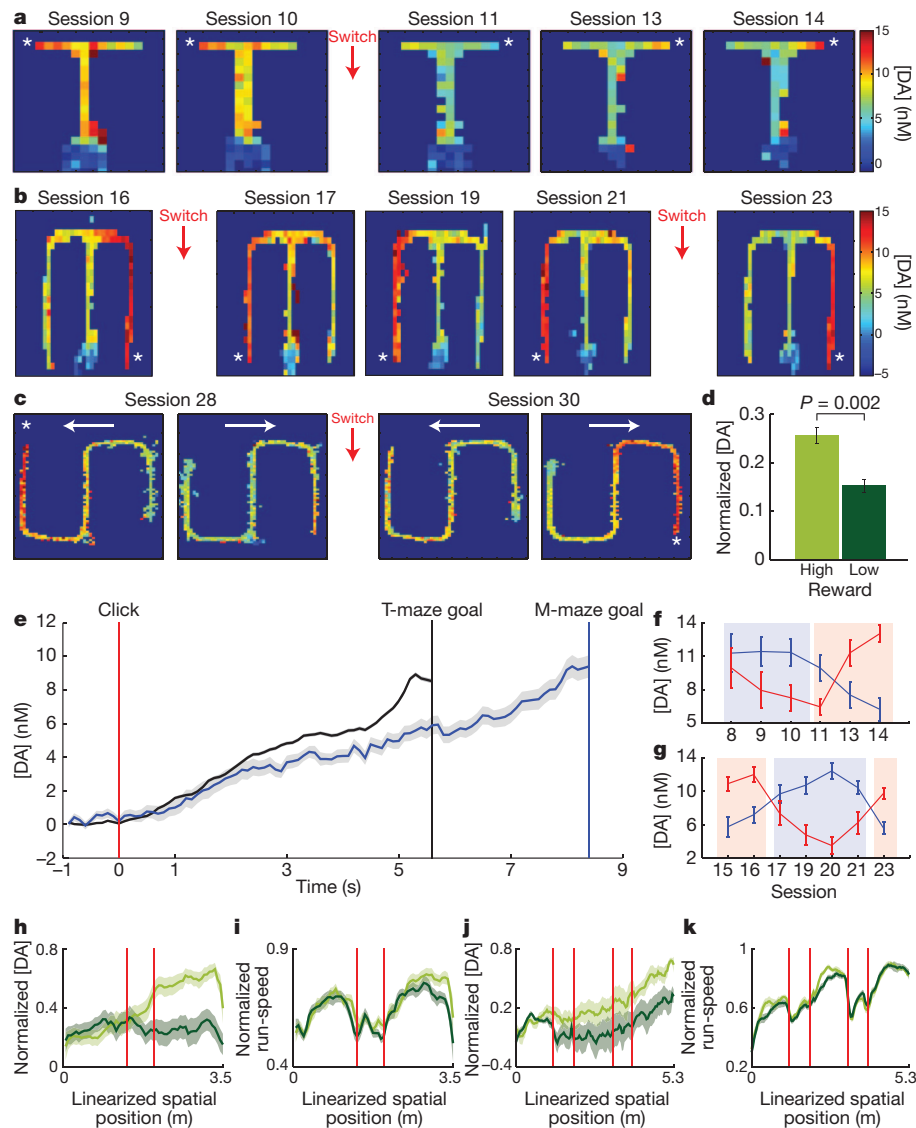


Figure 3 | Dopamine ramping is sensitive to reward magnitude.

a, b, Average dopamine signals from a VMS probe, for consecutive T-maze (**a**) and M-maze (**b**) sessions with asymmetric rewards. Asterisks indicate the goal with the larger reward; red arrows (and Switch) indicate reversal of reward amounts. **c,** Dopamine signals from a different rat running in the S-maze. White arrows indicate run direction. **d,** Average (\pm s.e.m.) peak dopamine across all value experiments ($n = 4$ rats). **e,** Average (\pm s.e.m.) VMS dopamine during T-maze ($n = 44$ sessions in 3 rats, black) and M-maze ($n = 17$, blue)

sessions in the same rats. **f, g,** Average (\pm s.e.m.) peak dopamine signals for the sessions plotted in **a** (**f**) and **b** (**g**) for trials to left (blue) and right (red) goals. Shading indicates arm with larger reward. **h, i,** Average normalized dopamine (**h**) and running speed (**i**) for runs to high (light green) and low (dark green) reward goals in the M-maze. Vertical lines indicate turns. **j, k,** Average normalized dopamine (**j**) and running speed (**k**) in the S-maze ($n = 9$ sessions in 2 rats), plotted as in **h** and **i**.

of turns and the lengths of the runs needed to reach the larger and smaller rewards were equivalent for both trajectories ($n = 2$ rats, 4 and 5 sessions per rat, Fig. 3c, j, k, and Extended Data Figs 8 and 9).

In rats performing the free-choice associative version of the T-maze task, robust dopamine signal biases existed in approximately 20% of sessions (Mann–Whitney U-test, $P < 0.05$) and significantly more often than chance overall (Z-test, $P < 0.00001$ versus bootstrapped variances; Methods and Fig. 4a, b, d). These biases were maintained across consecutive training sessions for individual animals (Fig. 4b), did not relate to run-speed biases (Fig. 4c, and Extended Data Fig. 10c) or recording hemisphere (Fig. 4d, and Extended Data Fig. 10a, b, f), and, notably, emerged gradually over days as performance improved and training progressed (Fig. 4e, f). Although not obviously related to imbalances in maze cues or differences in left–right performance, they displayed a weak association with right end-arm choice biases that developed late in training (Extended Data Fig. 10d, e, g). Thus, end-arm biases in the ramping dopamine signals could develop even in the absence

of experimentally imposed discrepancies in value, possibly reflecting developing internal value estimates (Supplementary Discussion).

Ramping spike-firing has been recorded for putative midbrain dopamine neurons in head-fixed primates under conditions of reward uncertainty²² and for nigral non-dopamine-containing neurons¹. We asked whether the magnitudes of the ramping dopamine signals that we recorded in the striatum changed as performance improved on the free-choice associative T-maze task (Fig. 4e). They did not (Pearson's $R = -0.08$, $P = 0.19$; Extended Data Fig. 10h–j), suggesting that uncertainty about reward probability was unlikely to have controlled the magnitude of the ramping signals²² (Supplementary Discussion).

Classic studies of dopamine neuron firing and striatal dopamine release have largely focused on transient responses associated with unpredicted rewards and reward-predictive cues. Here we demonstrate that, in addition to such transient dopamine responses, prolonged dopamine release in the striatum can occur, changing slowly as animals approach distant rewards during spatial navigation. These

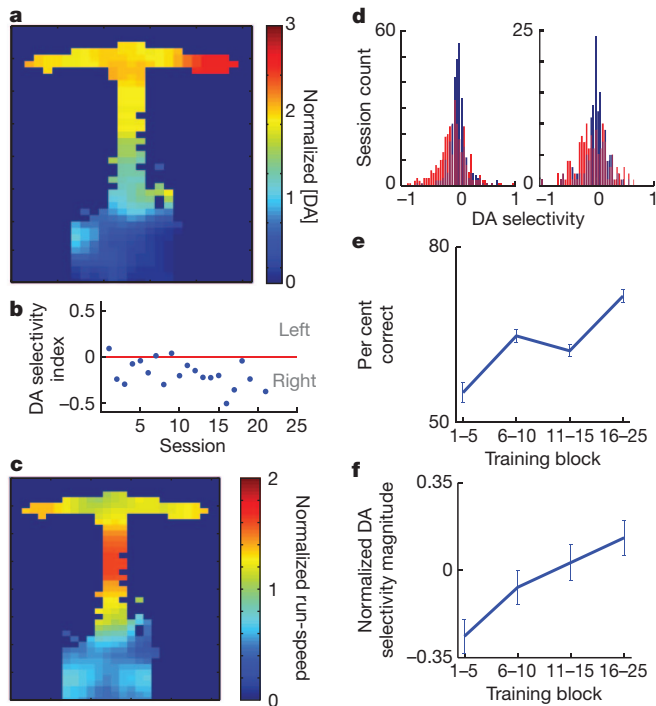


Figure 4 | Ramping dopamine selectivity can emerge with training without experimentally imposed reward discrepancies. **a**, Average normalized dopamine at a VMS site as a function of maze location ($n = 19$ sessions). **b**, Dopamine selectivity indices (Methods) for all individual sessions averaged in **a**. **c**, Average running speed for sessions in **a**. **d**, Selectivity indices for all VMS (left) and DLS (right) recordings (red) compared to shuffled data (blue) for all rats ($n = 9$). **e**, **f**, Average per cent correct performance (**e**) and average Z -score normalized dopamine selectivity (**f**) across training blocks. Error bars, s.e.m.

dopamine signals seem to represent the relative spatial proximity of valued goals, perhaps reflecting reward expectation²³. It remains unclear whether these signals represent goal proximity on the basis of environmental cues, effort, or internally scaled estimates of distance. However, the brain possesses mechanisms for representing both allocentric spatial context and relative distance from landmarks²⁴, which could, in principle, be integrated with dopaminergic signalling to produce such extended dopamine signals.

Transient dopaminergic responses to learned reward-predictive cues have been proposed to initiate motivated behaviours^{25,26}, but with this mode of signalling alone, it is difficult to account for how dopamine acts to maintain and direct motivational resources during prolonged behaviours (Supplementary Discussion). The ramping dopamine signals that we describe here, providing continuous estimates of how close rewards are to being reached, and weighted by the relative values of the rewards when options are available, seem ideally suited to maintain and direct such extended energy and motivation.

METHODS SUMMARY

Male Long Evans rats ($n = 9$) were deeply anaesthetized and were surgically implanted with headstages carrying voltammetry microensors placed in the VMS (anteroposterior $+1.5$ mm, mediolateral ± 2.1 mm, dorsoventral $6-7$ mm) and DLS (anteroposterior $+0.5$ mm, mediolateral ± 3.5 mm, dorsoventral $3.5-4.0$ mm) bilaterally ($n = 1$ rat) or in the left ($n = 5$) or right ($n = 3$) hemisphere. Triangular voltage sweeps (-0.4 V to 1.3 V) were applied at 10 Hz to the carbon fibre microsensor probes²⁷ relative to an Ag/AgCl reference electrode implanted in the neocortex. Current changes due to dopamine oxidation-reduction reactions were verified by principal component regression²⁸ using a training set of current profiles from evoked dopamine release and pH changes recorded in 5 rats in response to electrical stimulation of the medial forebrain bundle (MFB). Training on the T-maze task¹⁹ (Figs 1, 2 and 4) and chronic voltammetry recording began 4 weeks after surgery, and continued for 15 to 35 days for each rat. The position

of each rat was monitored continually by tracking software as the rat navigated the maze (Neuralynx), and the positions were used to trigger auditory cues and reward pumps when the animal entered specific maze locations. After T-maze training, a subset of rats ($n = 3$) was trained in an extended M-maze. Rats were required either to continue performing the associative tone-cued task (M31) or to visit one of the two end-arms on each trial by blockade of the other end-arm (M36 and M47). Another subset of rats ($n = 2$) underwent training on an S-maze version of the task in which they only had to run back and forth from end to end to retrieve rewards. Voltammetry data were analysed with in-house MATLAB codes (Mathworks), and probe positions were verified by standard histology¹⁹.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 January; accepted 17 July 2013.

Published online 4 August 2013.

- 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).
- Matsumoto, M. & Hikosaka, O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* **459**, 837–841 (2009).
- Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412**, 43–48 (2001).
- Bayer, H. M. & Glimcher, P. W. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**, 129–141 (2005).
- Schultz, W. Getting formal with dopamine and reward. *Neuron* **36**, 241–263 (2002).
- Redgrave, P. & Gurney, K. The short-latency dopamine signal: a role in discovering novel actions? *Nature Rev. Neurosci.* **7**, 967–975 (2006).
- Zweifel, L. S. *et al.* Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proc. Natl Acad. Sci. USA* **106**, 7281–7288 (2009).
- Wang, L. P. *et al.* NMDA receptors in dopaminergic neurons are crucial for habit learning. *Neuron* **72**, 1055–1066 (2011).
- Day, J. J., Jones, J. L., Wightman, R. M. & Carelli, R. M. Phasic nucleus accumbens dopamine release encodes effort- and delay-related costs. *Biol. Psychiatry* **68**, 306–309 (2010).
- Day, J. J., Roitman, M. F., Wightman, R. M. & Carelli, R. M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nature Neurosci.* **10**, 1020–1028 (2007).
- Gan, J. O., Walton, M. E. & Phillips, P. E. Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. *Nature Neurosci.* **13**, 25–27 (2010).
- 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).
- Braun, A. A., Graham, D. L., Schaefer, T. L., Vorhees, C. V. & Williams, M. T. Dorsal striatal dopamine depletion impairs both allocentric and egocentric navigation in rats. *Neurobiol. Learn. Mem.* **97**, 402–408 (2012).
- Salamone, J. D., Correa, M., Farrar, A. & Mingote, S. M. Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology (Berl.)* **191**, 461–482 (2007).
- Whishaw, I. Q. & Dunnett, S. B. Dopamine depletion, stimulation or blockade in the rat disrupts spatial navigation and locomotion dependent upon beacon or distal cues. *Behav. Brain Res.* **18**, 11–29 (1985).
- Salamone, J. D. & Correa, M. The mysterious motivational functions of mesolimbic dopamine. *Neuron* **76**, 470–485 (2012).
- Niv, Y., Daw, N. D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* **191**, 507–520 (2007).
- Barnes, T. D., Kubota, Y., Hu, D., Jin, D. Z. & Graybiel, A. M. Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature* **437**, 1158–1161 (2005).
- Niv, Y., Duff, M. O. & Dayan, P. Dopamine, uncertainty and TD learning. *Behav. Brain Funct.* **1**, 6 (2005).
- Tobler, P. N., Fiorillo, C. D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642–1645 (2005).
- Fiorillo, C. D., Tobler, P. N. & Schultz, W. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* **299**, 1898–1902 (2003).
- Hikosaka, O., Sakamoto, M. & Usui, S. Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J. Neurophysiol.* **61**, 814–832 (1989).
- Derdikman, D. & Moser, E. I. A manifold of spatial maps in the brain. *Trends Cogn. Sci.* **14**, 561–569 (2010).
- Flagel, S. B. *et al.* A selective role for dopamine in stimulus-reward learning. *Nature* **469**, 53–57 (2011).
- Berridge, K. C. The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology (Berl.)* **191**, 391–431 (2007).

27. Clark, J. J. *et al.* Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. *Nature Methods* **7**, 126–129 (2010).
28. Keithley, R. B., Heien, M. L. & Wightman, R. M. Multivariate concentration determination using principal component regression with residual analysis. *Trends Analyt. Chem.* **28**, 1127–1136 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Telian, L. Li, T. Lechina, S. Ng-Evans and N. Hollon for help, and M. J. Kim, K. S. Smith, T. J. Sejnowski, L. G. Gibb and Y. Kubota for their comments. This work was supported by US National Institutes of Health (NIH) grant R01 MH060379 (A.M.G.), National Parkinson Foundation (A.M.G.), CHDI Foundation

grant A-5552, the Stanley H. and Sheila G. Sydney Fund (A.M.G.), a Mark Gorenberg fellowship (M.W.H.), and NIH grants R01 DA027858 and R01 MH079292 (P.E.M.P.).

Author Contributions M.W.H. and A.M.G. designed the experiments, analysed the data, and wrote the manuscript. M.W.H. conducted the experiments. P.L.T. helped to develop the FSCV chronic recording setup. S.G.S. and P.E.M.P. provided technical training on FSCV and advice on data analysis. All authors contributed to discussion and interpretation of the findings.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M.G. (graybiel@mit.edu).

METHODS

All experimental procedures were approved by the Committee on Animal Care at the Massachusetts Institute of Technology and were in accordance with the US National Research Council Guide for the Care and Use of Laboratory Animals. Sample sizes were chosen based on signal variability estimates from other published studies using FSCV.

Implant procedures. Implantations were performed on deeply anaesthetized fully mature male Long Evans rats ($n = 9$; 380 to 450 g), in sterile conditions according to approved surgical procedures¹⁹ with headstages carrying 1 to 3 independently movable voltammetry probes targeting the DLS (anteroposterior +0.5 mm, mediolateral ± 3.5 mm, dorsoventral 3.5–4.0 mm), of the right ($n = 3$) or left ($n = 5$) hemisphere, or the DLS bilaterally ($n = 1$), 1 to 3 probes targeting the VMS of the same hemispheres (anteroposterior +1.5 mm, mediolateral ± 2.1 mm, dorsoventral 6–7 mm), and a unilateral Ag/AgCl reference electrode in the posterior cortex (anteroposterior –2.3 mm, mediolateral ± 3.5 mm, dorsoventral ~ 0.5 mm). Five rats that underwent maze training and three additional rats for acute stimulation experiments were implanted with tungsten bipolar stimulation electrodes (FHC) straddling the ipsilateral MFB (anteroposterior –4.6 mm, mediolateral ± 1.3 mm, dorsoventral 7–8 mm) to verify striatal dopamine release (see below).

Behavioural training. All behavioural training was conducted on a custom built 'grid maze' with fully reconfigurable tracks and walls. Training on an associative T-maze task with auditory instruction cues (Figs 1, 2, and 4)¹⁹ began 4 weeks after implantation. Voltammetric recordings began when rats learned to run smoothly down the track to retrieve reward. Early sessions with sporadic maze behaviour, such as wall rearing and sluggish initiation of maze running, were discarded. Daily behavioural sessions consisted of 40 trials. Trials began with a warning click, followed 0.5 s later by the lowering of a swinging gate, allowing the rat to run down the maze. Half-way down the long arm, a tone was triggered (1 or 8 kHz), indicating which end-arm to visit in order to retrieve chocolate milk reward (0.3 ml) delivered through automated syringe pumps (Pump Systems) upon the rat's arrival. The spatial position of each rat was monitored continually by video tracking (Neuralynx). Tone delivery and syringe pumps were controlled by in-house behavioural software written in MATLAB (Mathworks). After 15 to 35 T-maze sessions per rat, a subset of rats ($n = 3$) received 17 training sessions (4 to 6 sessions each) on the M-maze task in which the end-arms of the T-maze were extended (Fig. 3). These rats received a larger amount of reward (0.4 ml) at one goal site than at the other (0.1 ml for 2 rats and 0.2 ml for 1 rat). After 2 to 3 sessions with a given set of spatial reward contingencies, the reward amounts at the two goals were reversed. One rat (M31) was required to make turn choices in response to tones as in the previous T-maze task, whereas the other two rats (M36 and M47) were directed pseudo-randomly to one end-arm of the maze on each trial by removing the track to the opposite arm (20 trials to each arm) without tone presentation. Two rats were trained on the S-maze task (Fig. 3). These rats were required simply to run back and forth to retrieve a large volume of chocolate milk (0.4 ml) at one goal and a small volume (0.1 ml) at the other goal. Consecutive visits to the same reward site did not trigger the reward pumps.

Voltammetry data acquisition and analysis. Waveform generation and data acquisition for voltammetry recordings were done with two PCI data acquisition cards and software written in LabVIEW (National Instruments). Triangular voltage waveforms were applied to chronically implanted carbon fibre electrodes, relative to the reference electrode, at 10 Hz. Electrodes were held at –0.4 V between scans, and were ramped to 1.3 V and back to –0.4 V during each scan²⁷. Current produced by redox reactions was recorded during voltage scans.

We compiled a library of current versus applied voltage templates for dopamine and pH changes of varying magnitudes by stimulating the MFB (60 Hz, 24 pulses, 300 μ A) to induce dopamine release in the striatum in 5 rats maintained under isoflurane anaesthesia. We used these templates from all 5 rats as a training set to perform chemometric analysis²⁸ on voltammetry measurements obtained during behaviour with in-house MATLAB software. This procedure allowed us to distinguish changes in current due to dopamine release from changes due to pH or to other electroactive substances²⁸. In a separate set of rats, we stimulated the MFB (10 Hz, 60 pulses, 100–120 μ A) under urethane anaesthesia to mimic the slower, low amplitude ramping signals that we observed in behaving animals (Extended Data Fig. 2). Current changes were converted to estimated dopamine concentration by using calibration factors obtained from *in vitro* measurements of fixed dopamine concentrations. Behavioural video tracking was synchronized with voltammetry recordings by marker transistor–transistor logic signals sent to the voltammetry data acquisition system.

For each trial, voltammetry data were normalized by subtracting average background current at each potential measured during the 1-s baseline period before warning click. Session averaged traces (Figs 1 and 3, and Extended Data Figs 3 and 10) were computed by averaging the dopamine signals recorded in a single session across 40 trials, and then averaging these traces to obtain global averages across all

rats and electrodes. Each session-averaged trace (one from each probe from each session) was considered as an independent measurement for computing s.e.m. Concatenation of dopamine and proximity signals (see below) was performed by scaling the peri-event windows using the median inter-event intervals between consecutive events across all trials (Figs 1–3, and Extended Data Figs 3, 4 and 10). Traces between two consecutive events were plotted by taking data from each event to half of the median inter-event interval. Maze arm selectivity (Fig. 4 and Extended Data Fig. 10) was computed by the following equation:

$$\text{Dopamine selectivity index} = ([\text{DA}]_{\text{left}} - [\text{DA}]_{\text{right}}) / ([\text{DA}]_{\text{left}} + [\text{DA}]_{\text{right}})$$

where $[\text{DA}]_{\text{left}}$ and $[\text{DA}]_{\text{right}}$ represent dopamine concentration during trials to the left and right arms of the maze, respectively.

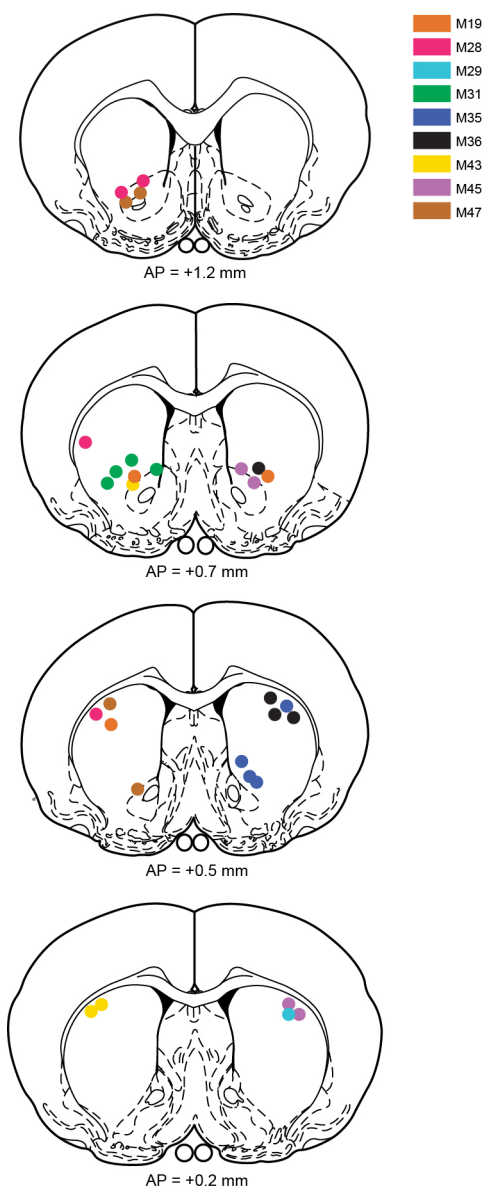
Session-averaged dopamine traces were identified as having positive or negative ramping characteristics (Extended Data Fig. 3) if they exhibited a significantly positive or negative linear regression coefficient (Pearson's, $R > 0.5$ or $R < -0.5$ and $P < 0.01$) over the entire trial period. Trials with phasic responses around the trial start (Extended Data Fig. 4) were identified by calculating the relative difference between consecutive time points (100 ms per sample) in a 1-s window with its centre sliding in 0.1-s steps for a 1-s period from 0.5 s to 1.5 s after warning click. For a given window position, if the differences were all positive values across the first half of the window (0.5 s) and negative across the second half, we determined that a significant inflection point was present in that trial. Comparison of dopamine signals on long and short trials (Fig. 2) was done by selecting trials that fell in the bottom third (short trials) and in the top third (long trials) of the trial-time distribution for each rat that displayed a dopamine peak within 0.5 s of goal-reaching. Trials with noisy video tracking data were discarded from this analysis. The simulations for the time-elapsed model (Fig. 2b, f) were made by calculating the average slope of the ramping signals across all trials on each session and by using linear extrapolation to predict the peak dopamine values on each trial within that session. Peak values were normalized to the median peak value for each session individually and averaged for short and long trials to generate the predictions in Fig. 2f. The multi-transient model (Extended Data Fig. 6) was implemented in MATLAB and tested using a range of physiologically realistic estimates for the slope and decay times for previously observed transient signals *in vivo*¹¹. For the model results shown in Extended Data Fig. 6a, c, simulations were run 100 times, and temporal offset times for the transients were determined by drawing randomly (normrnd function in MATLAB) from a normal distribution with means of 0.8 s (with standard deviation of 0.5 s) and 1.4 s (with standard deviation of 7 s) for short and long trials, respectively (Extended Data Fig. 6a, c). Spatial proximity to goal (Fig. 2c and Extended Data Fig. 7b, e) was calculated by summing the pixel differences in the x and y spatial dimensions for each recorded rat's position. These traces were averaged across all short and long trials separately to generate the traces shown in Fig. 2c. Session-by-session estimations of peak dopamine concentration were made by randomly generating peak trial values using the mean and standard deviation of peak values present in the experimental data. All peak values for short and long trials (Fig. 2f and Extended Data Fig. 6b, d) were normalized to the mean dopamine concentration for each session for both simulated and experimental data.

The presence of population selectivity in the dopamine signals on the associative T-maze task (Fig. 4) was determined by comparing the variance of selectivity indices from the T-maze recordings to the distribution of variances obtained by shuffling the dopamine concentrations on the two end-arms and bootstrapping 10,000 times (Z -test comparing variance of the data to the variances of the bootstrapped distributions, $P < 0.00001$). To identify changes in selectivity and ramping magnitude across training, Z -scores of selectivity indices (Fig. 4f) were computed for each rat by taking the absolute values of all selectivity indices, then by normalizing across sessions before combining all values across rats. Choice selectivity (Extended Data Fig. 10) was computed similarly to the selectivity score for dopamine:

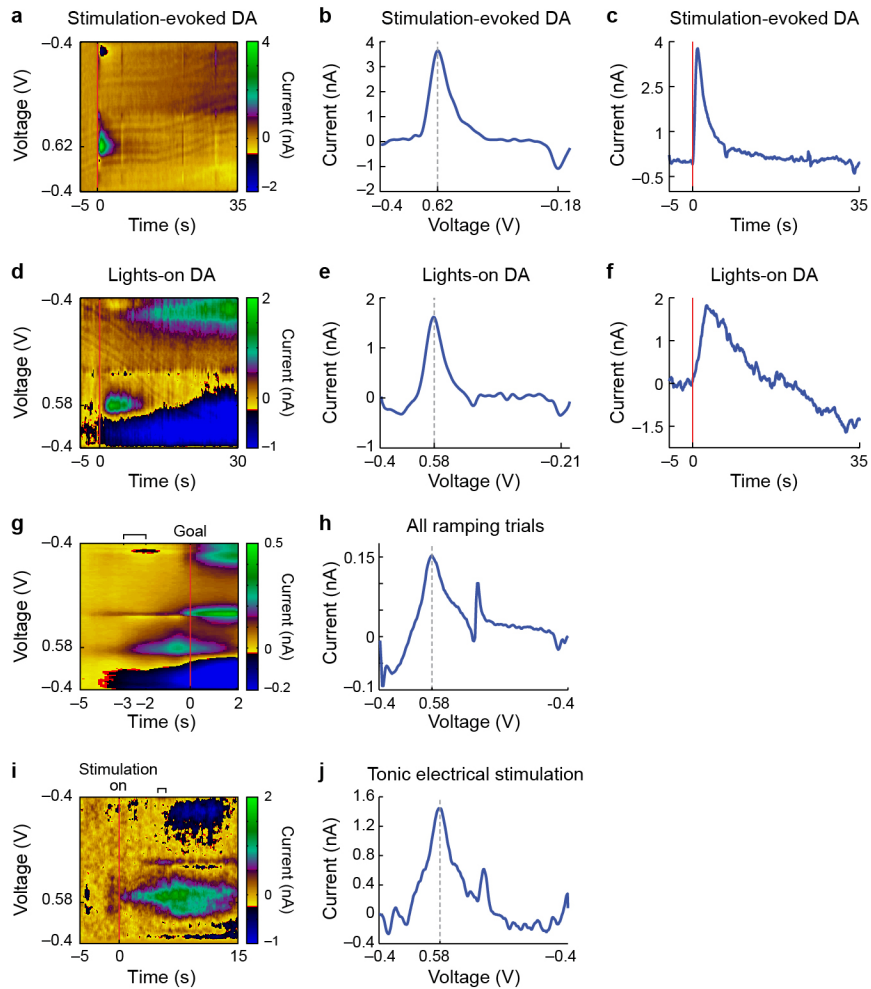
$$\text{Behavioural selectivity index} = (\text{left arm choices} - \text{right arm choices}) / (\text{left arm choices} + \text{right arm choices})$$

Run time biases and per cent correct biases (Extended Data Fig. 10) between the two arms were also computed in this way. Correlations between peak dopamine magnitude and per cent correct performance were calculated by normalizing the average peak dopamine values on each trial to the average peak value across all trials within that session.

Histology. Probe positions were verified histologically¹⁹. Brains were fixed by transcardial perfusion with 4% paraformaldehyde in 0.1 M NaKPO₄ buffer, post-fixed, washed in the buffer solution, and cut transversely at 30 μ m on a freezing microtome, and they were stained with cresylecht violet to allow reconstruction of the recording sites (Extended Data Fig. 1). For a subset of the probes, a constant current (20 mA, 20 s) was passed through the probe before fixation to make micro-lesions at probe-tip locations.

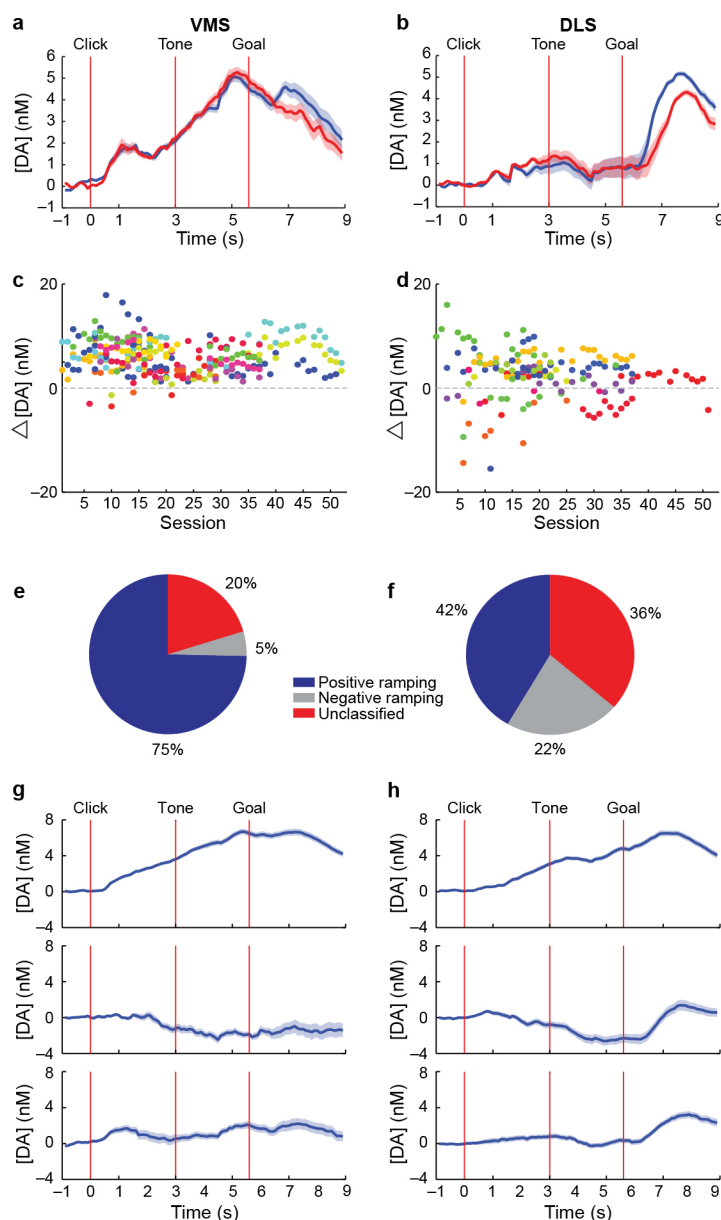


Extended Data Figure 1 | Schematic drawing of recording locations.
 Coloured dots indicate positions of probe tips as determined by post-experiment probe length measurements and histology (see Methods). Each colour represents probes from one rat.



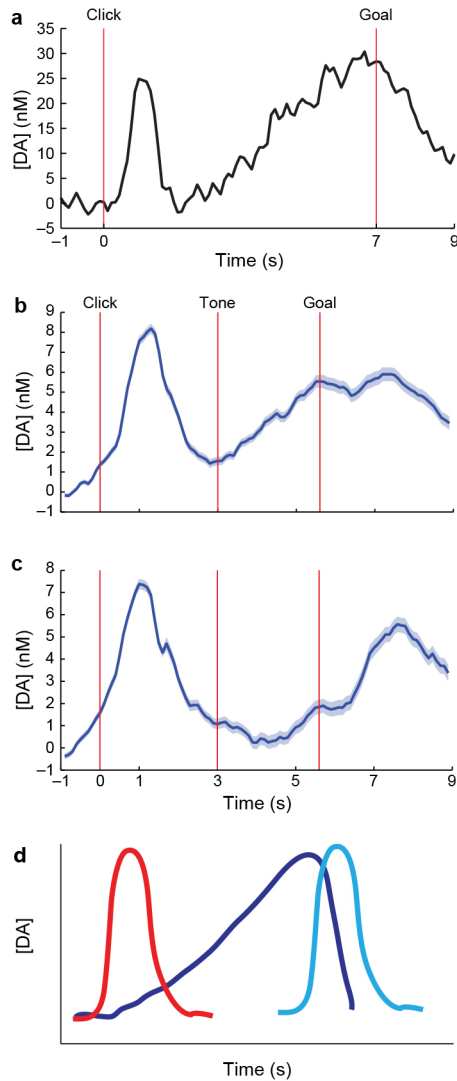
Extended Data Figure 2 | Comparisons of dopamine release in the VMS measured in behaving rats and evoked by electrical MFB stimulation in anaesthetized rats. **a**, Current measured at each potential during successive scans with electrical MFB stimulation (red line) in an anaesthetized rat. **b**, Current–voltage plot from the 5 scans during the 0.5-s period after MFB stimulation. The large increase in current around 0.6 V (dotted line, peak) corresponds to the dopamine redox potential measured *in vitro*. **c**, Time course of current at the dopamine redox potential around electrical MFB stimulation

(red line). **d–f**, Plots, as in **a–c**, illustrating dopamine response in the same rat, now awake, to room lights being turned on. **g**, Current changes averaged over all trials ($n = 4,418$ trials) in which ramping occurred (see Methods) during T-maze running. **h**, Average current–voltage plot for all identified ramping trials for the time period (–3 to –2 s) indicated by the brackets in the colour plot in **g**. **i**, Average current changes induced by tonic MFB stimulation in anaesthetized rats ($n = 3$). **j**, Average current–voltage plot from the bracketed time range (10 to 11 s) in **i** following the onset of stimulation.

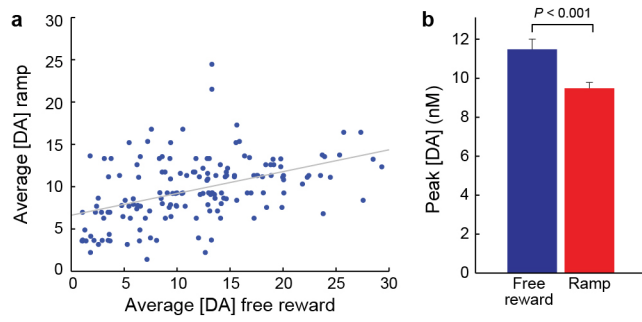


Extended Data Figure 3 | Patterns of dopamine signals in the VMS and DLS. **a, b**, Dopamine concentration recorded in the VMS (**a**, $n = 9$ rats) and in the DLS (**b**, $n = 8$ rats). Data were first averaged across trials to yield session average traces for each probe in each session. These traces were averaged within rat to obtain one average trace per rat, which were then averaged across rats. These plots differ from those in Fig. 1e, f, which considered session averages for each probe to be an independent measure. Shading represents s.e.m. across rats. **c, d**, Distribution of average peak dopamine values for all recordings in VMS

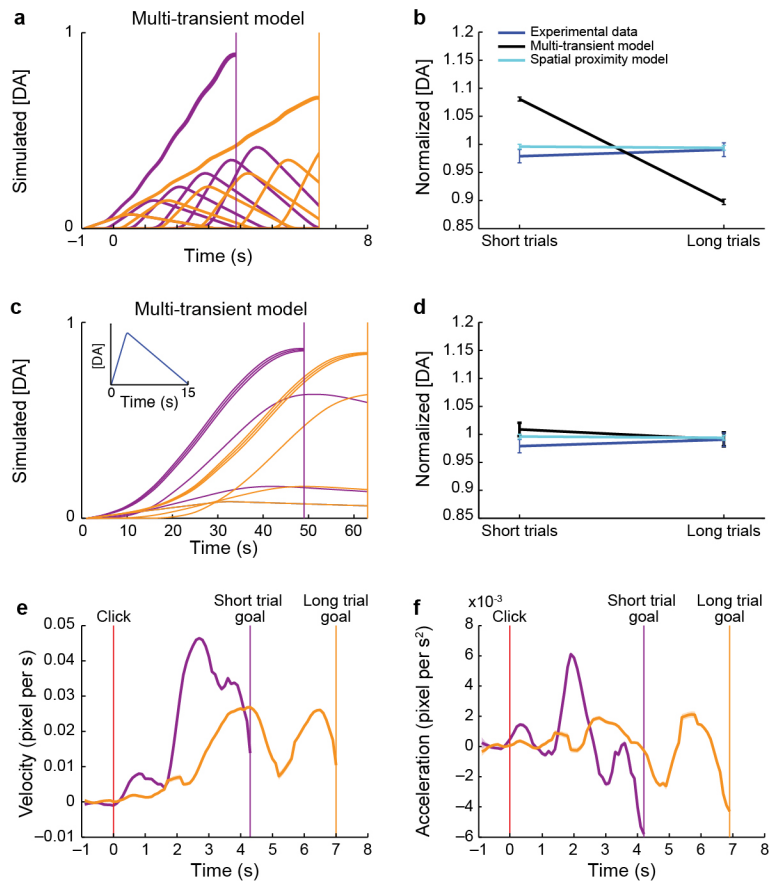
(**c**) and DLS (**d**). Each colour corresponds to an average peak dopamine concentration measured by a single probe in different sessions. **e, f**, Proportion of trial averaged dopamine recordings in the VMS (**e**, out of 300) and DLS (**f**, out of 262) that displayed a positive (blue) or negative (grey) ramping response during maze running, and an unclassified dopamine profile (red). **g, h**, Average dopamine concentration in the VMS (**g**) and DLS (**h**) for the positive ramping traces (top), the negative ramping traces (middle), and the unclassified traces (bottom). Shading represents s.e.m.



Extended Data Figure 4 | Transient responses to cues and rewards occur during the T-maze task performance and are distinct from the sustained ramping response. **a**, Dopamine concentration in a representative trial that included both a phasic response to warning click and a sustained ramping response to goal-reaching. **b**, **c**, Average normalized dopamine traces from VMS (**b**) and DLS (**c**) probes from all trials ($n = 890$ and 640 , respectively) that showed identified transients after warning click (see Methods). In the VMS, note the sharp increase in dopamine around warning click superimposed on the ramping response that followed the phasic click response. Shading represents s.e.m. calculated across trials. **d**, Model for dopamine release profiles in the T-maze task. Sharper transient responses are present at the start of maze running (red) and after goal-reaching (cyan). These responses can be superimposed on and modulated independently of the slower ramping signal related to goal proximity (dark blue).

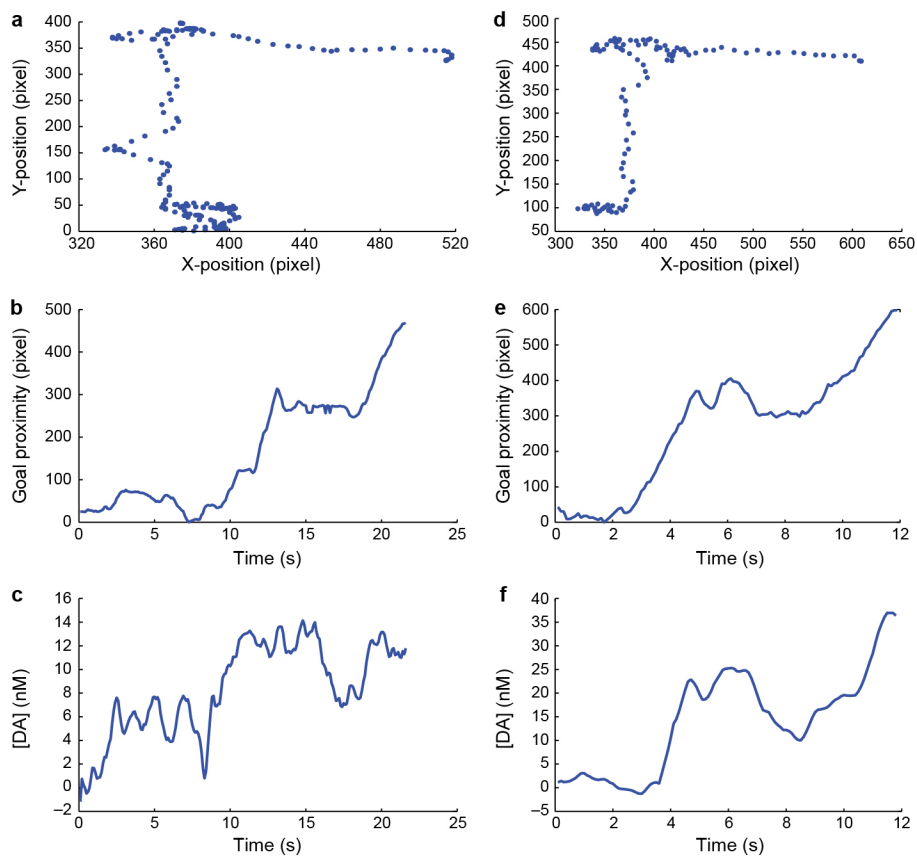


Extended Data Figure 5 | Dopamine signals in response to free reward outside of the task. **a**, Average peak dopamine to unexpected chocolate milk delivery outside the task is positively correlated with peak ramping dopamine measured from the same probes during preceding behavioural training in the maze ($n = 146$ sessions; Pearson's $R = 0.45$, $P < 0.0001$). **b**, Average peak dopamine concentration induced by unexpected free reward outside the task (blue) and peak amplitude of dopamine ramping during maze performance just before reward (red; paired t -test, $P < 0.001$).



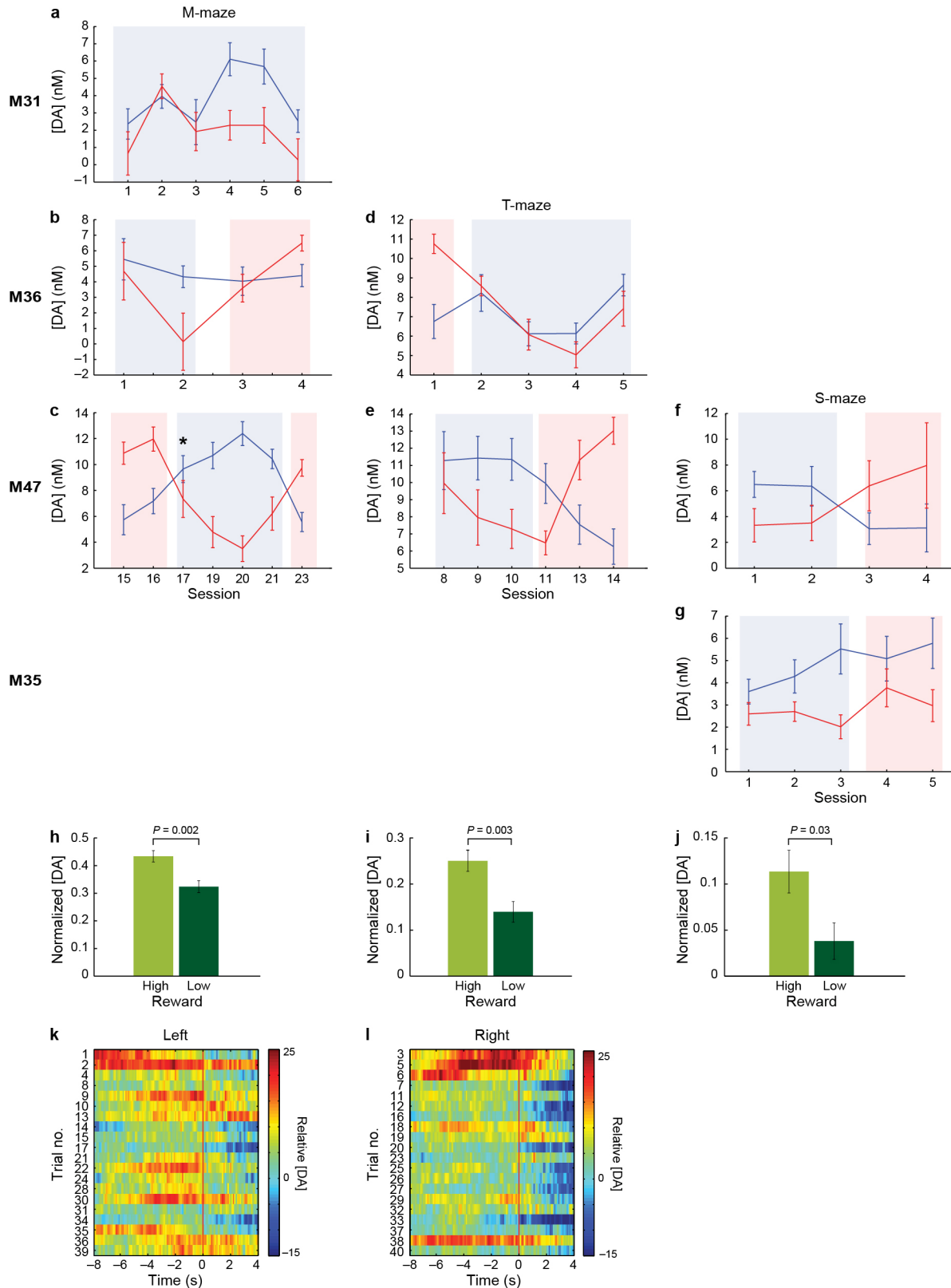
Extended Data Figure 6 | Peak ramping dopamine values do not vary as a function of trial time. **a**, Dopamine release modelled as a summation of four weighted transients in response to fixed maze events on short trials (purple) and long trials (orange). The overlap of the transients is reduced on the long trials, resulting in a lower peak dopamine level at the end of the maze run. Thick lines indicate overall average dopamine, and thin lines indicate the averages of each of the 4 transients across 100 simulation runs. **b**, Relative predicted peak dopamine levels on short and long trials calculated as a linear decay function of trial duration for the simulated model shown in **a** (black), for the spatial proximity model (light blue), and for the actual experimental data (dark blue).

c, An alternative multi-transient model in which the transients (3, inset) are heavily weighted towards the goal location, are highly variable in their time of occurrence, and display a long decay time-course. In this model, the difference between short and long trials is within the noise range of the data. The average of individual transients (inset) across multiple simulations is a smeared version of the single transients that is weighted towards the goal location. **d**, Data plotted as in **b**, for the alternative multi-transient model. **e**, **f**, Average run speed (**e**) and acceleration (**f**) during short (purple) and long (orange) trials, as shown in Fig. 2, for all animals and sessions.



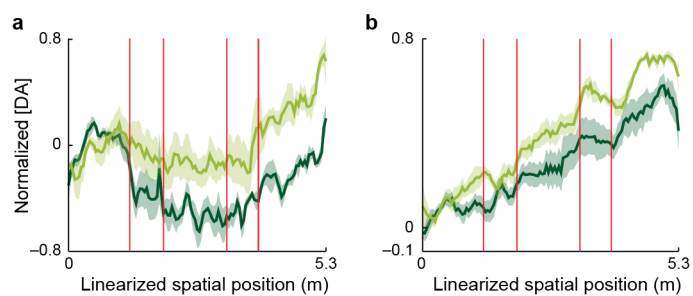
Extended Data Figure 7 | Examples of trials with pausing or exploration near the choice point of the T-maze. **a–c**, Video tracker traces (**a**), relative proximity to reward (**b**), and dopamine concentration (**c**) measured during a

single trial in which a rat paused near the choice point. **d–f**, Video tracker traces (**d**), relative proximity to reward (**e**) and dopamine concentration (**f**) measured during another trial from a different rat.

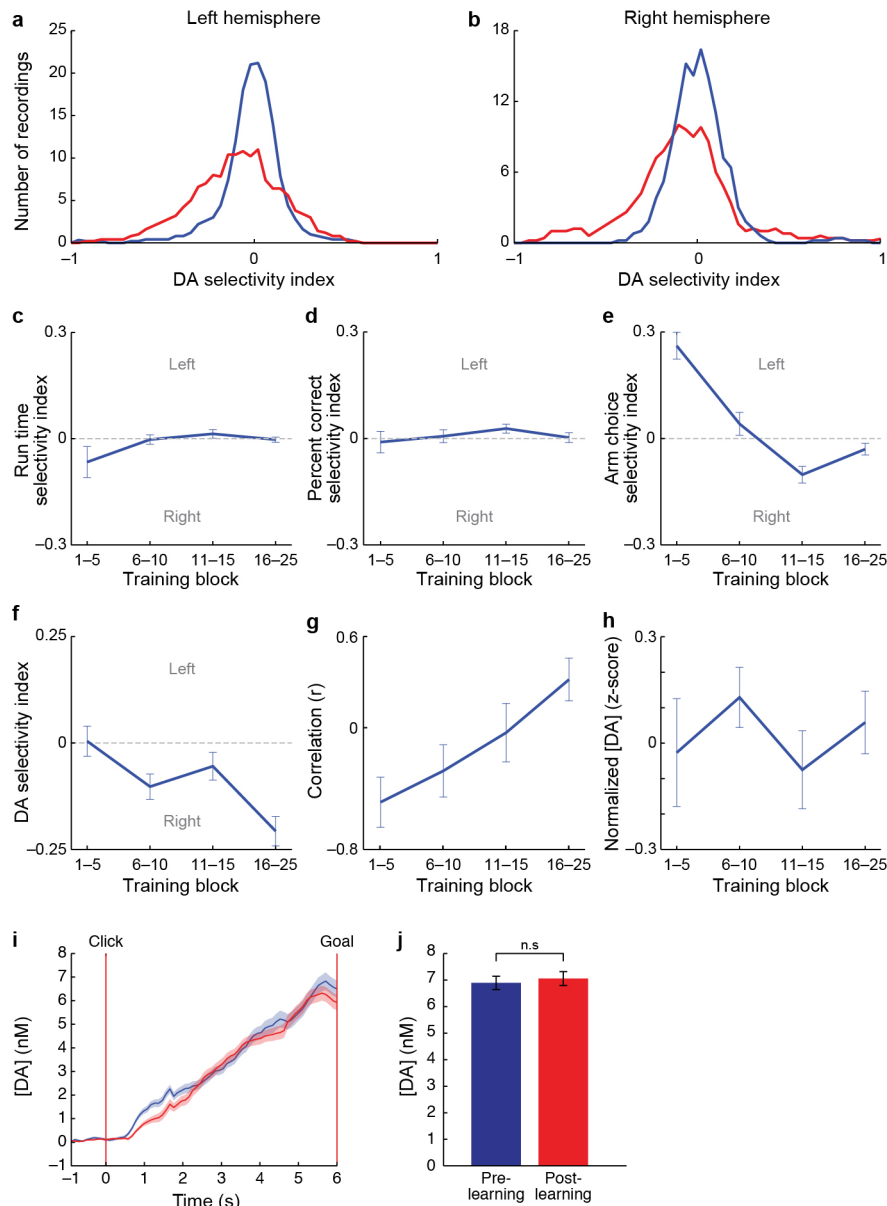


Extended Data Figure 8 | Value manipulation experiments in the T-maze, M-maze, and S-maze for all rats. **a–c**, Average peak dopamine levels for M-maze sessions for three individual rats (M31, M36, and M47) in the left (blue) and right (red) end-arms. Blue and red shading indicates sessions in which left and right arm contained the larger reward, respectively. Error bars indicate s.e.m. **d, e**, Average peak dopamine levels, as in **a–c**, for value bias T-maze sessions for two rats (M36 and M47). **f, g**, Average peak dopamine

levels, as in **a–c**, for value bias S-maze sessions for 2 rats (M47 and M35). **h–j**, Average normalized dopamine levels measured in the high reward arm (light green) and low reward arm (dark green) as rats performed the M-maze (**h**), T-maze (**i**) and S-maze (**j**) tasks. **k, l**, Dopamine concentration relative to left (**k**) and right (**l**) goal-reaching during the first session following a reversal of reward values (session 17 of M47, indicated by an asterisk in **c**).



Extended Data Figure 9 | Dopamine signals recorded in two rats trained on the S-maze task. **a**, Average normalized dopamine concentrations measured from VMS probes in one rat ($n = 5$ sessions) performing the S-maze task as in Fig. 3j. Light green line indicates runs to the higher reward goal, and dark green lines to the lower reward goal. Shading indicates s.e.m. Red vertical lines indicate turns. **b**, Traces, as in **a**, for the second rat trained on the S-maze task ($n = 4$ sessions).



Extended Data Figure 10 | Dopamine end-arm preferences become positively correlated with arm choices with training, but ramping dopamine signals do not change significantly with performance improvement. **a**, **b**, Distribution of selectivity indices, as in Fig. 4d, for all probes implanted in the left (**a**, $n = 5$ rats) and right (**b**, $n = 3$ rats) hemispheres. Note the bias in both groups of the selectivity preference towards negative selectivity indices (right bias, red) relative to the shuffled data (blue). **c–e**, Biases in average run time (**c**), percentage of correct responses (**d**), and arm choices (**e**) across training blocks. Negative values indicate biases towards the right end-arm. **f**, Raw average dopamine selectivity indices across training blocks. Note emergence of right bias with training. **g**, Correlation coefficients (Pearson's R) computed for each training block between arm choice selectivity indices and dopamine

selectivity indices. Error bars indicate confidence limits of the correlations. **h**, Normalized peak magnitudes of dopamine signals averaged in a 0.5-s window before goal-reaching in sessions with significant (Mann-Whitney U -test, $P < 0.05$) pre-goal increases. Data are averaged across rats for each training block (for which, left to right, $n = 48, 101, 113$ and 179 trial averaged recordings, respectively). **i**, Average (\pm s.e.m.) dopamine concentration from ramping dopamine sessions in which percentage of correct trials fell above (red, $n = 179$) or below (blue, $n = 92$) the learning criterion for T-maze task acquisition (72.5% correct, chi-square test, $P < 0.05$). **j**, Average (\pm s.e.m.) peak dopamine levels from the sessions plotted in **i**, showing no significant difference between pre- and post-learning periods (t -test, $P = 0.44$).