Article

Predictive coding of reward in the hippocampus

https://doi.org/10.1038/s41586-025-09958-0

Received: 11 September 2024

Accepted: 24 November 2025

Published online: 14 January 2026

Check for updates

Mohammad Yaghoubi[1,2] ✉, M. Ganesh Kumar[3,4,5], Andres Nieto-Posadas[1], Coralie-Anne Mosser[1], Thomas Gisiger[1], Émmanuel Wilson[1], Cengiz Pehlevan[3,4,5], Sylvain Williams[1,2] & Mark P. Brandon[1,2] ✉

Anticipating future outcomes is a fundamental task of the brain[1–3]. This process requires learning the states of the world as well as the transitional relationships between those states. In rodents, the hippocampal spatial cognitive map is thought to be one such internal model[4]. However, evidence for predictive coding[5,6] and reward sensitivity[7–10] in the hippocampal neuronal representation suggests that its role extends beyond purely spatial representation. How this reward representation evolves over extended experience remains unclear. Here we track the evolution of the hippocampal reward representation over weeks as mice learn to solve a cognitively demanding reward-based task. We find several lines of evidence, both at the population and the single-cell level, indicating that the hippocampal representation becomes predictive of reward as the mouse learns the task over several weeks. Both the population-level encoding of reward and the proportion of reward-tuned neurons decrease with experience. At the same time, the representation of features that precede the reward increases with experience. By tracking reward-tuned neurons over time, we find that their activity gradually shifts from encoding the reward itself to representing preceding task features, indicating that experience drives a backward-shifted reorganization of neural activity to anticipate reward. We show that a temporal difference model of place fields[11] recapitulates these results. Our findings underscore the dynamic nature of hippocampal representations, and highlight their role in learning through the prediction of future outcomes.

The hippocampus represents a mixture of spatial (such as place[12], landmark[13] and so on) and non-spatial (such as time[14], sound frequency[15] and so on) environmental features. The collective encoding of environmental features and their relationships, known as a cognitive map[4], is thought to support spatial navigation and memory-related behaviours. From an evolutionary standpoint, an animal's survival depends on using these cognitive abilities to efficiently learn and remember rewarding experiences, such as navigating to home, safety and food. This computation is supported by an experience-dependent spatial cognitive map in the hippocampus[7,16]. Therefore, the hippocampal representation of the environment is expected to undergo considerable changes once an animal has learned how to navigate to rewarding locations or has mapped environmental features associated with rewards[17].

Previous studies have shown that the hippocampus encodes reward-related events at multiple time points, including reward approach, onset, location and history[7]. During reward approach, running toward a known goal induces place-specific firing patterns along the path that differ from firing during random foraging in the same environment[9]. Place fields also cluster near reward sites, generating a reward over-representation[8,18]. At reward arrival, a distinct group of hippocampal neurons consistently encodes reward delivery, independent of location or context, indicating that reward signals can be separated from place coding[10]. The hippocampus also encodes reward history: after probabilistic reward delivery and after leaving the reward site, neuronal firing changes depending on the reward outcome[19]. Although these studies describe how hippocampal representations change before and after learning reward locations, how these dynamics emerge and evolve with experience over days, weeks or months remains unknown.

The hippocampus has been shown to support predictive models in various species[5,20]. We propose that a reorganization of hippocampal representations—in particular, reward representation—during learning of a reward-based task will occur to support reward prediction. We examine this hypothesis by tracking the evolution of the hippocampal representation across weeks as mice perform a reward-based task.

Calcium imaging of CA1 neurons

We used a one-photon miniaturized head-mounted fluorescent microscope[21] to perform calcium imaging of CA1 of dorsal hippocampus in seven mice (Fig. 1a). Mice were injected with a viral construct to express GCaMP6f in dorsal CA1 and were implanted with a gradient refractive

[1]Douglas Hospital Research Center, McGill University, Montreal, Quebec, Canada. [2]Integrated Program in Neuroscience, McGill University, Montreal, Quebec, Canada. [3]Center for Brain Science, Harvard University, Cambridge, MA, USA. [4]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. [5]Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Cambridge, MA, USA. ✉e-mail: mohammad.yaghoubi@mail.mcgill.ca; mark.brandon@mcgill.ca

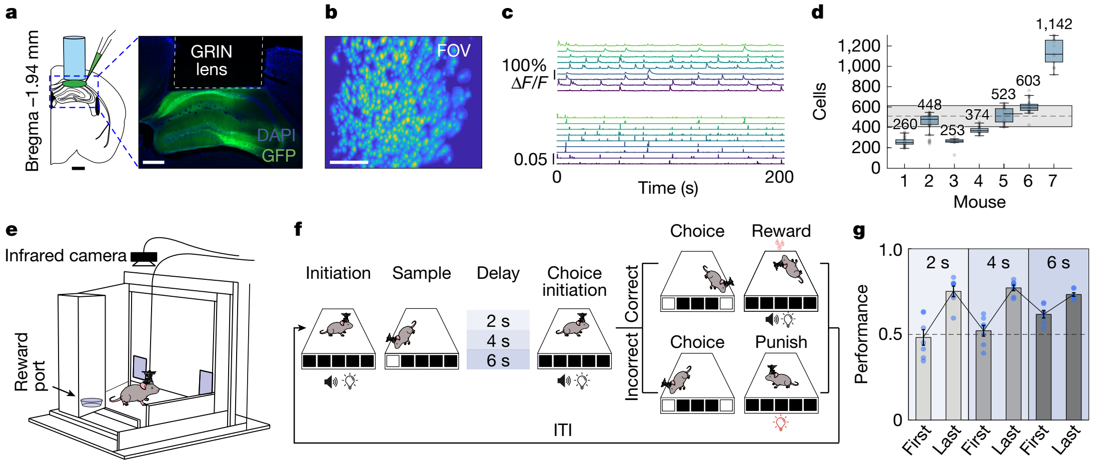

Fig. 1 | Imaging of CA1 neuronal activity in mice while they perform a reward-based task. a, Schematic of the surgical strategy. Right, magnification of the boxed region. Scale bars, 1 mm (left); 200 µm (right). **b**, Example CA1 field of view (FOV) with identified cells. Scale bar, 200 µm. **c**, Extracted calcium traces of nine representative cells (top rows) and their corresponding deconvolved traces (bottom rows). **d**, Number of identified cells across mice (number of sessions per mouse from left to right: 33, 53, 17, 18, 24, 23, 13). Box plots show median (centre line), 25th–75th percentiles (box) and range within 1.5 × interquartile range (IQR; whiskers); points beyond whiskers are outliers. The dashed line and the shading represent the mean ± s.d. (504 ± 101) of the number of cells. **e**, Schematic of the touchscreen chamber. **f**, Schematic of the task. **g**, Mouse performance for the first and last day for each delay ($n = 7$ mice). Bar graph and error bars show mean ± s.e.m. Dashed line shows the chance level. The schematic in **a** was created using CorelDRAW; illustrations in **e**,**f** were created using Affinity Designer.

index (GRIN) lens targeting CA1 (see 'Surgeries' in Methods). Calcium recording data were preprocessed to correct for motion artefacts[22], segment cells, extract calcium transients[23] and deconvolve the traces (Fig. 1b,c and Extended Data Fig. 1). We recorded 504 ± 101 (mean ± s.d.) neurons across sessions and mice (Fig. 1d). We used a 20 × 18-cm automated touchscreen recording box[24,25] to monitor mouse behaviour. The box consists of a touchscreen in front, a reward port in the back and an infrared camera on top to record behaviour (Fig. 1e). Mice were trained on a delayed non-matching-to-location task, where a sample appeared randomly on the left or right screen after trial initiation. After a nose poke to the sample, the delay starts. At the end of the delay, a tone and light cue signalled the mouse to move to the back of the cage and break a beam to initiate the choice phase. During the choice phase, two white squares are displayed and mice must choose the non-matching square to receive the reward (Fig. 1e,f). Mice performed one session per day. When the mouse reached a high level of performance, we increased the delay between the sample and the choice phase to make the task more challenging. This served two purposes: (1) to separate the effects of experience (session number) from learning (performance) on hippocampal activity; and (2) to keep mice continuously learning, engaging relevant neuronal circuits throughout recordings. Mice exhibited increased performance, for each delay duration, over time (Fig. 1g and Extended Data Fig. 1). The following sections focus on the encoding of reward. An extended analysis of spatial tuning and decoding reveals that hippocampal neurons are involved in representing multiple aspects of the task (Extended Data Figs. 2–4).

Reward encoding decreases with experience

We investigated the dynamics of reward representation in the hippocampus as mice learn to solve the delayed non-match-to-location task. The learning period varied, taking a few weeks depending on each mouse's learning rate (Extended Data Fig. 1). To quantify the reward-encoding signal across sessions, we measured reward information at the population level using an information-theoretic analysis in the CEBRA-derived latent space (see 'CEBRA embedding' in Methods). This framework enabled us to track reward representation changes with experience. At the single-cell level, we used a shuffle-control approach to identify reward cells per session and tracked their percentage across days. Both analyses indicate that reward representation declines mainly with experience, not performance.

Our data suggest that a dedicated subpopulation of cells is responsive to reward (Fig. 2a). Notably, distinct subpopulations encode the reward depending on whether the mouse approached the reward from the left or right choice on the touchscreen. The sorted calcium traces show that reward neurons are not necessarily tuned to the reward onset[10] but form a reliable sequence spanning the entire duration of reward consumption (Fig. 2a and Extended Data Fig. 4).

Using CEBRA[26], we projected our deconvolved calcium traces into a 32-dimensional latent space. To quantify the information content of the reward representation, we used a fivefold cross-validation approach to decode the reward moments from latent space. The cross-fold-averaged mutual information (MI) between the decoded reward traces and the actual reward traces was regarded as the reward information content for each session (Fig. 2b and Extended Data Fig. 5). Correlating reward information content (we call it reward MI) with session number (day) and mouse performance indicates a negative correlation with session number and a weak correlation with mouse performance (Fig. 2b). The result is consistent across mice (Fig. 2c). A linear model (see 'Linear modelling of information content' in Methods) showed that variance in reward MI is explained mainly by experience, not by performance (Fig. 2c).

At the single-cell level, we used a shuffle-control procedure to identify reward cells (see 'Identification of cell types' in Methods) (Fig. 2d). This resulted in 8.5 ± 1.5% of the cells being identified as reward cells (Fig. 2e). Reward-cell tuning curves show two features: (1) responses depend on the mouse's approach direction to the reward port; and (2) cells are tuned to distinct moments of reward consumption, extending beyond reward onset (Fig. 2d and Extended Data Fig. 4). Furthermore, reward cells exhibit greater firing during task engagement than during inter-trial intervals (ITIs) (Extended Data Fig. 4). Notably, consistent with the population-level analysis, the percentage of reward cells declined with session number but showed only a weak correlation with performance (Fig. 2f), a pattern consistent across mice (Fig. 2g). A linear model indicates that a significant amount of the variance in the dynamics of reward-cell recruitment is attributed to session number rather than to mouse performance (Fig. 3g). Both population and single-cell level analysis reveal that the reward representation decreases with

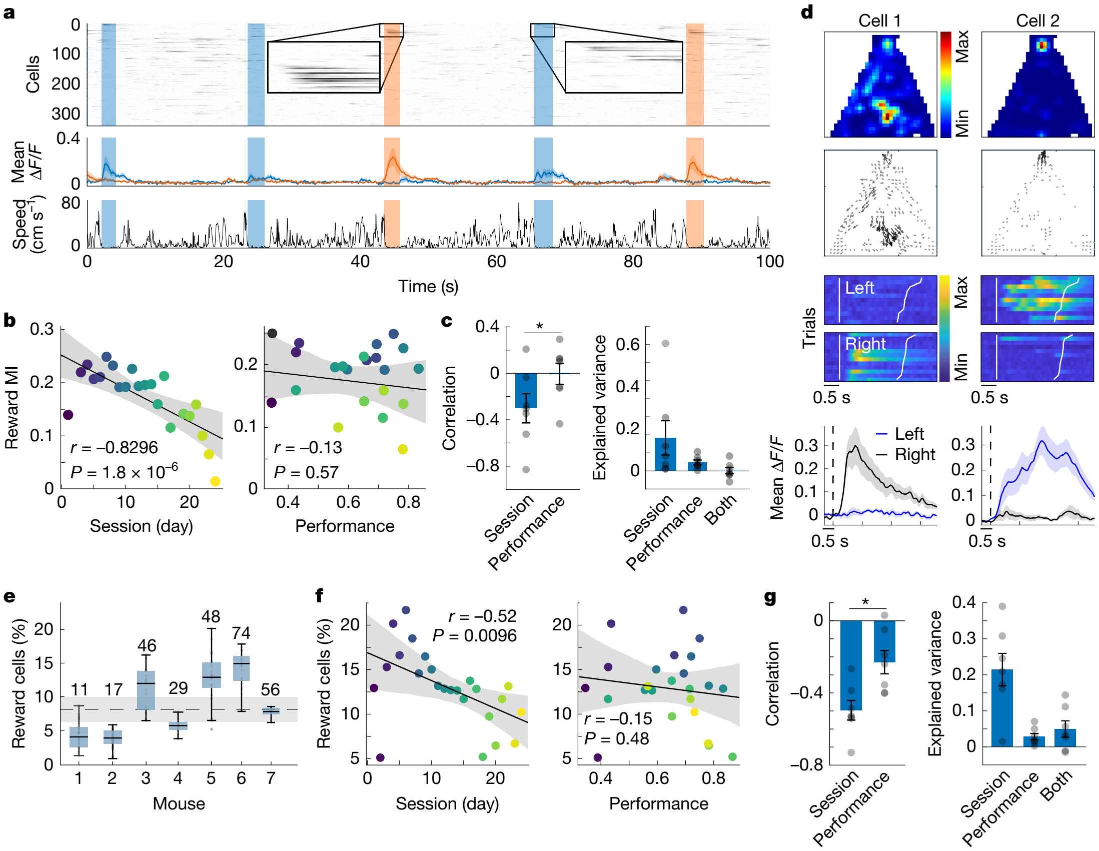

Fig. 2 | Dynamics of reward encoding during learning. a, Top, raw calcium traces from 350 neurons in one session. Blue and orange bands mark reward consumption time after right (blue) and left (orange) approaches. First 45 reward-responsive cells are sorted by peak activity; insets show reward-cell responses for right (blue, cells 1–20) and left (orange, cells 21–45) approaches. Middle, averaged activity of left- and right-preferring reward cells. Bottom, mouse speed. **b**, Reward mutual information (MI) declines with session number ($r = -0.83$, two-sided *t*-test $P = 0.0001$) but shows weak correlation with performance ($r = -0.13$, two-sided *t*-test $P = 0.57$). **c**, The effect is consistent across mice ($n = 7$) (two-sided Wilcoxon signed-rank, $P = 0.0469$); linear modelling confirms that session number is the dominant factor explaining reward MI dynamics. **d**, Tuning of two reward cells. First row, place fields. Second row, vectorized place fields. Third row, trial-by-trial calcium activity aligned to reward onset. First and second white lines indicate the start and offset of reward consumption, respectively. Trials are sorted by reward consumption duration. Fourth row, average calcium traces across trials (mean ± s.e.m.). **e**, Percentage of identified reward cells across mice. The numbers for each mouse show the cross-session average number of reward cells. The dashed line and the shading represent the mean number of reward cells ± s.e.m. (8.5 ± 1.5%) across mice. Box plots show median (centre line), 25th–75th percentiles (box) and range within 1.5 × IQR (whiskers); points beyond whiskers are outliers. **f**, Percentage of reward cells decreases with session number ($r = -0.52$, two-sided *t*-test $P = 0.0096$) but not with performance ($r = -0.15$, two-sided *t*-test $P = 0.48$). **g**, The effect is consistent across mice ($n = 7$) (two-sided Wilcoxon signed-rank, $P = 0.0312$). Linear modelling confirms that session number is the main factor explaining variance in reward-cell recruitment. Bar graphs and error bars in **c**,**g** show mean ± s.e.m. In **b**,**f**, the solid line shows the linear regression fit (least-squares) and the shaded error band represents the 95% confidence interval.

experience. To rule out potential preprocessing effects, we identified reward cells using both deconvolved traces and the area under the curve (AUC) of raw calcium signals, finding similar dynamics in reward MI and cell recruitment (Extended Data Fig. 6). Additional analyses confirmed that the decline in reward representation was not due to task difficulty (delay length) or behaviour variability (running speed before reward) (Extended Data Fig. 7). The gradual reorganization of reward representation prompted us to examine hippocampal dynamics for other task features, such as the pre-reward epoch.

Pre-reward encoding increases with learning

In this section, we apply the same methodology used to measure hippocampal reward representation to quantify the evolution of hippocampal encoding of pre-reward moments. Specifically, we analysed two pre-reward events: (1) screen, the [−150, 150]-ms window around the choice touch; and (2) reward approach, the interval between a choice and a reward as the mouse runs to the port. Using the same methods as for reward, we assessed both population- and single-cell-level encoding of these events to track how their representations evolve with time.

Distinct neuron subpopulations encoded left versus right choices at the touchscreen (Fig. 3a). We applied the same analysis used for reward to measure population-level screen information content. In contrast to that observed for reward, screen information increased with both session number and mouse performance (Fig. 3b). A linear model indicates that both factors contribute significantly to explaining the variance in the dynamics of the screen information content (Fig. 3b). A similar analysis for reward-approach encoding indicates a similar positive correlation for the reward-approach information content (Fig. 3c,d).

At the single-cell level, screen and reward-approach cells were identified using a shuffle-control procedure (see 'Identification of cell types'

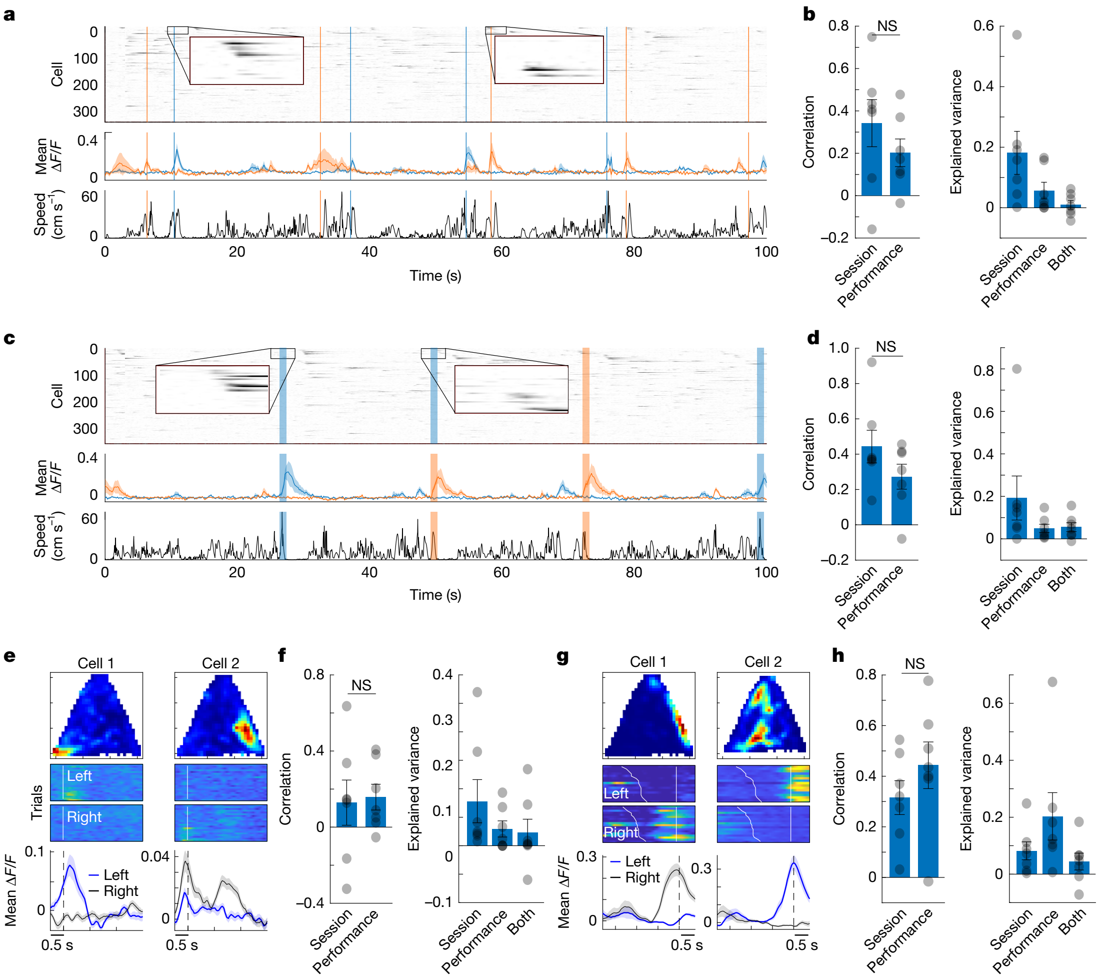

Fig. 3 | Dynamics of pre-reward encoding across learning. a, Top, raw calcium traces. Blue and orange lines indicate poke moments to the right (blue) and left (orange) screens. Middle, average calcium activity for left- and right-screen-responsive cells. Bottom, mouse speed. **b**, Screen MI positively correlates with both experience and performance ($n = 7$ mice). Two-sided Wilcoxon signed-rank test $P = 0.4688$ (NS, not significant). A linear model shows that both session number and performance have key roles in explaining the dynamics of screen MI. **c**,**d**, Analysis as in **a**,**b**, but for reward-approach moments. Reward-approach MI positively correlates with both experience and performance ($n = 7$ mice). Two-sided Wilcoxon signed-rank test $P = 0.2188$ (NS). **e**, Tuning curves of two screen cells. Top row, place fields. Middle row, trial-by-trial calcium activity for left (top) and right (bottom) screen choice. White line indicates screen poke time. Bottom row, average calcium traces (mean ± s.e.m.) for left (blue) and right (grey) screen pokes. **f**, The percentage of screen cells increases with both session number and performance ($n = 7$ mice). Two-sided Wilcoxon signed-rank test $P = 1$ (NS). A linear model shows that both session number and performance significantly explain changes in screen cell recruitment. **g**,**h**, Analysis as in **e**,**f**, but for reward-approach cells. In the middle row in **g**, the left white line shows the screen poke, and the right white line indicates the reward onset. Like screen cells, reward-approach cells increase with session number and performance. In **h**, two-sided Wilcoxon signed-rank test $P = 0.2188$ (NS). Bar graphs and error bars in **b**,**d**,**f**,**h** show mean ± s.e.m.

in Methods). We identified 7.5 ± 0.7% of the cells as screen cells and 5.7 ± 0.7% as reward-approach cells (Extended Data Fig. 5). The percentage of identified cells for both screen and reward-approach cells shows a positive correlation with both session number and mouse performance (Fig. 5b,e). A linear model reveals that both session number and performance contribute significantly to the dynamics of recruitment of screen and reward-approach cells (Fig. 5c,f). Finally, we compared calcium response amplitudes of reward-approach cells during approaches to the main reward versus the smaller incentive given during the delay, and found that reward magnitude modulated their activity significantly (Extended Data Fig. 8).

Together, these results show distinct dynamics: with experience, measures of reward encoding decline at both population and single-cell levels, whereas measures of screen and reward-approach encoding increase.

Backward shift of reward coding during learning

Across all mice, we were able to track 1,814 neurons (see ‘Tracking cells’ in Methods and Extended Data Fig. 9). Out of 1,814 cells, 225 were reward cells (12.4%), 225 were screen cells (12.4%) and 53 were reward-approach cells (2.9%). The remaining 1,311 cells (72.3%) are labelled as non-classified cells. Next, we examine the functional properties and evolution of these cells across sessions.

Our data reveal that a significant number of reward cells exhibit a backward shift across sessions from reward to the reward approach and screen, termed as backward-shifting reward cells (Fig. 4a–c and Extended Data Figs. 10, 12 and 13). Specifically, we report a significant negative correlation between the response timing and the session number for reward cells and reward-approach cells (Fig. 4d). Using a shuffle-control method, we found that 21% (47 out of 225) of

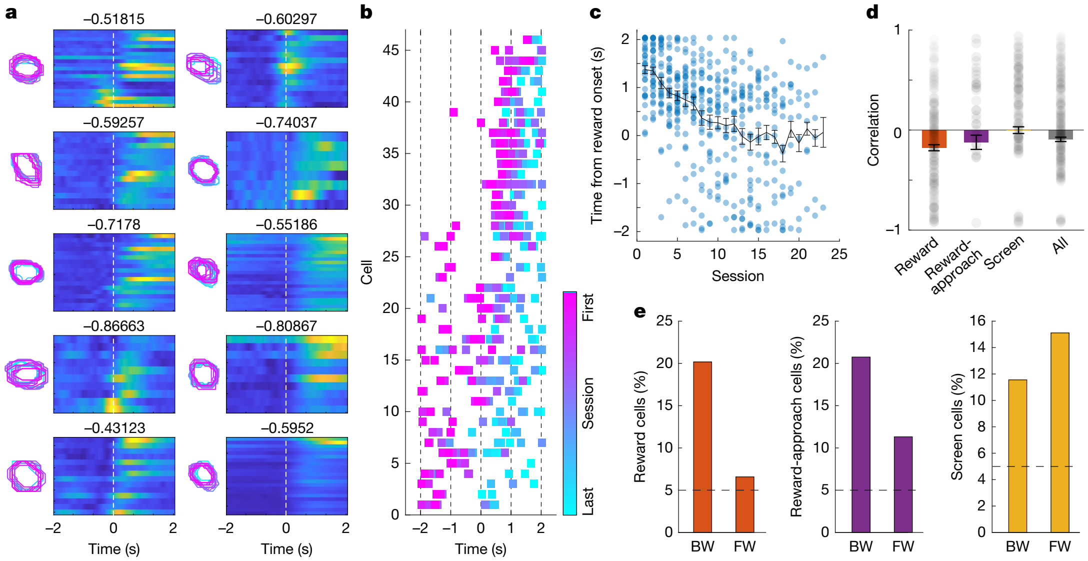

Fig. 4 | Weeks-long backward shift of reward encoding during learning. **a**, Ten representative backward-shifting reward cells. Rows show average calcium activity at reward from first (top) to last (bottom) session. Numbers indicate the correlation between peak timing and session number. Overlapping contours of cell bodies across sessions are shown on the left. **b**, Peak activity timing relative to reward onset across sessions for all backward-shifting reward cells, sorted by mean peak timing. **c**, Each point represents the timing of peak activity for a single backward-shifting reward cell in a given session. Scatter plot shows 47 backward-shifting reward cells (each tracked across an average of 10 sessions; total $n = 473$ data points). Black line indicates the within-session average of peak times, and error bars are s.e.m. **d**, Correlation between peak activity timing and session number across all tracked cells for each cell type ($n = 228$ reward cells, 53 reward-approach cells, 225 screen cells and 1,308 non-classified cells). Bar graphs and error bars show mean ± s.e.m. **e**, Proportion of cells identified as forward-shifting (FW) and backward-shifting (BW) for each cell type. Dashed line indicates chance level.

tracked reward cells showed backward shifting—well above the 5% chance level (Fig. 4e). A substantial portion (60%; 28 out of 47 cells) of backward-shifting reward cells shifted enough to be classified as screen or reward-approach cells in later sessions (Fig. 4b). Detection of forward-shifting reward cells was at chance levels (Fig. 4e). Unlike reward cells, we found that screen and reward-approach cells exhibited a mixture of backward and forward shifting (Fig. 4e).

We also examined whether neuron response amplitudes changed across sessions. Using a similar approach to that used for temporal shifts, we correlated each neuron's peak amplitude with session number instead of peak timing. Many neurons across all cell types show declining activity over sessions, suggesting that, alongside backward shifts, reduced firing of some reward cells contributes to the population-level decrease in reward representation (Extended Data Figs. 11, 12 and 14).

A TD error model recapitulates the backward shift

The marked similarity of the backward-shifting reward cells to the reward prediction error (RPE) response observed in midbrain dopamine neurons motivated us to see whether a temporal difference (TD) learning model of the hippocampal representation could explain our observations. We focused on the segment from choice at the screen to reward, modelling it as a one-dimensional (1D) navigation task: the agent moves from state 1 (choice at screen) through to state 7 (reward port nose poke) and receives a reward at terminal state 8 (Fig. 5a,b).

In our model, at initiation, 1,000 place cells uniformly tile the 1D state space with each cell's state selectivity described by a Gaussian radial basis function. The place-cell population activity is passed to a critic (v) for value estimation and TD (δ) computation. The objective is to minimize the TD error by updating both the value function and place-cell peaks (see 'Simulations for TD-error-modulated place-cell model' in Methods). This causes backward shifting of TD error from the reward to the start state (Fig. 5b,c), driving backward updates in state-value estimates and correspondingly backward shifts in place-cell peaks (Fig. 5d–i) (see model details in Supplementary Information and 'Simulations for TD-error-modulated place-cell model' in Methods). Three main reorganization patterns appear: (1) reward-proximal cells shift monotonically backward; (2) reward-approach cells first move towards the reward, then shift backwards; and (3) screen-proximal cells shift forwards late in learning (Fig. 5f). In addition, we extended the model to a policy-learning agent, in which place cells evolve to maximize rewards, mirroring animal behaviour. Despite the added complexity (Extended Data Fig. 15), spatial selectivity still shifts as in Fig. 5. In the early learning phase, the model replicates the over-representation of the reward state by place cells[8,27], as observed in previous experiments (Fig. 5j), and in the later phase, a gradual decrease is observed, consistent with our experimental results (Fig. 2f).

Our modelling underscores the crucial role of reward predictability. Specifically, the backward shift is seen only when the reward discount factor (γ), which determines the influence of future state values in the TD error calculation, is greater than 0.1 (Fig. 5k). When γ is less than 0.1, place cells remain over-represented at the reward without shifting backwards. This indicates that incorporating future state-value estimates into the TD error is essential for driving the backward shift, supporting the idea that a RPE-like signal underlies the dynamics observed in our experiments.

Discussion

We combined large population recordings[21] of mouse CA1 neurons with an automated touchscreen reward-based task[24,25] to investigate the long-term dynamics of reward encoding in the hippocampus. Our data revealed a reduction in reward signal and an increase in the response to the cues that anticipate the reward. This was further supported by tracking individual cells that are at first tuned to the reward and gradually shift backwards to encode aspects of the task that are reward predictors. This backward shift in coding can be explained by a temporal difference reinforcement learning (TDRL) model of

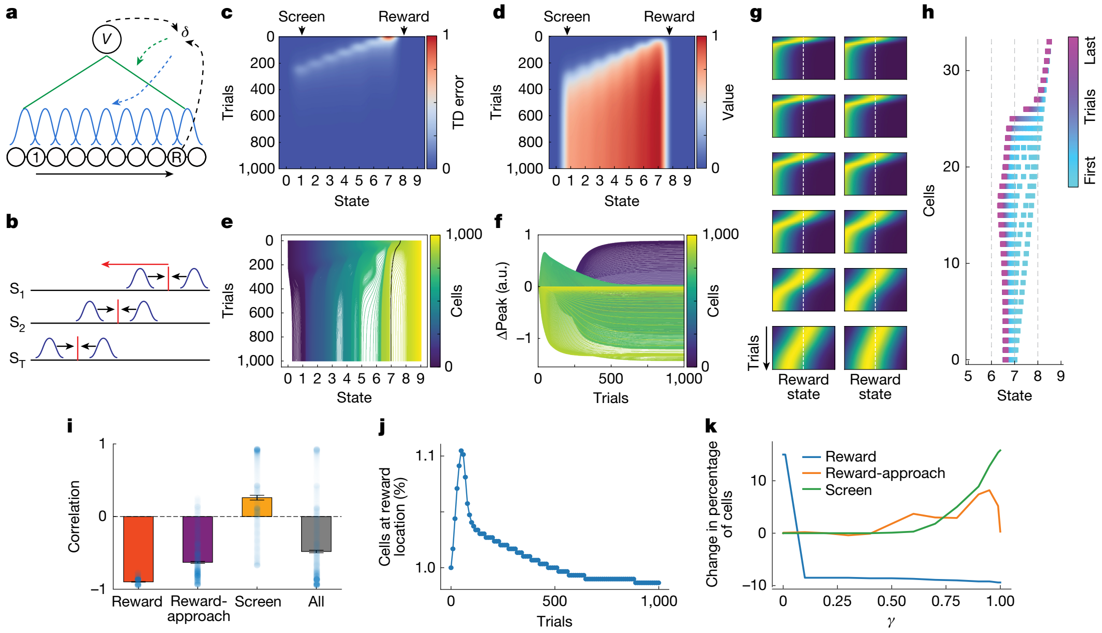

Fig. 5 | TD error drives backward shifting of place fields. a, An agent follows a fixed path to a terminal reward state 8. Place cells (Gaussian basis functions) project to a critic computing value ($v(s)$) and TD error (δ), which updates both value estimates and place-field peaks. **b**, TD error (red vertical line) causes place fields to shift towards its location; as it propagates backwards from the reward, fields shift progressively earlier. S1, S2, and ST denote session 1, session 2, and session T, respectively, and are used to illustrate session-by-session shifts in place fields. **c**–**f**, Place-cell dynamics for an agent that minimizes value estimation. **c**, TD error propagates backwards from state 7 to state 1 across learning trials. **d**, Value estimates update accordingly. **e**, Place fields, uniform at first, over-represent the reward then shift backwards (black line represents the average peak of reward-coding cells). **f**, Three shift patterns: (1) reward cells (yellow) shift backwards; (2) approach cells (green) shift forwards then backwards; (3) screen cells (purple) shift forwards later. a.u., arbitrary units. **g**, Peak trajectories of 12 example reward cells show consistent backward shifts (dashed line represents reward). **h**, Cells near reward exhibit reliable backward shifts. **i**, Correlation between peak timing and session number replicates (Fig. 4d) ($n = 280$ reward cells, 443 reward-approach cells, 258 screen cells, 1,000 all cells). Bars and error bars represent mean ± s.e.m. **j**, The number of cells in the reward zone increases rapidly during early trials, then declines gradually with continued experience. **k**, The change in percentage of cells indicates the difference in the number of cells at the reward, approach and screen states between trial 0 and trial 1,000. For higher discount factors we observe a reduced reward-cell prevalence and an increase in approach and screen cells. The figure shows the dynamics of one agent, because there is no stochasticity in learning value estimation.

hippocampal place fields. These results highlight a dynamic reorganization of hippocampal representations that supports learning by gradually shifting its coding toward cues that best predict future reward. Previous studies have revealed that hippocampal place fields move towards goal locations early in learning, probably contributing to what others have observed as an over-representation of rewarded locations[8,27]. We also observe an over-representation of the reward location early in our recordings (Extended Data Fig. 3). Other work has shown a backward skew of hippocampal place fields, independent of reward locations, on a faster, within-session timescale[28]. Together, these outcomes suggest that the hippocampal representation over-represents rewarded locations at first, and that this is followed by a slower, weeks-long shift to represent the cues that predict these rewards. Notably, our TD model also over-represents reward at first (Fig. 5k), followed by a backward shift of reward-tuned cells with experience (Fig. 5e–i).

The dynamics observed in our CA1 data mirror those of the dopaminergic output of the ventral tegmental area (VTA). This system is central to reward learning by RPE[29–31], as formalized by TDRL[31–35]. TDRL has profoundly shaped our understanding of dopaminergic reward coding, a concept that has also influenced our understanding of hippocampal physiology[36]. Prevalent implementations of TDRL make two key predictions: (1) a gradual decrease in reward response coupled with a gradual increase in response to reward-predicting cues during learning[1,37]; and (2) a gradual backward temporal shift of the error signal from reward to cues during learning[38]. Both of these are well-documented in dopamine neurons, and are also evident in our data. This resemblance suggests that the dynamics of hippocampal reward representations emerges from interactions within a broader circuit involving the hippocampus and VTA.

The model presented here extends TDRL by using Gaussian basis functions as spatial features, which reorganize through the TD error to improve state-value estimation and policy learning for reward maximization[11]. Because these functions are modulated by the backward-shifting TD error, the resulting place fields also shift backwards from the reward. The successor representation algorithm also exhibits a backward shifting of fields in the presence of a reward[11], although it tends to maintain or increase field density at the reward location, which differs from the decrease we observe in our experimental data. Although a TDRL model captures key aspects of the observed dynamics, future work could consider alternative predictive coding objectives[39–41] and develop more biologically plausible[40,42–46] hippocampus–dopamine models beyond backpropagation-based implementations.

In conclusion, our study uncovers a dynamic and organized backward shift of the hippocampal reward representation during extended experience. Far from serving as a stationary spatial map, the hippocampus exhibits predictive coding, progressively tuning its representation to anticipate future rewards. These insights advance our understanding of the role of the hippocampus in learning, highlighting its crucial contribution to the brain's overarching objective of forecasting and optimizing future rewards.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09958-0.

1. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
2. Keller, G. B. & Mrsic-Flogel, T. D. Predictive processing: a canonical cortical computation. *Neuron* **100**, 424–435 (2018).
3. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
4. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon Press, 1978).
5. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017).
6. Levenstein, D., Efremov, A., Eyono, R. H., Peyrache, A. & Richards, B. Sequential predictive learning is a unifying theory for hippocampal representation and replay. Preprint at *bioRxiv* https://doi.org/10.1101/2024.04.28.591528 (2024).
7. Sosa, M. & Giocomo, L. M. Navigating for reward. *Nat. Rev. Neurosci.* **22**, 472–487 (2021).
8. Lee, I., Griffin, A. L., Zilli, E. A., Eichenbaum, H. & Hasselmo, M. E. Gradual translocation of spatial correlates of neuronal firing in the hippocampus toward prospective reward locations. *Neuron* **51**, 639–650 (2006).
9. Aoki, Y., Igata, H., Ikegaya, Y. & Sasaki, T. The integration of goal-directed signals onto spatial maps of hippocampal place cells. *Cell Rep.* **27**, 1516–1527 (2019).
10. Gauthier, J. L. & Tank, D. W. A dedicated population for reward coding in the hippocampus. *Neuron* **99**, 179–193 (2018).
11. Kumar, M. G., Bordelon, B., Zavatone-Veth, J. A. & Pehlevan, C. A model of place field reorganization during reward maximization. *Proc. Mach. Learn. Res.* **267**, 31892–31929 (2025).
12. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
13. Deshmukh, S. S. & Knierim, J. J. Influence of local objects on hippocampal representations: landmark vectors and memory. *Hippocampus* **23**, 253–267 (2013).
14. Kraus, B. J., Robinson, R. J. 2nd, White, J. A., Eichenbaum, H. & Hasselmo, M. E. Hippocampal 'time cells': time versus path integration. *Neuron* **78**, 1090–1101 (2013).
15. Aronov, D., Nevers, R. & Tank, D. W. Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit. *Nature* **543**, 719–722 (2017).
16. Sosa, M., Plitt, M. H. & Giocomo, L. M. A flexible hippocampal population code for experience relative to reward. *Nat. Neurosci.* **28**, 1497–1509 (2025).
17. Kaufman, A. M., Geiller, T. & Losonczy, A. A role for the locus coeruleus in hippocampal CA1 place cell reorganization during spatial reward learning. *Neuron* **105**, 1018–1026 (2020).
18. Dupret, D., O'Neill, J., Pleydell-Bouverie, B. & Csicsvari, J. The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nat. Neurosci.* **13**, 995–1002 (2010).
19. Lee, S.-H. et al. Neural signals related to outcome evaluation are stronger in CA1 than CA3. *Front. Neural Circuits* **11**, 40 (2017).
20. Lisman, J. & Redish, A. D. Prediction, sequences and the hippocampus. *Philos. Trans. R. Soc. B* **364**, 1193–1201 (2009).
21. Aharoni, D. & Hoogland, T. M. Circuit investigations with open-source miniaturized microscopes: past, present and future. *Front. Cell. Neurosci.* **13**, 141 (2019).
22. Pnevmatikakis, E. A. & Giovannucci, A. NoRMCorre: an online algorithm for piecewise rigid motion correction of calcium imaging data. *J. Neurosci. Methods* **291**, 83–94 (2017).
23. Zhou, P. et al. Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data. *eLife* **7**, e28728 (2018).
24. Mosser, C.-A. et al. The McGill-Mouse-Miniscope platform: a standardized approach for high-throughput imaging of neuronal dynamics during behavior. *Genes Brain Behav.* **20**, e12686 (2021).
25. Bussey, T. J. et al. New translational assays for preclinical modelling of cognition in schizophrenia: the touchscreen testing method for mice and rats. *Neuropharmacology* **62**, 1191–1203 (2012).
26. Schneider, S., Lee, J. H. & Mathis, M. W. Learnable latent embeddings for joint behavioural and neural analysis. *Nature* **617**, 360–368 (2023).
27. Xu, H., Baracskay, P., O'Neill, J. & Csicsvari, J. Assembly responses of hippocampal CA1 place cells predict learned behavior in goal-directed spatial tasks on the radial eight-arm maze. *Neuron* **101**, 119–132.e4 (2019).
28. Mehta, M. R., Barnes, C. A. & McNaughton, B. L. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proc. Natl Acad. Sci. USA* **94**, 8918–8921 (1997).
29. Berke, J. D. What does dopamine mean?. *Nat. Neurosci.* **21**, 787–793 (2018).
30. Glimcher, P. W. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proc. Natl Acad. Sci. USA* **108**, 15647–15654 (2011).
31. Watabe-Uchida, M., Eshel, N. & Uchida, N. Neural circuitry of reward prediction error. *Annu. Rev. Neurosci.* **40**, 373–394 (2017).
32. Dayan, P. Improving generalization for temporal difference learning: the successor representation. *Neural Comput.* **5**, 613–624 (1993).
33. Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A. & Sederberg, P. B. The successor representation and temporal context. *Neural Comput.* **24**, 1553–1568 (2012).
34. Maes, E. J. P. et al. Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors. *Nat. Neurosci.* **23**, 176–178 (2020).
35. Kim, H. R. et al. A unified framework for dopamine signals across timescales. *Cell* **183**, 1600–1616 (2020).
36. Lisman, J. E. & Grace, A. A. The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron* **46**, 703–713 (2005).
37. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
38. Amo, R. et al. A gradual temporal shift of dopamine responses mirrors the progression of temporal difference error in machine learning. *Nat. Neurosci.* **25**, 1082–1092 (2022).
39. Foster, D. J., Morris, R. G. & Dayan, P. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* **10**, 1–16 (2000).
40. Kumar, M. G., Tan, C., Libedinsky, C., Yen, S.-C. & Tan, A. Y.-Y. One-shot learning of paired association navigation with biologically plausible schemas. Preprint at https://doi.org/10.48550/arXiv.2106.03580 (2021).
41. Fang, C. & Stachenfeld, K. L. Predictive auxiliary objectives in deep RL mimic learning in the brain. In *The 12 International Conference on Learning Representations* (ICLR, 2024).
42. Lillicrap, T. P., Cownden, D., Tweed, D. B. & Akerman, C. J. Random synaptic feedback weights support error backpropagation for deep learning. *Nat. Commun.* **7**, 13276 (2016).
43. Miconi, T. Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks. *eLife* **6**, e20899 (2017).
44. Murray, J. M. Local online learning in recurrent networks with random feedback. *eLife* **8**, e43299 (2019).
45. Nøkland, A. Direct feedback alignment provides learning in deep neural networks. In *Proc. 30th International Conference on Neural Information Processing Systems* 1045–1053 (NIPS, 2016).
46. Overwiening, J., Kumar, M. G. & Sompolinsky, H. TeDFA-δ: Temporal integration in deep spiking networks trained with feedback alignment improves policy learning. In *8th Annual Conference on Cognitive Computational Neuroscience* (CCM, 2025).

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

Methods

Mice

Eight naive male mice (C57BL/6 mice, Charles River) were housed individually and maintained under a 12-h light–dark cycle at 22 °C and 40% humidity with water ad libitum. Owing to signs of infection observed during the behavioural testing phase, one of the mice that had been recorded for seven days was excluded from the analysis. The infection was considered to be likely to affect task performance and behavioural outcomes, rendering the data unreliable. All experiments were performed during the light part of the light–dark cycle and were in accordance with the guidelines of the McGill University and Douglas Hospital Research Center Animal Use and Care Committees (protocol 2015–7725) and with Canadian Institutes of Health Research guidelines.

Surgeries

Mice underwent three surgeries under isoflurane (1.5–2%, v/v). In addition, carprofen (10 mg kg^{-1}) and saline (0.5 ml) were administered at the beginning of each surgery. We injected 400 nl of either AAV9.syn.GCaMP6f.WPRE.SV40 (University of Pennsylvania Vector Core, 3.26×10^{14} genome copies per ml) diluted 1:1 with PBS or AAV5-CaMKII-GCaMP6f.WPRE.SV40 (Addgene, 2.3×10^{13} genome copies per ml)) diluted with PBS 1:2 into dorsal CA1 (−1.8 mm from Bregma, 1.5 mm mediolateral, 1.45 mm dorsoventral). Two weeks after viral injections, a GRIN lens (Edmund Optics, 1.8 mm in diameter, 0.25 pitch, 4.31 mm in length) was implanted above the previous injection site. In brief, a 1.8-mm craniotomy above the injection site was done followed by aspiration of cortical tissue directly below the craniotomy. The GRIN lens was lowered to the area of interest and two stainless-steel screws were threaded into the contralateral skull. Both the GRIN lens and the screws were fixed with dental cement (C&B Metabond). Silicone adhesive was used to cover the lens until the next surgery. Two to three weeks after the GRIN lens implantation, an aluminium baseplate was attached with dental cement to the mouse's skull and covered with a plastic cap to protect the lens.

Apparatus

Mice were trained in the Bussey-Saksida automated touchscreen operant chamber (Lafayette Instruments)[25,47]. In brief, this trapezoidal-shaped apparatus features a touch-sensitive LCD computer monitor (12.1-inch screen, 800 × 600 resolution) at one end and a reward collection magazine (20 cm height × 18 cm length × 6–24 cm width) at the other, tapering from the touchscreen to the magazine. The arena's walls are made of black Perspex, and the floor is perforated stainless steel with a stainless-steel waste tray underneath. The entire set-up is housed in a sound- and light-attenuating box equipped with a house light, a tone generator and a ventilating fan.

Above the arena, a house light (3 W) and a video camera are mounted. A peristaltic pump is positioned centrally behind the touchscreen unit to deliver the liquid reward; in our experiment, we used strawberry-flavoured milkshake (Québon, Agropur) as the food reward. An infrared beam detects entries into the reward-delivery magazine, which is equipped with a light and a small speaker. In addition, two infrared beams cross the arena to detect locomotor activity.

To minimize unintended screen touches and to demarcate screen response locations, a black Perspex mask with five response windows (each consisting of 4 × 4-cm square aperture, 1.5 cm above the grid floor) covered the touchscreen. The task schedules were designed and managed and the events recorded using Whisker Server and ABETTII software (Campden Instruments).

Behaviour

Mice were deprived of food until they reached 85–90% of their original weight. Before starting the delayed nonmatch-to-location task, the mice underwent several behavioural training stages, as previously described[48].

Pretraining

The mice were first habituated to human handling in the touchscreen chamber room for three days. After this, they were acclimated to the chamber itself, with rewards presented in the reward tray. They could progress to the next stage once they finished the reward within 20 min, typically within 1–2 days.

After the habituation phase, the mice were trained to touch the screen when a white square stimulus was presented pseudo-randomly in one of five possible locations on the screen. A reward was given when the mouse touched the screen while the sample was displayed. The mice progressed to the next stage after completing 30 trials within 60 min.

The next stage required the mice to touch the white square on the screen to receive a reward, with the same completion criterion of 30 trials within 60 min. Subsequently, the mice had to learn to initiate trials by moving to the back of the chamber and breaking the infrared beam near the reward magazine.

In the final pretraining stage, a touch to blank windows resulted in a five-second timeout, signalled by the illumination of the house light. Correction trials, which repeated the same trial after a five-second ITI, were administered until the mouse made a correct response. However, these correction trials were not included in the performance calculation. Reward collection initiated a 15-second ITI before the next trial began.

Task

The delayed nonmatch-to-location task consists of two phases: the sample phase, which is an encoding phase, in which the mouse learns the location of the cue; and a retrieval phase, in which it has to remember the cue location and choose the non-matching one. The first stage of training is designed to teach the non-matching rule, requiring the mouse to identify the novel location as the correct choice (Fig. 1f).

During the sample phase, one of five locations on the touchscreen is illuminated. After a nose poke to this location, the mouse is directed to the back of the chamber by the illumination of the reward tray (an 800-ms pulse delivering 20 µl of milkshake) and an auditory tone. To maintain the mouse's engagement in the task, during the sample phase, a small incentive (one-quarter of the total reward) is delivered in one-third of the trials, selected randomly. The smaller magnitude of the incentive provides the opportunity to compare the reward-encoding properties during the incentive and the actual reward.

The delay length is maintained at 2 s during learning of the non-matching rule and is then increased by an increment of 2 s during specific probe trials. Once the back infrared beams are broken after the delay period, the original sample and a novel correct location are presented simultaneously on the touchscreen.

If the mouse makes an incorrect response to the original sample location, a correction trial loop is initiated until the correct response is made. Correction trials are repeated presentations of the same sample and choice locations after an incorrect response. Mice were trained until they reached an average of 70% correct over 2 sessions of 36 trials. Once the mice reached the criterion for trials with a two-second delay, they progressed to trials with a four-second delay, and so on.

Data acquisition

In vivo calcium videos were recorded using a UCLA miniscope[21] (v.3; http://miniscope.org) equipped with a monochrome CMOS imaging sensor (MT9V032C12STM, ON Semiconductor). This sensor was connected to a custom data acquisition (DAQ) box (Miniscope) with a lightweight, flexible coaxial cable. The DAQ box was linked to a PC using a USB 3.0 SuperSpeed cable and operated with Miniscope custom acquisition software. Videos of mouse behaviour were recorded with an infrared camera positioned above the touchscreen. The DAQ simultaneously acquired behavioural and cellular imaging streams at 30 Hz as uncompressed avi files and all recorded frames were time-stamped for post hoc alignment. The touchscreen chamber also provides task-related information such as trial initiation timing, nose pokes to the screen and reward onset. A touchscreen chamber time stamp is also provided for follow-up alignment with neuronal and behavioural data.

Data preprocessing

We have used a UCLA miniscope to simultaneously record several hundred neurons in a freely moving mouse[21]. This provides the possibility of monitoring hundreds of neurons that are located inside of our field of view. The output of this recording in our experimental set-up is a video with 30 frames per second temporal resolution and 2–3 µm spatial resolution. The temporal resolution is sufficient to capture the slow dynamics of calcium transients and the spatial resolution is sufficient to capture the cell bodies. The main steps for analysing the calcium recording videos are as follows: (1) within-session motion correction to address small displacements and shakes during recording; (2) detecting cell bodies; (3) extracting calcium traces for each cell body by measuring the average fluorescent emission from the detected cell body; (4) inferring the likelihood of spikes from the raw calcium traces.

Calcium imaging data were preprocessed before analyses using a pipeline of open-source MATLAB (MathWorks; v.R2021a) functions to correct for motion artefacts[22], segment cells and extract transients[22,23,49]. A second-order autoregressive model is used to infer the likelihood of spiking events through the deconvolution of the transient trace as described previously[49]. The resulting time series is used to measure the 'firing rate'.

DeepLabCut, a deep-learning-based pose-estimation tool, was used to track multiple body parts of the mouse during behaviour[50]. The tracking is used to estimate position, heading direction, speed and other behavioural features.

Identification of cell types

To identify each cell type (reward, reward-approach and screen cells), the averaged neuronal response of each cell to each of the three features was calculated. Averaged neuronal activity at reward: average deconvolved traces during the reward consumption period. Averaged neuronal activity at reward approach: average deconvolved traces between the correct choice and the onset of the reward. Neuronal activity at the screen: average deconvolved traces at a window of [−150 ms, 150 ms] around screen pokes during choice period. The averaged neuronal activity for each cell is compared to the distribution of averaged neuronal responses made by 1,000 circular shuffles. Cell types were identified as those whose neuronal activity exceeded the 99th percentile of the corresponding shuffled distribution.

Tracking cells

To identify cells and track them across days, we first used the constrained non-negative matrix factorization (CNMFe) toolbox[23,51] to simultaneously identify neuron locations, separate spatially overlapping components and denoise and deconvolve spiking activity from the slow dynamics of the calcium indicator. Once the cells are identified, we use CellReg to track cells across sessions on the basis of their spatial footprints[52].

Although the effectiveness of this method has been shown in previous studies[53], we performed additional analyses to verify the reliability of our cell identification and cell tracking and to ensure that potential errors do not influence our results for cell registration and tracking procedures.

To do this, first, we made a detailed visualization of the tracked cells with a close look at their functional properties across sessions. Extended Data Fig. 9 shows the neuronal footprints of different cell types. Extended Data Fig. 9b shows the tracking of one reward cell across sessions. The green contour outlines the detected cell body, and the sessions in which CellReg[52] has failed to track the cell are in red. This analysis provides visual proof of the reliable tracking of cells across days, with their tuning properties being preserved around the reward onset. Extended Data Fig. 9biii shows the response of the neurons across sessions as presented in the main manuscript. To assess the impact of potential tracking errors, we introduced controlled imperfections: in Extended Data Fig. 9biv, we artificially replaced each tracked cell with a randomly selected nearby neuron (within five cell diameters). As shown, both the consistent response pattern and the structured backward shifting seen in Extended Data Fig. 9biii disappeared, supporting the idea that these effects rely on accurate cell identification. We test this null hypothesis more systematically in the next analysis. For further illustration, Extended Data Fig. 9c presents a second example neuron, showing similarly reliable tracking and reward-related tuning across sessions.

Furthermore, we tested the robustness of our results against a null hypothesis to assess how potential misalignment or inaccuracies in cell registration across days might affect the backward shifting of reward-cell activity. To simulate registration errors, we randomly selected a proportion of sessions for each tracked reward cell. We replaced the identified cell for each of those sessions with a randomly selected nearby neuron (within five cell diameters). This process introduces a controlled misalignment in cell registration. For each reward cell with the new modified tracking, we computed the average reward response of the cell across sessions and correlated the timing of peak activity with the session number. Similar to the analysis in the main text of the paper, a negative correlation indicates a backward shift in response timing relative to the reward onset. If the observed backward shift is a genuine effect, we expect this correlation to weaken as the proportion of misaligned sessions increases. We systematically varied the percentage of misaligned sessions from 0% (true data with no misalignment) to 100% (complete misalignment across all sessions) (Extended Data Fig. 9). The analysis was performed for all tracked reward cells and also for only backward-shifting reward cells. Both groups showed a substantial decrease in the absolute value of the negative correlation as we increased the level of misalignment. This result supports the idea that the structured reorganization of reward cells relies on accurate cell identification.

Across 6 mice (the tracking quality for mouse 2 was poor, so it was excluded from the analysis), we successfully tracked a total of 3,165 neurons, defined as cells that were tracked for at least 5 sessions. To ensure data quality, we excluded any sessions in which a cell's activity variance was below the 50th percentile of the overall distribution (calculated across all cells and sessions). After applying this filtering criterion, we retained 1,814 neurons with sufficiently high variance in activity traces across all tracked sessions. These cells form the basis for the broader functional analyses presented in the main text and in Extended Data Figs. 10 and 11. Out of the 1,814 cells that passed our quality and tracking criteria, we identified 228 reward cells (12.6%), 225 screen cells (12.4%) and 53 reward-approach cells (2.9%). The remaining 1,308 cells (72.1%) did not meet the classification criteria for any of these 3 functional groups, and we refer to them as non-classified cells. More analysis on the reliability of our cell tracking is presented in Extended Data Fig. 9.

Naive Bayes spatial decoding

To decode the positions of mice from calcium traces within each session, we divided the spatially binned position (using spatial bins of 1 cm along each of the axes) and our deconvolved calcium traces into fivefold splits. The binned positions were converted into a one-hot vector. Using the Gaussian Naive Bayes method from the scikit-learn Python library, we predicted positions on the withheld data using maximum likelihood estimation, as follows:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}}\exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right),$$

where x_i corresponds to the predicted position at i and y to the respective calcium traces. We assumed a flat prior (equal likelihood at all positions) and the scikit-learn default σ value of 1×10^{-9}. The decoded position is assigned to the bin with the highest probability. The decoding error was then estimated as the Euclidean distance between the mouse's predicted and actual spatial bin position on withheld data.

CEBRA embedding

CEBRA is an algorithm that optimizes neural networks that map neural activity onto an embedding space[26]. This algorithm uses contrastive learning[54] and a generalized InfoNCE loss function[26] to learn representations, where similar data points are pulled closer together and dissimilar data points are pushed apart within the embedding space. CEBRA has three different modes: CEBRA-Time (fully unsupervised or self-supervised), CEBRA-Behaviour (supervised) and CEBRA-Hybrid.

In our study, we used CEBRA-Time, so our input data will be unlabelled and there will be no behavioural assumptions that influence neuronal activity. CEBRA embedding is used to project the deconvolved neuronal traces into a 32-dimensional latent space[26]. We set all parameters as default and used the same set of parameters across all sessions.

MI content of task features

CEBRA embedding is used to project the deconvolved neuronal traces into a 32-dimensional latent space[26]. The feature of interest (screen, reward approach or reward) was presented as a binary vector in which for each frame, if the mouse is in that condition it is 1; otherwise it is 0. A fivefold cross-validation is implemented in which for each fold one fold of the data is held out, and a scikit-learn-based linear decoder is trained on the latent presentation to decode the class of the binary target. The MI[55,56] (using sciki-learn) between the predicted and the actual target for the held-out data is calculated. The averaged MI across five folds is considered the MI for each recording session. MI measures the dependency between two variables. It is equal to zero if and only if two random variables are independent, and higher values mean higher dependency. MI between two discrete variables (as in our case) X and Y can be calculated as follows:

$$\mathrm{MI}(X, Y) = \sum_{i,j} p(i,j) \log \frac{p(i,j)}{p_x(i) p_y(j)},$$

where i covers the values in X and j covers the values in Y. $p(i,j)$ is the joint distribution of the two variables X and Y. $p_x(i)$ is the distribution function of X and $p_y(j)$ is the distribution function of Y.

Linear modelling of information content

To compute the contribution of each feature (session number and mouse performance) in the evolution of hippocampal representation, we used a MATLAB function, fitlm, to model each of the measures of interest denoted by Y, by session number and performance. The contribution of each of the features is measured by their contribution to the explained variance of Y:

Session number = $\mathrm{var}_{\mathrm{full}} - \mathrm{var}_{\mathrm{performance}}$

Performance = $\mathrm{var}_{\mathrm{full}} - \mathrm{var}_{\mathrm{session}}$

Session number and performance = $\mathrm{var}_{\mathrm{full}} - \mathrm{var}_{\mathrm{session}} - \mathrm{var}_{\mathrm{performance}}$

$\mathrm{var}_{\mathrm{session}}$: model's explained variance when Y is modelled only by session number.

$\mathrm{var}_{\mathrm{performance}}$: model's explained variance when Y is modelled only by performance.

$\mathrm{var}_{\mathrm{full}}$: model's explained variance when Y is modelled by both session number and performance.

Temporal shifting score

We defined a temporal shifting score by comparing each cell's true correlation between peak activity timing and session number to a shuffled distribution of correlations, providing a standardized measure of temporal shift.

$$\text{Temporal shifting score} = \frac{\mathrm{corr}_{\mathrm{true}} - \mu_{\mathrm{shuffle}}}{\sigma_{\mathrm{shuffle}}},$$

where $\mathrm{corr}_{\mathrm{true}}$ is the observed correlation coefficient between session number and the timing of peak activity, and μ_{shuffle} and $\sigma_{\mathrm{shuffle}}$ are the mean and standard deviation of the correlation values obtained from a shuffled distribution. The shuffled distribution was generated by randomly permuting session numbers 1,000 times.

Amplitude change score

To quantify changes in response amplitude, we defined an amplitude change score using the same approach as for the temporal shifting score, except that we correlated session number with the peak activity amplitude of each neuron instead of its timing.

$$\text{Amplitude change score} = \frac{\mathrm{corr}_{\mathrm{true}} - \mu_{\mathrm{shuffle}}}{\sigma_{\mathrm{shuffle}}},$$

where $\mathrm{corr}_{\mathrm{true}}$ is the observed correlation coefficient between session number and the response amplitude, and μ_{shuffle} and $\sigma_{\mathrm{shuffle}}$ are the mean and standard deviation of the correlation values obtained from a shuffled distribution. The shuffled distribution was generated by randomly permuting session numbers 1,000 times.

Identification of backward- and forward-shifting cells

Within-session averaged calcium traces were calculated for each of the cells. The correlation between the session number and the time of peak activity was calculated. A shuffled distribution was calculated by correlating the time of peak activities and the shuffled session numbers. A cell was identified as a backward-shifting cell if its correlation was less than the 5th percentile of the shuffled distribution. Similarly, a cell was identified as a forward-shifting cell if its correlation was more than the 95th percentile of the shuffled distribution. This criterion sets the chance level as 5%.

Identification of declining and inclining cells

Similar to backward- and forward-shifting cells, here, within-session averaged calcium traces were calculated for each of the cells. The correlation between the session number and the amplitude of peak activity of raw calcium traces was calculated. A shuffled distribution was calculated by correlating the amplitude and the shuffled session numbers. A cell was identified as a declining cell if its correlation was less than the 5th percentile of the shuffled distribution. Similarly, a cell was identified as an inclining cell if its correlation was more than the 95th percentile of the shuffled distribution. This criterion sets the chance level as 5%.

Rate maps

To calculate the rate map for each neuron, we binned the x and y axis each into 30 bins. The rate value assigned to each bin was simply calculated by the sum of neuronal activity (deconvolved traces) normalized by the time the mouse spent in that bin. For visualization, we used a Gaussian filter of size 5 × 5 bins and $\sigma = 1$ bin.

Identification of place cells

We computed the spatial information of all cells using the unsmoothed-event rate map of each cell, as previously described[57].

$$\text{Spatial information} = \sum_i p_i \left(\frac{r_i}{\underline{r}}\right) \log_2 \left(\frac{r_i}{\underline{r}}\right),$$

where p_i is the probability of the mouse being in the ith bin (time spent in ith bin/total running time); r_i is the Ca^{2+} event rate in the ith bin; and

$\underline{r}$ is the overall Ca^{2+} event rate. We then performed 1,000 distinct shuffles of mouse locations during Ca^{2+} events and calculated the spatial information for each shuffle. Cells with spatial information higher than that of 99% percentile of their shuffles were identified as place cells.

Box-plot statistics

Box plots were generated using the seaborn.boxplot function. Each box represents the interquartile range (IQR), defined by the 25th percentile (Q1) and 75th percentile (Q3), with a horizontal line indicating the median. The whiskers extend to the most extreme data points within 1.5 × IQR from the quartiles. Data points outside this range are considered outliers and are shown as individual markers. This visualization provides a summary of the data distribution, including central tendency, spread and outlier values for each group or condition.

Measuring the size of the place fields

After identifying place cells and determining their rate maps, we masked these rate maps by setting all bins with values below the 90th percentile of values across all bins to zero. This operation creates islands of non-zeros surrounded by bins of zero value. We used the MATLAB function bwconncomp to detect these islands and used regionprops to calculate the area of each of these islands. The island with the maximum area was used to calculate the place-field size. The size of the place field was then calculated as:

$$\text{Characteristic size of place field} = \sqrt{\text{Place field area}}.$$

Classification decoding analysis

A linear support vector machine classifier from MATLAB (fitcecoc) was used to decode contextual information in the task. For example, we want to decode the correctness of the trial at different moments of the task. Population vectors of deconvolved calcium traces were used to train and test the classifier. Given the limited number of samples, we used a leave-one-out approach by training our classifier on all samples except one and testing it on the excluded sample, repeating this process for each sample point. In each iteration, we ensured that the training dataset contained an equal number of samples from each class by randomly downsampling the class with a larger number of samples. For each decoding, we repeated the process five times and averaged the decoding accuracy across these five iterations. The classifier's decoding performance was compared to the accuracy obtained from shuffled interaction, in which the class labels were randomly shuffled.

Reward over-representation score

To measure the reward over-representation score within each session, we first generated the spatial rate map for all cells. We identified the location of peak activity for each rate map, detecting the spatial bin with the highest firing rate for all cells. A density plot was then generated to represent the density of peaks in each spatial bin. This matrix provides a representation of each spatial bin. The reward over-representation was calculated as the average representation of the 10% of spatial bins closest to the reward port, normalized by the average representation across all spatial bins. In this context, a score of 1 indicates an even distribution of reward representation compared to the baseline, whereas a value greater than 1 signifies an over-representation of spatial bins near the reward port.

Statistical analysis

For visualization, we used error bars (or shaded areas for line plots) to show s.e.m. To compare two distributions, we used the two-sample Kolmogorov–Smirnov test, through the 'kstest2' command in MATLAB. For comparisons of a distribution against zero, we used the Wilcoxon signed-rank test, implemented using MATLAB's 'signrank' command. Significance levels for all tests were set at $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$.

Simulations for the TD-error-modulated place-cell model

Anatomical relevance. The model's architecture is motivated by hippocampal–striatal–VTA circuitry, in which projections from the hippocampus to the ventral striatum indirectly regulate VTA dopamine activity[36,58–61]. The ventral striatum encodes value ramps and VTA encodes RPE-like signals[1,28,59]. Dopaminergic feedback from the VTA to both the hippocampus and the striatum modulates plasticity for learning[1,36,62]. This circuit organization motivates a feed-forward structure in which place cells drive RPE computation, and a feedback structure in which RPE signals modulate plasticity underlying both value computation[40,63,64] and place-cell spatial selectivity[11].

Abstracted navigation task. We modelled navigation as a Markov decision process with 10 discrete states (state 0 to state 9) on a 1D track. State transitions were deterministic, with the agent receiving reward $R = 1$ by being at state 7 and choosing the action right to reach terminal state 8. After this, the trial ended immediately, and a new trial began with the agent starting at state 1. The state space had absorbing boundaries (no circular topology) so that the agent could not reach state 0 from taking a step from state 9. The results were robust with 5 or 15 states.

Neural representation. Following a previous study[11], each place cell's spatial selectivity φ_i is modelled as a Gaussian radial basis function:

$$\phi_i(s_t) = \exp\left(-\frac{(s_t - \lambda_i)^2}{2\sigma^2}\right),$$

where s_t is the current state; λ_i is each place cell's peak activity location, which was initialized uniformly across states 0 to 9; and $\sigma = 0.5$ is the spread of each place cell. The agent estimates the value of a location using a linear readout of $N = 1{,}000$ place cells:

$$v_\pi(s) = \frac{1}{N}\sum_i^N w_i^v \phi_i(s),$$

using the vector w^v. Increasing or decreasing the number of place cells N did not significantly change the dynamics of place-cell peak reorganization as the agent is initialized in a rich-feature learning regime[65,66] instead of a lazy-feature learning regime[11].

Learning algorithm. We assume that the central objective for which animals are optimizing is reward maximization[11]. However, we first consider the simplified case of value estimation, in which the agent has an optimal policy π^*, and the objective for the agent is to learn to estimate the value of the state given a policy. This simplification is to aid our understanding of the intuition and visualize how the TD error learning signal directly modulates each place cell's peak shifts. Hence, the value estimation objective is to minimize the loss function, described by the TD error:

$$\begin{aligned}\mathcal{L}(w^v, \lambda) &= E_{\tau - \pi^*}\left[\sum_{t=0}^{T-1} \frac{1}{2}(r(s_t, a_t) + \gamma v(s_{t+1}; w^v, \lambda) - v(s_t; w^v, \lambda))^2\right]\\ &= E_{\tau - \pi^*}\left[\sum_{t=0}^{T-1} \frac{1}{2}\delta(s_t, a_t)^2\right],\end{aligned}$$

starting with a commonly used reward discount factor ($\gamma = 0.95$). Changing the discount factor influences the backward-shifting dynamics. We optimize only the critic's weights (w^v) and each place cell's centre of mass (λ) to study how this minimization influences peak activity. This results in the critic's weights being updated by:

$$\Delta w_t^v = \frac{1}{N}\delta(s_t)\phi(s_t),\ w_{t+1}^v = w_t^v + \eta\Delta w_t^v,$$

where $\eta = 0.1N$. Using chain rule, each place-cell centre is updated according to:

$$\Delta\lambda_t = \frac{1}{N}\delta(s_t, a_t)w_t^v\phi(s_t)\left(\frac{s_t - \lambda_t}{\sigma^2}\right),\ \lambda_{t+1} = \lambda_t + \eta\Delta\lambda_t$$

to minimize the loss function (see Supplementary Information for derivation). We visualized the value estimation objective in the main text to understand the intuition of how the TD error modulates each place cell's peak shifts. We also investigated how policy learning for reward maximization (see model details in Supplementary Information) influences place-cell shifts (Extended Data Fig. 15), which more closely replicates the objective of animals and the stochastic shifting dynamics that were observed in the experimental results.

Reporting summary

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

Data availability

The complete dataset for all experiments is available at McGill University Dataverse (https://doi.org/10.5683/SP3/877CGZ). The dataset should not be used for republication without prior consent from the authors.

Code availability

All source codes used in the current study are available on request to the corresponding authors.

47. Heath, C. J., Phillips, B. U., Bussey, T. J. & Saksida, L. M. Measuring motivation and reward-related decision making in the rodent operant touchscreen system. *Curr. Protoc. Neurosci.* **74**, 8.34.1–8.34.20 (2016).
48. Kim, C. H. et al. Trial-unique, delayed nonmatching-to-location (TUNL) touchscreen testing for mice: sensitivity to dorsal hippocampal dysfunction. *Psychopharmacology* **232**, 3935–3945 (2015).
49. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.* **13**, e1005423 (2017).
50. Lauer, J. et al. Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nat. Methods* **19**, 496–504 (2022).
51. Pnevmatikakis, E. A. et al. Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2016).
52. Sheintuch, L. et al. Tracking the same neurons across multiple days in Ca^{2+} imaging data. *Cell Rep.* **21**, 1102–1115 (2017).
53. Yang, W. et al. Simultaneous multi-plane imaging of neural circuits. *Neuron* **89**, 269–284 (2016).
54. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A simple framework for contrastive learning of visual representations. In *International Conference on Machine Learning* 1597–1607 (PMLR, 2020).
55. Kraskov, A., Stogbauer, H. & Grassberger, P. Estimating mutual information. *Phys. Rev. E* **69**, 066138 (2004).
56. Ross, B. C. Mutual information between discrete and continuous data sets. *PLoS ONE* **9**, e87357 (2014).
57. Markus, E. J., Barnes, C. A., McNaughton, B. L., Gladden, V. L. & Skaggs, W. E. Spatial information content and reliability of hippocampal CA1 neurons: effects of visual input. *Hippocampus* **4**, 410–421 (1994).
58. Floresco, S. B., Todd, C. L. & Grace, A. A. Glutamatergic afferents from the hippocampus to the nucleus accumbens regulate activity of ventral tegmental area dopamine neurons. *J. Neurosci.* **21**, 4915–4922 (2001).
59. Barnstedt, O., Mocellin, P. & Remy, S. A hippocampus-accumbens code guides goal-directed appetitive behavior. *Nat. Commun.* **15**, 3196 (2024).
60. Kalivas, P. W., Churchill, L. & Klitenick, M. A. GABA and enkephalin projection from the nucleus accumbens and ventral pallidum to the ventral tegmental area. *Neuroscience* **57**, 1047–1060 (1993).
61. Ibrahim, K. M. et al. Dorsal hippocampus to nucleus accumbens projections drive reinforcement via activation of accumbal dynorphin neurons. *Nat. Commun.* **15**, 750 (2024).
62. Russo, S. J. & Nestler, E. J. The brain reward circuitry in mood disorders. *Nat. Rev. Neurosci.* **14**, 609–625 (2013).
63. Kumar, M. G., Tan, C., Libedinsky, C., Yen, S.-C. & Tan, A. Y. Y. A nonlinear hidden layer enables actor-critic agents to learn multiple paired association navigation. *Cereb. Cortex* **32**, 3917–3936 (2022).
64. Krishnan, S., Heer, C., Cherian, C. & Sheffield, M. E. J. Reward expectation extinction restructures and degrades CA1 spatial maps through loss of a dopaminergic reward proximity signal. *Nat. Commun.* **13**, 6662 (2022).
65. Bordelon, B. & Pehlevan, C. Self-consistent dynamical field theory of kernel evolution in wide neural networks. *J. Stat. Mech.* **2023**, 114009 (2023).
66. Vyas, N. et al. Feature-learning networks are consistent across widths at realistic scales. *Adv. Neural Inf. Process. Syst.* **36**, 1036–1060 (2023).
67. Paninski, L. & Cunningham, J. P. Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience. *Curr. Opin. Neurobiol.* **50**, 232–241 (2018).
68. Jazayeri, M. & Ostojic, S. Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr. Opin. Neurobiol.* **70**, 113–120 (2021).
69. Urai, A. E. et al. Large-scale neural recordings call for new insights to link brain and behavior. *Nat. Neurosci.* **25**, 11–19 (2022).
70. Yu, B. M. et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* **102**, 614–635 (2009)
71. Hollup, S. A. Molden, S, Donnett, J. G., Moser, M. B. & Moser,E. I. Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *J. Neurosci.* **21**, 1635–1644 (2001).

Acknowledgements We thank the members of the M.P.B. laboratory for discussions and for providing inputs during the analysis of the data; in particular, Z. Ajabi and J. Q. Lee. We thank M. Iordanova, B. Richards, E. J. P. Maes, J. Quinn Lee, H. Nagaraj, Z. Haqqee and A. Sharma for comments on the first draft of this manuscript, and A. Peyrache and B. Richards for their guidance on data analysis. M.G.K. and C.P. were supported by an NSF award (DMS-2134157). C.P. was also supported by an NSF CAREER award (IIS-2239780), a Sloan Research Fellowship and the William F. Milton Fund from Harvard University. This work was made possible in part by a gift from the Chan Zuckerberg Initiative Foundation to establish the Kempner Institute for the Study of Natural and Artificial Intelligence. This work was also supported by funding from Fonds de Recherche du Québec – Santé (FRQS) postdoctoral fellowships awarded to A.N.-P. and C.-A.M.; CIHR project grants 463403 and 480510 to M.P.B.; and a Core Facilities and Technology Development grant from the Canada First Research Excellence Fund, 'Health Brains for Health Lives', to S.W. and M.P.B.

Author contributions M.Y., A.N.-P., S.W. and M.P.B. conceptualized the project. A.N.-P. performed surgeries and recordings. M.Y. organized the raw data and did the preprocessing, analysis, modelling and data visualization. M.Y., T.G and É.W. did the preprocessing of the data. M.G.K. developed the model and ran the simulations. M.G.K. and M.Y. analysed the simulated data. C.P. supervised the modelling section. M.Y., M.G.K. and C.-A.M. wrote the initial draft. M.Y., M.G.K., C.-A.M., C.P., S.W. and M.P.B. contributed to editing and revising the first draft of the paper. M.P.B. guided and supervised all stages of experiments and data analysis.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09958-0.
Correspondence and requests for materials should be addressed to Mohammad Yaghoubi or Mark P. Brandon.
Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.