

# Rapid Neural and Behavioral Adaptation in a Updated Theory of the Cognitive Map and its Neural Implementation

Suheecho<sup>1,\*</sup> and James L. McClelland<sup>1,\*</sup>

<sup>1</sup>Stanford University, Department of Psychology, Stanford, CA 94306, USA

\*{suheecho,jlmcc}@stanford.edu

## ABSTRACT

Animals rapidly and flexibly adapt their behavior based on experience by forming an internal cognitive map of their environment. The cognitive map has been studied at multiple levels, including normative, representational, and mechanistic accounts. Here, we propose extensions of existing normative theory and a mapping of this theory to the brain, incorporating biologically motivated learning mechanisms, including behavioral time scale plasticity (BTSP) and spike-time dependent plasticity during replay, as well as experimentally attested findings about factors that affect place-cell formation, to present a model that captures rapid adaptations in neural representations, replay processes, and observed behavior. In our model, CA3 uses a symmetric variant of BTSP to encode spatiotemporal proximity experienced during exploration, CA1 uses a temporally asymmetric variant of BTSP to encode a salience-weighted and predictive successor representation, and downstream areas learn successor feature predictions that can guide behavior toward environmental features (such as food or water) relevant to the animal's current motivational state (such as hunger or thirst). Through simulations, we show that the network rapidly learns cognitive maps that overweight salient locations in accordance with experimental findings. The learned connection weights underlying the learned map then support off-line replay, which further extend the learned predictive representations and propagate salience information to unvisited states, enhancing behavioral flexibility. Behavioral simulations show how these processes give rise to rapid behavioral adaptation, including one-trial avoidance learning and motivational-state dependent choice.

## Introduction

Animals adapt quickly to their experience in spatial environments. This ability is closely tied to the concept of a "cognitive map"—a putative representation that receives locally available information, integrates it with internal state information, and generates signals that guide behavior at any given moment<sup>1–3</sup>. Classic maze experiments show that rodents rapidly learn to associate spatial and reward-related cues to make efficient decisions (Fig. 1a), motivating three central questions: (1) **What is learned?** In what form are spatial and reward signals integrated in the brain's representation of the cognitive map? (2) **How is it learned?** What neural mechanisms give rise to the cognitive map? (3) **How is it used?** How does the output of the cognitive map contribute to evaluating different possible choices and guiding behavior?

To address the question *What is learned?*, we build on the compelling normative account of the cognitive map provided by the successor representation (SR)<sup>4,5</sup>. Rather than encoding environments purely in terms of spatial distance, the SR encodes the expected future occupancy of each location based on the animal's movement history (Fig. 1b, top). This normative account is supported by characteristics of place cells in the hippocampus<sup>5</sup> such as predictive skewing of place fields<sup>6,7</sup>. However, it has been noted<sup>8</sup> that the classic version of SR

cannot explain stronger encoding of behaviorally important but infrequently visited locations, such as those associated with aversiveness<sup>9</sup>, motivational salience<sup>10</sup>, or novelty<sup>11</sup>.

These factors, collectively defined as *perceived salience* (PS) in prior work<sup>12</sup>, provide crucial input to the hippocampus and are especially important given its role in rapid adaptation and flexible learning. Existing experimental evidence supports the idea that hippocampal SR is modulated by PS, as reflected in increased numbers of place cells representing the salient locations<sup>13–16</sup> as well as the acceleration of their formation<sup>17</sup>. Taken together, these findings support the idea that the hippocampus integrates PS signals with spatial input, forming a PS-weighted SR that is more adaptive than a purely spatial SR.

To answer *How is it learned?*, we probe further toward the goal of understanding the acquisition of the PS-weighted SR through local synaptic plasticity mechanisms. Over the past decade, a new variant of such a mechanism in the hippocampus, called behavioral time-scale synaptic plasticity (BTSP) has emerged<sup>18</sup>. BTSP is particularly important because it provides a mechanism giving rise to large changes in synaptic connections among neurons activated within a time scale of seconds, modulated by novelty, salience and reward<sup>19–22</sup>. We build on recent work linking BTSP to the formation of SR<sup>23</sup>

as a part of the framework we propose here, in which BTSP underlies the real-time formation of the PS-weighted SR.

A distinct form of learning occurs when the animal is not actively exploring but is instead at rest or asleep. During this “offline” phase, spontaneous hippocampal activity replays trajectories through the environment<sup>24</sup>. This phenomenon, known as hippocampal replay, involves place cells firing in sequential patterns during periods of immobility, and is often thought to contribute to the stabilization and consolidation of learning<sup>25–27</sup>. Remarkably, these sequences can represent trajectories that the animal has never physically traversed<sup>28</sup>, suggesting that replay helps complete the internal spatial map<sup>29,30</sup>. Although the precise mechanisms by which replay influences synaptic change are not yet fully understood, we include a proposed implementation based on spike-time dependent plasticity to support learning during replay in the hippocampal circuitry.

Finally, to answer *How is it used?* we consider how the learned PS-weighted SR guide decisions. Like the original SR, the PS-weighted SR provides an advantage for state-value evaluation that is a key computation in decision making: because it encodes expected future state occupancies, it allows values to be estimated by multiplying each state’s reward<sup>4</sup>. However, this formulation offers a limited account of how internal states influence decisions, which often depend on specific features rather than entire states. For instance, water is valuable when the animal is thirsty but not when satiated. How such feature-specific motivational needs are integrated into state value computation remains unclear.

Successor features (SFs) have been proposed as a way of addressing this issue<sup>31</sup>. Unlike SR, which predicts the *future occupancy of states*, SF predicts the *future occurrence of features* (Fig. 1b, bottom). Inspired by this work, we propose that these predicted features are combined with their associated values to compute the overall value of each state, allowing SFs to guide behavior. Importantly this allows feature-associated values to be flexibly modulated by the animal’s transient needs.

Building on these ideas, we begin by formalizing a unified normative model of the cognitive map, and then consider its mechanistic implementation in the brain and its application to decision making (Fig. 1c). In our model, the hippocampus first merges spatial and PS inputs to encode the PS-weighted SR, and then relays its output to connected areas representing SFs. The SFs are then combined with the animal’s current motivational state to determine the overall values of the states (Fig. 1d). To demonstrate how this representation emerges, we simulated a rodent maze-learning scenario consisting of online and offline phases. During the online phase, the network received spatial inputs and PS signals and be updated using BTSP rule (Fig. 1e). The following offline phase spontaneously generates replay activities in which these representations were strengthened (Fig. 1f). Subsequently, we simulated behavioral paradigms to show how the learned representations compute values of possible decisions and guided subsequent actions (Fig. 1g). In summary, our work presents a integrated

normative and mechanistic model of the cognitive map, emphasizing BTSP during online learning and replay during offline consolidation, and flexible use of the map to guide behavior.

## Results

We begin by presenting the formal statement of our normative theory and the assumptions we make in mapping from this theory to neural mechanisms. We then describe a series of simulations demonstrating how the model can address the findings that motivated the theory’s development.

### Formal Statement of Theory and Its Neural Correlates

As originally formulated<sup>4</sup>, the SR for a starting state  $s$  and a target state  $s'$  is the expected discounted future occupancy of  $s'$  across time  $t$ , given  $s_0 = s$ :

$$M(s, s') = \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t \mathbb{I}(s_t = s') \mid s_0 = s \right], \quad (1)$$

where  $\gamma \in [0, 1]$  is a discount factor. Thus  $M(s, s')$  weights nearer visits more heavily than distant ones.

In our model, we focus instead on SFs—predictions of future occurrence of environmental features (e.g., cues, rewards, or hazards)<sup>31</sup>. Let  $\phi(s', f)$  denote the expected presence of feature  $f$  at state  $s'$ . The SF for feature  $f$  at start state  $s$  is defined as the expected discounted future occurrence of that feature:

$$\begin{aligned} \psi(s, f) &= \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t \phi(s_t, f) \mid s_0 = s \right] \\ &= \sum_{s'} M(s, s') \phi(s', f). \end{aligned} \quad (2)$$

We further incorporate perceived salience (PS),  $\omega(s')$  into our framework. First, we define the PS-weighted SR,

$$M^{\text{PS}}(s, s') = \omega(s') M(s, s'). \quad (3)$$

That is, PS amplifies states by  $\omega(s')$  in the SR. With this scaling, the SF can be expressed as a linear map from  $M^{\text{PS}}$ :

$$\psi(s, f) = \sum_{s'} M^{\text{PS}}(s, s') \cdot W_{\psi, M^{\text{PS}}}(s', f), \quad \text{where} \quad (4)$$

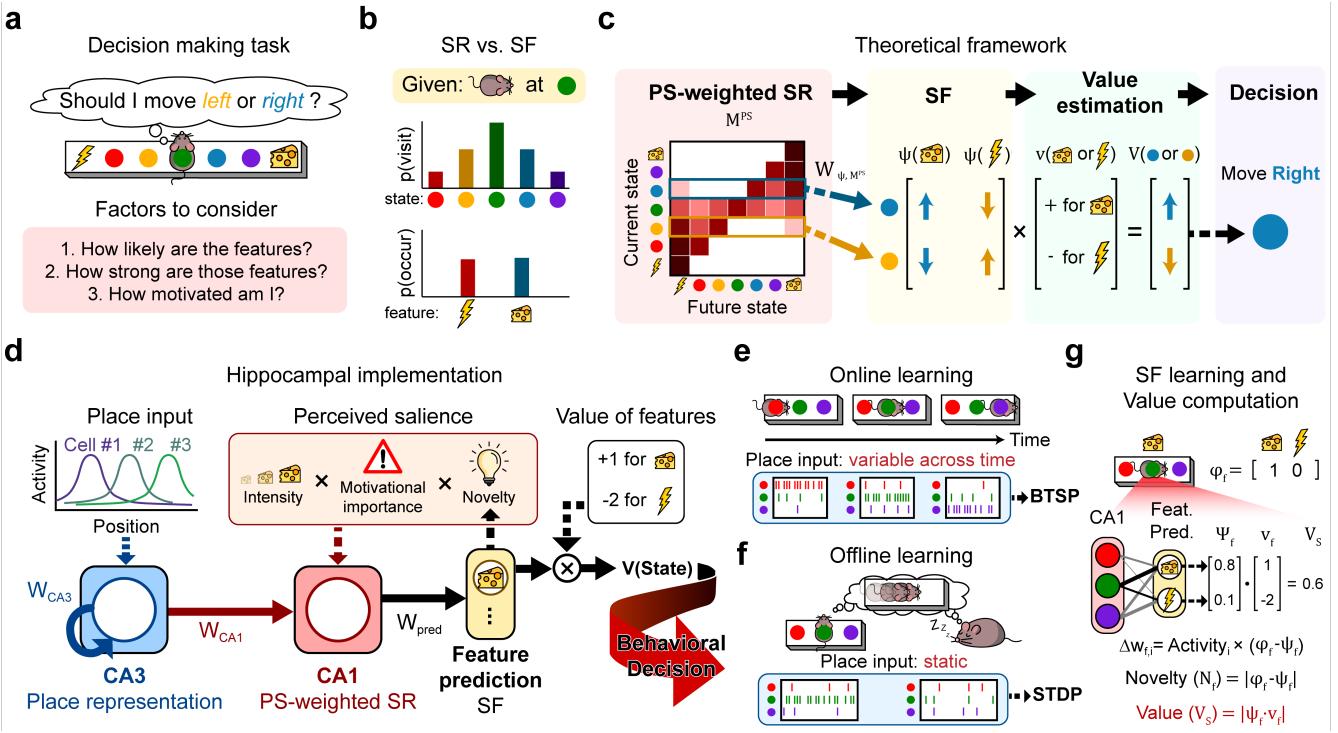
$$W_{\psi, M^{\text{PS}}}(s', f) = \frac{\phi(s', f)}{\omega(s')}. \quad (5)$$

Here, the mapping to features are divided by  $\omega(s')$  to recover the original feature mappings.

Finally, the value of state  $s$  is defined as the sum of successor features weighted by their value:

$$V(s) = \sum_f \psi(s, f) v(f), \quad (6)$$

where  $v(f)$  is the value of feature  $f$  at the time when  $V(s)$  is estimated. This allows the same predictive representation  $\psi$  to



**Figure 1.** **a**, Rodent decision-making in a maze. **b**, Two ways of representing a state: successor representation (SR) and successor features (SF). **c**, Schematic of the theoretical framework. A perceived-salience-weighted SR is relayed to encode successor features, which are then used to compute value and guide decisions. **d**, Hippocampal implementation of the model. CA3 receives place input and encodes place representations. CA1 receives input from CA3 along with perceived salience, and encodes the salience-weighted SR. This is passed to the feature prediction layer, which encodes SF and is ultimately used to determine the subjective value of states. **e-f**, Two forms of learning that train the networks. **e**, Online learning occurs when the animal actively explores the environment and encounters features. Place input varies with position, and the hippocampal network is updated using the BTSP rule. **f**, Offline learning occurs while the animal rests or sleeps. Place input is static, and hippocampal networks are updated using symmetric STDP. **g**, Schematic of feature prediction layer learning.

flexibly support different decisions, depending on the agent's internal state (e.g., thirst making water valuable, but not food).

We link these theoretical quantities to neural processes and behavior as follows:

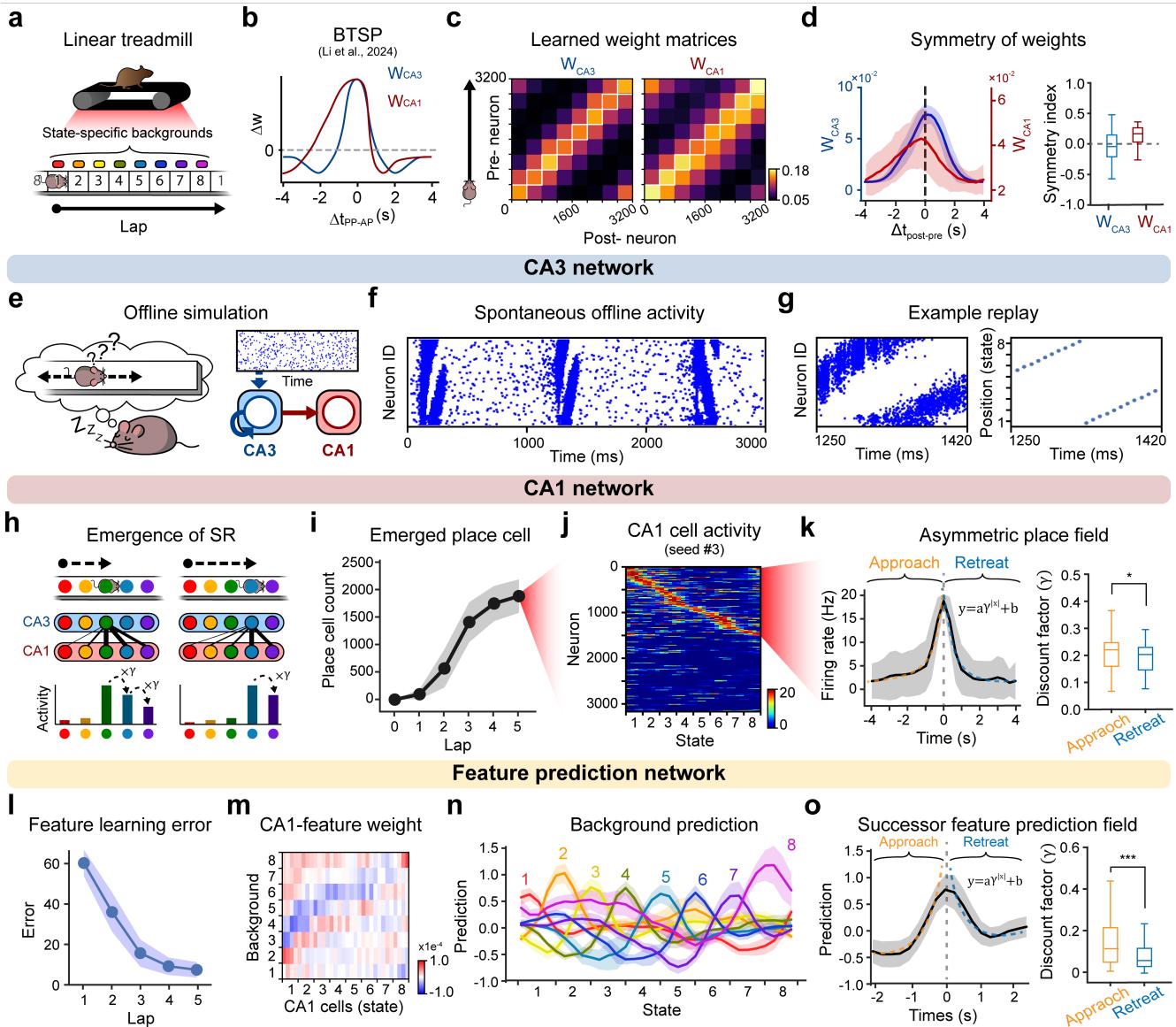
- An estimate of the **PS-weighted SR**,  $M^{PS}(s, s')$ , is acquired through BTSP in hippocampal areas CA3 and CA1 during exploration and is further reinforced through replay-driven plasticity.
- **Perceived salience**,  $\omega(s)$ , is determined based on the intensity, novelty, and motivational significance of features present at state  $s$  and is incorporated by modulating BTSP.
- The **mapping from PS-weighted SR to SF**,  $W_{\psi, M^{PS}}$ , is instantiated by learned synaptic weights from CA1 place cells to downstream feature-coding populations, resulting in estimates of  $\psi(s, f)$ . The weighting of the SR by perceived salience allows features associated with high PS states to be learned quickly, as we discuss below.
- **State value**,  $V(s)$ , is computed in extra-hippocampal circuits integrating feature predictions with feature-specific value signals  $v(f)$ , which depend on the animal's physiological needs during behavior.

- The resulting  $V(s)$  guides **decision making**: when selecting actions, animals preferentially transition to states with higher  $V(s)$ .

### Specialized representations of distinct hippocampal subregions under neutral conditions emerges through BTSP

First, we demonstrated that the hippocampal circuit can implement our normative framework in an environment without salient outcomes, with behavioral time-scale synaptic plasticity (BTSP) as the synaptic learning mechanism. We constructed an artificial neural network with three components: a CA3-like recurrent layer with connection weights ( $W_{CA3}$ ), a CA1-like feedforward layer with connection weights ( $W_{CA1}$ ) receiving input from CA3, and a "feature-prediction layer" representing downstream areas involved in decision making—all initialized randomly (see Methods). For simplicity, each CA3 cell had a predetermined excitatory input field centered at a specific location in the environment, with the centers evenly distributed throughout the environment (See Methods and Supplementary Fig. 1). Navigation was treated as uniform motion through the environment, producing a uniform progression through the centers of the place fields.

Using this model, we divided the linear treadmill into eight consecutive regions or 'states' (states 1-8; Fig. 2a) that



**Figure 2.** **a**, Linear treadmill simulation setup. **b**, BTSP weight update kernels for CA3 and CA1. **c**, Weight matrices with place cells sorted by field location, binned into 8 groups and averaged. Left:  $W_{CA3}$ . Right:  $W_{CA1}$ . **d**, Left: Averaged synaptic weights of postsynaptic neurons aligned by presynaptic index, for  $W_{CA3}$  (blue) and  $W_{CA1}$  (red). The curve is divided into weights from presynaptic neurons with smaller indices ("behind") and larger indices ("ahead"). Right: Normalized area of behind vs. ahead weights. **e-h**, Results from CA3. **e**, Offline simulation schematic. **f**, Example spontaneously generated offline activity (3 s). **g**, Left: Example replay activity. Right: Decoded position. **h-k**, Results from CA1. **h**, Successor representation emerges in CA1 due to asymmetry of  $W_{CA1}$ . **i**, Number of CA1 place cells formed after each lap. **j**, CA1 place fields after lap 5. **k**, Left: Average firing rate across place fields. Right: Fitted discount factors for approach vs. retreat trajectories relative to field centers. Paired t-test,  $N_{seed} = 10$ ,  $*P = 0.041$ . **l-o**, Results from the feature prediction layer. **l**, Average prediction error for background feature presence. **m**,  $W_{pred}$  after lap 5. **n**, Predictions for each background feature across positions after lap 5. **o**, Average predicted presence for the current state vs. non-current states. Paired t-test,  $N_{seed} = 10$ ,  $***P = 6.6 \times 10^{-4}$ . All data are presented as mean  $\pm$  s.d.

the animal was treated as moving through at a constant rate. Each state was associated with a distinct environmental feature (or "background") that was neutral in valence but served to distinguish spatial locations. Consistent with experimental findings<sup>32</sup>, CA3 cells were trained using the BTSP learning rule with a symmetric time kernel, whereas CA1 cells were trained using an asymmetric kernel (Fig. 2b). This distinction in learning kernels resulted in different structures in  $W_{CA3}$  and  $W_{CA1}$ . To illustrate this, we assigned each CA3 and CA1 neuron an ID based on the spatial position where it exhibited peak firing. Then, we plotted  $W_{CA3}$  and  $W_{CA1}$ , sorting both presynaptic and postsynaptic neurons by their IDs. For clarity, neurons were binned into 8 groups, the average weight within each bin was calculated, and the results were plotted (Fig. 2c). The results showed that  $W_{CA3}$  encoded spatial relationships between place fields in a direction-independent manner, resulting in a symmetric structure (Fig. 2d, blue). In contrast,  $W_{CA1}$  displayed a skewed distribution: connections from place cells representing positions behind the animal's current location were more potentiated than those ahead (Fig. 2d, red). We confirmed this by comparing the area under the "behind" and "ahead" portions of the curves (Fig. 2d, right).

Next, we examined the functional consequences that ensued from these distinct representations. We first showed that the CA3 network can spontaneously generate replay-like activity during offline periods after learning driven by BTSP. To test this, we delivered a uniform, non-spatial input for three seconds to the trained  $W_{CA3}$  and  $W_{CA1}$  to mimic sleep conditions (Fig. 2e, right). We simulated the resulting spontaneous activity using biophysically plausible parameters and an adaptive exponential integrate-and-fire neuron model<sup>33</sup> (for details, see Supplementary 6). As a result, CA3 neurons fired in sequence according to their place-field locations (Fig. 2f). These sequential events were best decoded as a linear travel along the track and therefore qualify as replay events<sup>34</sup> (Fig. 2g). Strikingly, the replay activity emerged after only a single lap of exploration<sup>35</sup>. In contrast, networks trained with STDP rule required approximately 400 laps before exhibiting comparable replay dynamics (Supplementary Fig. 3).

Next, we showed that the asymmetric structure of  $W_{CA1}$  endows CA1 neurons with SR-like activity patterns (Fig. 2h). To verify this, we examined the activity of individual CA1 neurons, which were randomly initialized. We classified a neuron as a place cell if its peak firing rate exceeded 15 Hz (75% of the model's theoretical maximum of 20 Hz) at any position along the track (see Methods). As a result, we found that after five laps of exploration, an average of 1,867 CA1 place cells had emerged (Fig. 2i). The resulting place cells exhibited fields distributed across the environment (Fig. 2j; also see Supplementary Fig. 2). As expected from  $W_{CA1}$ , these place fields were asymmetric: neurons began firing earlier as the animal approached the field's peak location and shut off more rapidly as it moved away (Fig. 2k, left). Within the SR framework, this profile can be interpreted as expressing a discounted probability of future visitation. By fitting the

place cell activity during the "approach" phase to the curve  $y = a \times \gamma^{|x|} + b$ , we estimated the empirical discount factor  $\gamma$  (Fig. 2k, right). The discount factor fell between 0 and 1, indicating that the network balanced memory of the immediate versus more distant past. For comparison, we calculated the discount factor during the "retreat" phase, when the animal moved away from the field peak, which revealed the asymmetry.

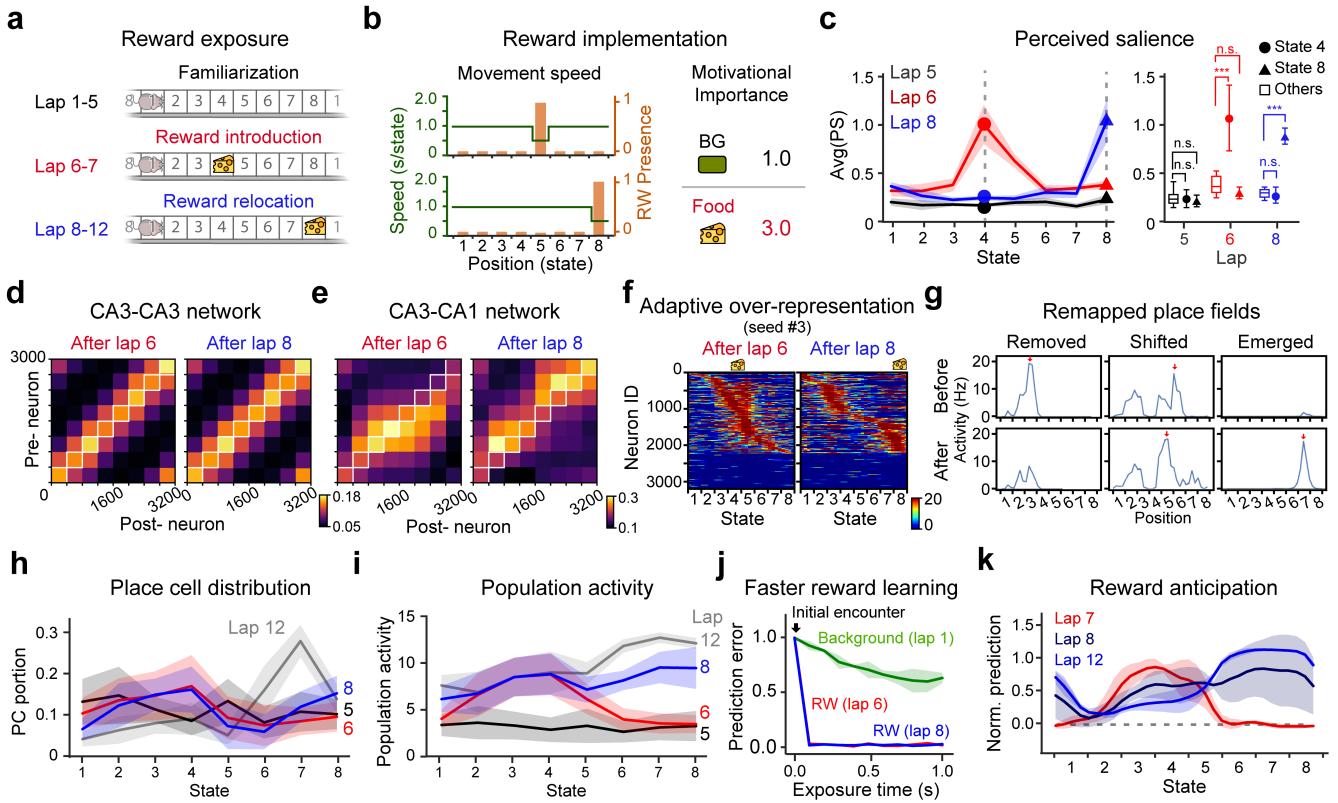
Finally, we examined whether the feature prediction network successfully learned the SFs. During online simulations, the eight units in this layer were trained using a simple local error correcting learning rule to predict the presence of the eight background features associated with the track states. As expected, the mean prediction error decreased across the five laps. This indicated that as CA1 developed place fields, the prediction layer increasingly relied on them to infer the corresponding features (Fig. 2l). By the end of the fifth lap, the feedforward weights from CA1 to the prediction units ( $W_{pred}$ ) had become state-specific, with CA1 place cells preferentially projecting to the feature unit associated with their place field (Fig. 2m). Consequently, each background feature unit exhibited strongest prediction when the animal occupied its corresponding location (Fig. 2n). Because CA1 place cell activity is asymmetric, the learned feature fields were also asymmetric: feature predictions rose rapidly as the animal approached the relevant state and decayed quickly as it moved away. Fitting the discount factors of approach and retreat phases confirmed this asymmetry (Fig. 2o).

### Salient features are adaptively encoded and updated in CA1 and feature prediction network

Next, we demonstrated that the SR encoded within the CA1 layer reflects the introduction and relocation of salient outcomes. To test this, we included a salient positive feature in the simulation corresponding to a food reward. After five familiarization laps without salient features, the food reward was placed at state 4 for two laps. The food was then relocated to state 8 for the five additional laps (Fig. 3a).

We treated the introduction of food reward as affecting two key factors that operated during exploration of the environment (Fig. 3b). First, the motivational importance of food was set to 3, compared with 1 for background states. This significantly increased the perceived salience of the food location (Fig. 3c). Second, the animal's travel speed was halved when passing through the food state to model consumption<sup>18,22</sup>.

We next examined whether the SR encoded within  $W_{CA1}$  captured the salience information introduced by the food. As expected, food exposure did not meaningfully alter  $W_{CA3}$  (Fig. 3d). In contrast,  $W_{CA1}$  adapted rapidly to changes in food location: after lap 6, weights projecting from CA3 neurons representing state 4 were selectively strengthened, and after lap 8, weights from neurons representing the new food location at state 8 increased sharply (Fig. 3e). As observed in empirical studies, CA1 place cells correspondingly reorganized to over-represent the vicinity of the food site (Fig. 3f,



**Figure 3.** **a**, Simulation scheme for introducing a salient reward. **b**, Two factors modulated by the introduction of reward: decreased movement speed and the motivational importance of the reward. **c**, Left: Average perceived salience of each state on laps 5, 6, and 15. The rewarded state shows higher perceived salience when the reward is present. Right: Perceived salience at state 4, 5 and the rest experienced during lap 5, 6 and 8. Paired t-test,  $N_{seed} = 10$ , n.s.  $P > 0.05$ , \*\*\* $P < 0.005$ . **d**, Weights of  $W_{CA3}$  after lap 6 (initial reward introduction) and lap 8 (reward relocation). Average value across 10 different seeds are plotted. **e**, Weights of  $W_{CA1}$  after lap 6 and 8. Average value across 10 different seeds are plotted. **f**, CA1 cell activity after lap 6 ad 8. Place cells clustered around the reward states, which representation adapts based on reward introduction and relocation. Average value across 10 different seeds are plotted. **g**, Example place field revisions after reward removal: emergence of new place cells (left), elimination of existing cells (middle), and shifts of field peaks (right). Red arrows indicate place field peaks. **h**, Number of place cells representing states on the track after lap 5, 6, 8 and 12. **j**, Novelty of the reward decreases much faster than for background feature 5 when first encountered at the same state. **k**, Predicted presence of the reward across track positions. All data are presented as mean±s.d.

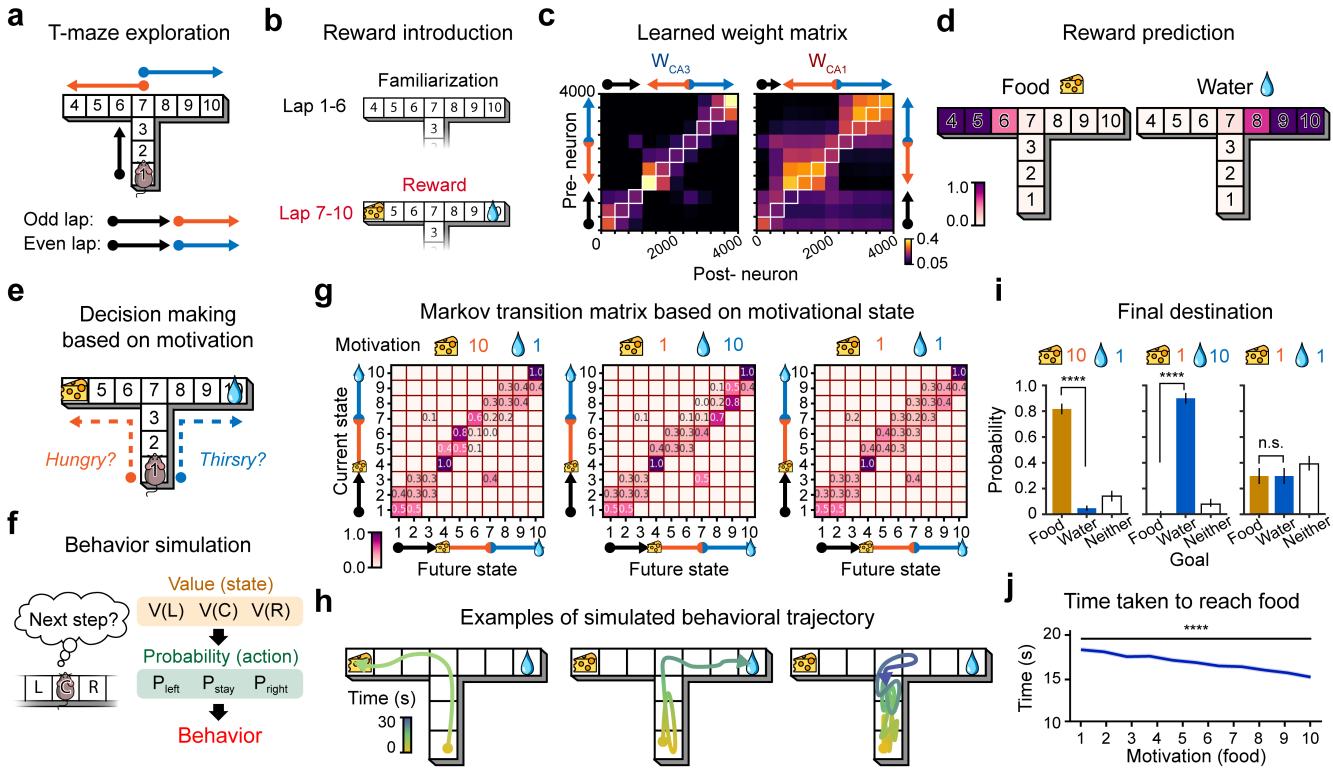
Supplementary Fig. 2). This population-level shift arose from diverse cell-specific adaptations, caused by adaptability of BTSP; the resulting connection strength of synapses trained by BTSP is modulated by the strength of plateau potentials (Supplementary Fig. 4). Thanks to these adaptive characteristics, some CA1 cells lost their place fields, others shifted to new positions, and some acquired place fields following the change in food location (Fig. 3g). Overall, the proportion of CA1 place cells representing each state changed dynamically with the food's location (Fig. 3h). As a result, average CA1 activity, or population activity, was stronger when passing through the salient state (Fig. 3i). This, in turn, accelerated learning of the associated feature. Indeed, the prediction error for the food decreased much faster upon first exposure compared to background features during initial exploration (Fig. 3j). Consequently, the network rapidly developed reliable predictions of the food. After two laps of initial reward exposure at state 4, the network showed maximum prediction of food at and before state 4. Following the relocation of the food to state 8, the network updated its prediction accordingly (Fig. 3k).

## The network recapitulates behavioral preferences based on motivational state

We next tested how the learned framework supports decision making based on an animal's transient motivational state. We trained the networks on a two-dimensional T-maze with two distinct arms. In every odd lap, the animal started at the stem (state 1) and moved toward the left arm (state 4), whereas in every even lap it moved toward the right arm (state 10) (Fig. 4a). After six laps (three exposures to each arm), two distinct reward features were introduced at the arm ends (food and water) and the animal was exposed to each reward twice (laps 7–10) (Fig. 4b).

After the lap 10,  $W_{CA3}$  encoded the structural layout of the T-maze, while  $W_{CA1}$  over-represented both arms, reflecting the presence of food and water (Fig. 4c). Correspondingly, the feature prediction layer strongly predicted each reward at its respective location (Fig. 4d).

We then simulated decision-making under different motivational needs for food and water (Fig. 4e). In each simulation, the animal started at the stem (state 1) and advanced one state per second. At each step, it evaluated all possible next states



**Figure 4.** **a**, T-maze learning scheme. **b**, After 6 familiarization laps, two distinct rewards (food and water) are introduced at each arm for 4 laps. **c**, Resulting  $W_{CA3}$  and  $W_{CA1}$  after lap 10. **d**, Predictions for food and water across maze positions. **e**, Animal's decision to pursue either arm of the maze based on its motivational state. **f**, Behavioral decision-making simulation scheme. **g**, Markov state transition matrices under three different motivational scenarios for food and water. **h**, Example simulated trajectories for each scenario. **i**, Goal reached within 30 seconds of movement for each scenario. Paired t-test,  $N_{seed} = 10$ ,  $****P < 0.0005$ , n.s.  $P = 0.98$ . Data are presented as mean $\pm$ s.d. **j**, Time to reach the goal as a function of motivational strength. Mann-Kendall test,  $N_{trial} = 500$ ,  $****P = 5.16 \times 10^{-5}$ ,  $Z = -4.05$ . Data are presented as mean $\pm$ s.e.m.

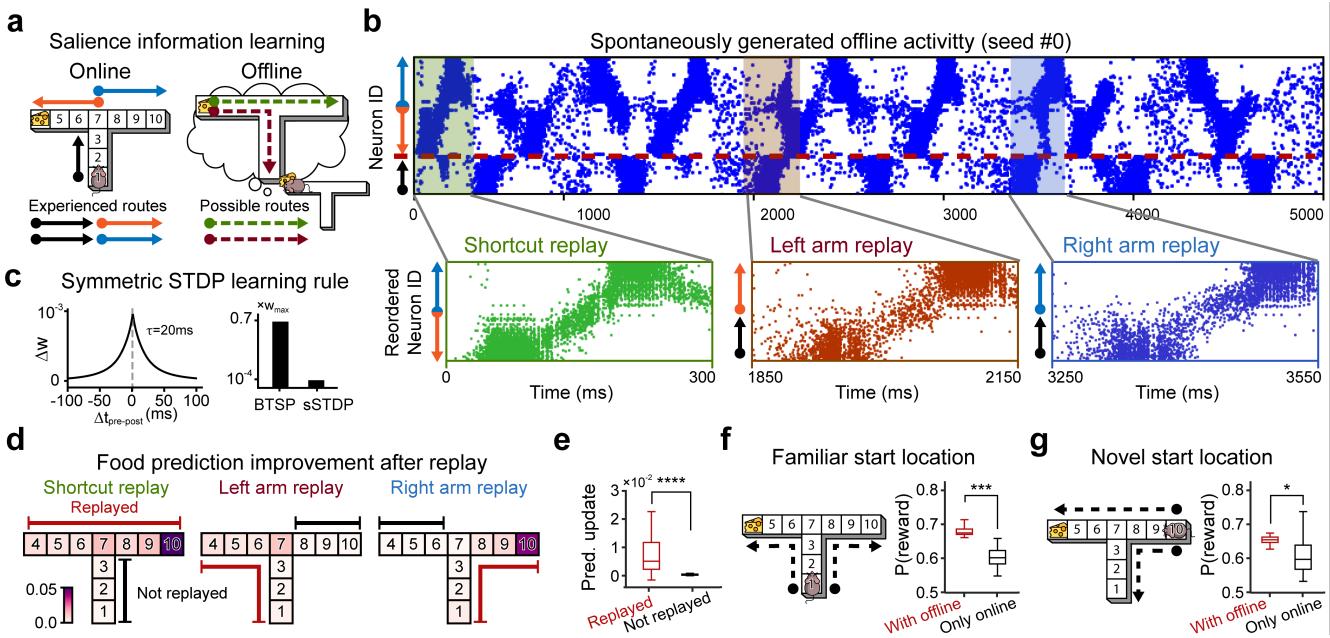
including neighboring states and the current state, and translated these values into movement probabilities using a softmax policy (Fig. 4f). Each behavioral trial lasted up to 60 seconds; if neither goal was reached, the trial ended.

We simulated three motivational conditions: (1) hungry (food need = 10, water need = 1), (2) thirsty (food need = 1, water need = 10), and (3) neutral (food need = 1, water need = 1). For each condition, we computed the Markov transition matrix (Fig. 4g). The three matrices notably differed at state 7, the intersection of the stem and both arms. When hungry, the animal was more likely to move toward the food arm (left), whereas when thirsty it preferred the water arm (right). Under neutral conditions, movements left, right, or back to the stem were roughly balanced. Accordingly, the behavioral simulations showed that the animal moved quickly toward food when hungry and toward water when thirsty (Fig. 4h). Across 500 simulation trials, the animal predominantly reached the food when hungry, the water when thirsty, and distributed its choices more evenly under neutral conditions (Fig. 4i). Finally, we found that stronger motivational drives accelerated goal-directed behavior: as the motivational need for food increased from 1 to 10 (with water fixed at 1), the time required to reach the food decreased (Fig. 4j).

### Offline replay propagates salience information to unexperienced paths

We next investigated how offline learning enhances the network's prediction of salient outcomes and improves behavior. Specifically, we hypothesized that replay activity can include trajectories not experienced during online exploration<sup>28</sup>, thereby propagating salience information across the maze to positions never directly visited (Fig. 5a). To test this, we simulated an offline phase corresponding to when an animal consumes a food reward by providing place input for state 4 to CA3 cells, then observed the resulting spontaneous network activity. Sequential activity of place cells spontaneously emerged (Fig. 5b, top). Notably, these replay events included not only previously experienced trajectories, such as left- and right-arm replays, but also a novel "shortcut replay" connecting the two arms that had never been physically traversed in sequence (Fig. 5b, bottom).

We then examined how each replay type contributed to propagating salience information from the reward location to distant states. To do so, we updated the network using a symmetric STDP rule<sup>33</sup>, which adjusts synaptic weights after every spike but with smaller magnitude than BTSP (Fig. 5c). After a single shortcut replay event, food anticipation increased, indicating that the food could now be predicted at previously unconnected positions (Fig. 5d, left). Similarly, left-arm re-



**Figure 5.** **a**, Propagation of reward information through online and offline learning. **b**, Spontaneously generated offline activity. Replay events are grouped into three categories based on their trajectories: shortcut, left-arm, and right-arm. **c**, Left: sSTDP rule used in offline learning. Right: Largest possible weight update of BTSP and sSTDP rules. **d**, Normalized shock prediction improvement after replay. **e**, Normalized shock prediction update of states that were replayed (red) and those that were not replayed (black). Independent t-test,  $N_{replayed} = 70$  vs.  $N_{notreplayed} = 30$ ,  $****P = 2.28 \times 10^{-4}$ . **f**, Behavioral simulation results when the animal starts from the same location as during the online phase. Paired t-test,  $N_{seed} = 10$ ,  $***P = 3.72 \times 10^{-3}$ . **g**, Behavioral simulation results when the animal starts from a different location than in the online phase. Paired t-test,  $N_{seed} = 10$ ,  $*P = 0.01$ . All data are presented as mean $\pm$ s.d.

plays enhanced food prediction along the replayed trajectory, including the left arm and stem (Fig. 5d, middle). By contrast, right-arm replays had minimal effect, as their trajectories did not include the food location (Fig. 5d, right). Overall, each replay type improved food prediction across the states it traversed relative to non-traversed states (Fig. 5e).

We next tested how this replay-induced propagation of salience affected behavior. In the first scenario, the food reward was positioned at the left arm (state 4), and the animal started at the stem (state 1) with a motivational need of 5.0 (Fig. 5f, left). Each trial ended when the animal reached either end of the maze or failed to reach a goal within 30 steps. We found that including offline learning significantly increased the probability of reaching the food goal (Fig. 5f, right).

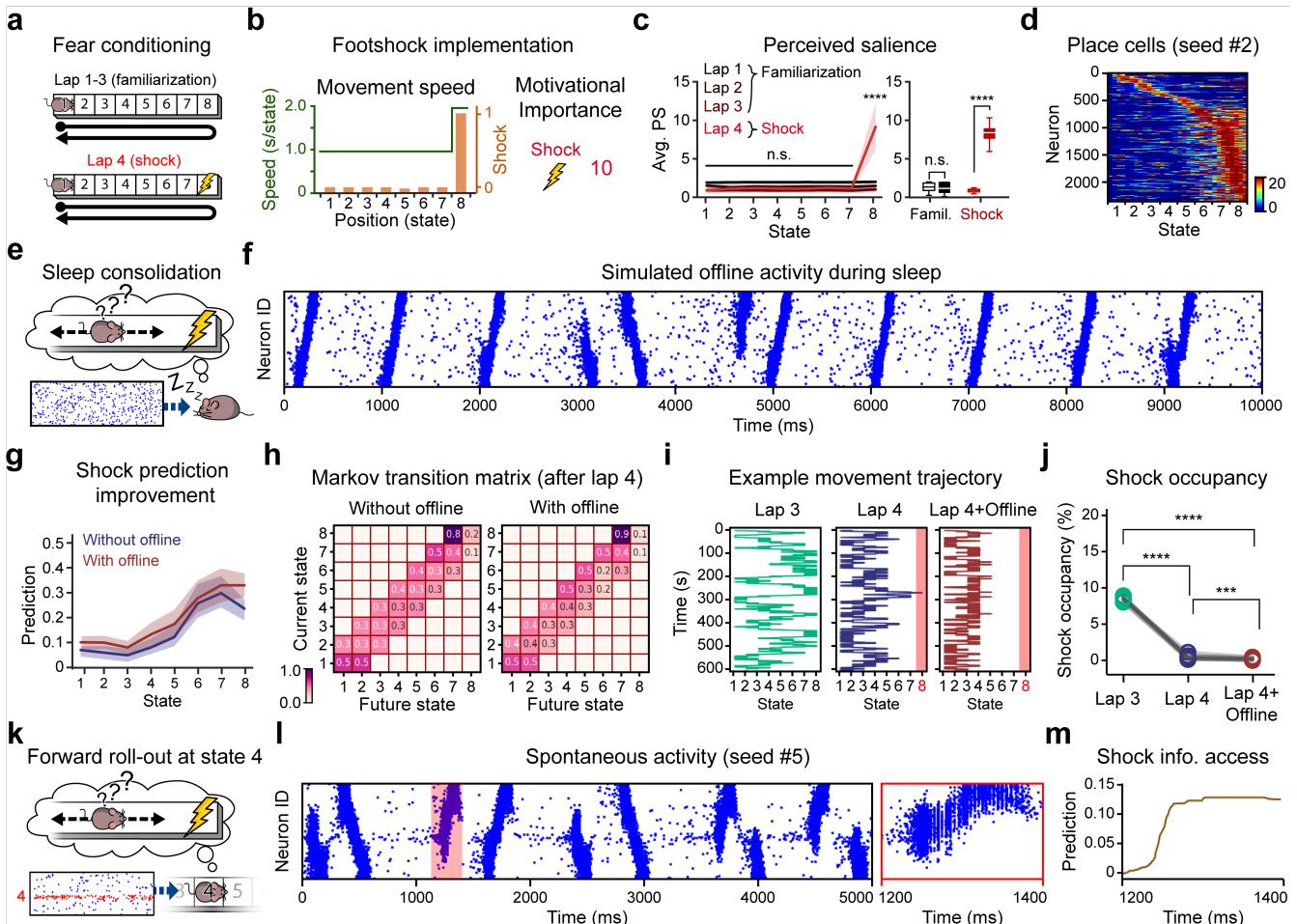
Strikingly, offline learning also enhanced adaptability when starting from a novel location. We simulated a paradigm where the animal began at the right end of the maze (state 9), choosing between two possible goals (state 4 and state 1) with the food still located at state 4 (Fig. 5g, left). The inclusion of an offline learning phase again increased the likelihood of reaching the reward within the time limit (Fig. 5g, right). We further demonstrated that the stronger spatial input during the offline phase increases the frequency of replay events (Supplementary Fig. 5). The intensity of the spatial input during the offline phase can possibly be modulated by perceived salience, as proposed by previous studies<sup>37</sup>.

### The model learns aversive outcomes, predicts their presence through forward rollouts, and exhibits avoidance behavior

We next asked whether the model could also account for the learning of aversive features. This differs from reward learning (Fig. 3–5) because animals spend minimal time at aversive sites, limiting the opportunity to develop over-representation through prolonged dwell time. Moreover, "on-site" replay, where salience information is propagated during offline learning while the animal remains at the location, is not possible. At the behavioral level, even a single exposure to a negative outcome typically repels the animal, reducing future visits and further constraining learning.

To simulate this, we followed the design of a previous study<sup>9</sup>, in which mice freely explored a linear track before encountering a shock at the track end. In our simulation, the animal completed three laps without salient outcomes (from state 1 to 8 and back), followed by a fourth lap in which a shock occurred at state 8 (Fig. 6a). During the shock, the animal's speed doubled and the motivational importance of the shock was set to 10 (Fig. 6b). Compared to familiarization laps, perceived salience sharply increased at state 8 during shock exposure (Fig. 6c), inducing over-representation of that location despite the short dwell time (Fig. 6d).

We then simulated an offline phase in which the animal was removed from the environment by providing random, low-intensity input to CA3 cells for 10 seconds (Fig. 6e). As a



**Figure 6.** **a**, Aversive feature learning scheme. **b**, Two factors modulated by the introduction of aversive outcome: increased movement speed and the motivational importance of the aversive outcome. **c**, Average perceived salience. Perceived salience of each state on lap 1 to 4. The states without salient outcome does not show significant difference across different laps, whereas the perceived salience of the punished state significantly increased after the introduction of the footshock. Left; Paired t-test, mean perceived salience of the first three laps vs. fourth lap,  $N_{seed} = 10$ ,  $****P = 0.01$ . Right; Independent t-test, perceived salience of the non-salient states vs. perceived salience of the salient state,  $N_{seed} = 10$ , n.s. $P = 0.40$ ,  $****P = 2.16 \times 10^{-34}$ . **d**, CA1 place cells clustered around the punished state, over-representing it. **e**, Offline simulation at state 4. **f**, Example spontaneously generated offline activity (10 s). **g**, Elevated anticipated shock stimulus after the offline learning phase. **h**, Markov state transition matrices before and after the offline learning. **i**, Simulated movement trajectories before shock introduction (left), after shock introduction (middle), and after offline learning (right). **j**, Shock state occupancy of each scenario. Paired t-test,  $N_{seed} = 10$ ,  $***P = 3.42 \times 10^{-3}$ ,  $****P < 2.3 \times 10^{-13}$ . **k**, Forward roll-out at state 4, simulated by inputting state 4-specific input. **l**, Left: Example spontaneously generated offline activity (5 s). Right: Zoomed-in roll-out activity. **m**, Shock presence prediction output from feature-prediction layer during the roll-out activity. All data are presented as mean $\pm$ s.d.

result, spontaneous replay-like activity emerged, spanning the entire track (Fig. 6f), upon which the network was updated using the symmetric STDP rule<sup>33</sup>.

Because exposure to aversive stimuli is brief and rarely repeated, online learning alone was insufficient to produce reliable shock prediction. Shock presence was predicted less strongly at the actual shock site (state 8) than at a neighboring state (state 7) (Fig. 6g, blue). Incorporating offline learning resolved this deficit: shock prediction increased at nearby states, with the strongest response at the true shock location (Fig. 6g, brown). Markov transition matrices showed that improved prediction reduced the probability of entering not only the shock state but also adjacent states (state 5–7) (Fig. 6h). Simulating 600 seconds of free movement revealed that offline learning decreased the animal’s tendency to approach the shock zone

(Fig. 6i). This effect was quantified as shock-state occupancy, defined as the proportion of time spent in the shock state. We showed that shock-state occupancy dropped sharply after shock introduction and decreased further following offline learning (Fig. 6j).

A previous study reported that, after shock learning, mice paused midway along the track and exhibited forward-directed replay events toward the shock zone<sup>9</sup>. We hypothesized that such "forward roll-outs" allow the animal to mentally access information about the shock without physically entering it. To test this, we simulated an offline phase with place input from the middle of the track (state 4) (Fig. 6k). The network spontaneously generated sequential place cell activity extending toward both ends of the track (Fig. 6l, left), including forward roll-out events that included the shock state (Fig. 6l, right).

Consistent with our hypothesis, internal shock prediction increased during these roll-out events (Fig. 6m).

Together, these results show that aversive stimuli can be learned reliably despite minimal exposure and suggest potential roles for both off-site and on-site replay in supporting avoidance behavior.

## Discussion

In this study, we proposed a theoretical framework of the cognitive map and decision making, bridging from synaptic level findings to behavioral findings. Specifically, we showed how the framework can be implemented in the hippocampal network: CA3 encodes a salience-neutral representation of environmental structure, CA1 integrates this input with salience information to form a salience-weighted successor representation, and a feature-prediction layer uses this to estimate successor features. These features can then be combined with temporally associated values to compute the overall value of available choices and guide behavior. This framework can be acquired through BTSP during online exploration and further strengthened by STDP during offline replay, even in the absence of informative external input. Importantly, it was updated adaptively when environmental contingencies changed, such as when rewards were relocated. We also showed how this framework explains a range of behaviors reported in experiments, providing a theoretical grounding that links synaptic mechanisms to network-level representations and decision making. In particular, it accounts for the rapid learning of salient environmental features, regardless of valence, and for behavioral flexibility driven by transient motivational states.

A key innovation of our framework is the separation of salience and value in predictive representations. This has two advantages relative to traditional accounts. First, our model explains how hippocampal circuits over-represent states associated with salient features<sup>19–22</sup>, thereby facilitating faster learning. The mechanisms driving this over-representation may differ for rewarding versus aversive stimuli: in our simulations, positive outcomes were associated with moderate salience relative to neutral states but longer dwell times, whereas negative outcomes had even higher salience but shorter exposures. Second, our model can flexibly incorporate transient motivational signals<sup>10</sup>, because our model predicts features rather than states *per se*. This enables rapid adjustments in behavioral preference according to internal states, such as hunger or thirst, without requiring relearning of the environment's structure.

Our findings posit distinct computational roles for CA3 and CA1 subregions of the hippocampus. First, CA3 encoded the objective structure of the environment, yielding a representation that was generalizable and reusable even when salience information shifted<sup>38</sup>. It further supported attractor dynamics<sup>32</sup>, enabling rapid emergence of replay during offline phases<sup>35,39</sup>. In contrast, CA1 encoded more transient, salience-related, and subjective information, with representations that adapted rapidly to reward relocation and other transient changes<sup>11,38</sup>. This division is consistent with some

experimental findings showing that CA1 place fields are more plastic and context-dependent<sup>17</sup>. However, it remains possible that learning in CA3, along with CA1, can also be affected by salience related factors, and such influences could be incorporated if substantiated in future research.

Our simulations also provide insight into the long-debated function of hippocampal replay. We found that offline replay consolidates and strengthens online learning<sup>25,26</sup>, while also filling in gaps not directly experienced during exploration<sup>30</sup>. Shortcut replays<sup>28</sup>, for example, transmitted salience information to states that were never sequentially paired with salient features, thereby improving prediction at otherwise unrelated locations. Beyond strengthening networks, replay may support value computation by transiently activating representations of non-current states<sup>9</sup>. This enables prediction of features and estimation of state values necessary for planning<sup>34,40,41</sup>. Nevertheless, important questions remain. The relative importance of replay for strengthening newly acquired salient memories<sup>37</sup>, maintaining other memories to protect them from interference during new learning<sup>42,43</sup>, and for planning and decision making remain ongoing topics for discussion and debate. In addition, regarding replay content and frequency, our model hinted that increased salience signals can enhance replay frequency, but we did not explore how replay content is selected or modulated<sup>10,44</sup>. Future work should investigate these issues in greater detail.

Our framework also has implications for artificial intelligence. Current AI systems often require large amounts of data and struggle with generalization, and this may arise in part from the learning procedure they use, which may contribute to their failure to extract and reuse structural regularities across tasks. Our work was motivated in part by seeking solutions to these issues in AI systems; rapid on-line learning through BTSP and subsequent offline replay may provide alternatives to the gradient based learning methods used in AI and may help address AI's limitations<sup>45,46</sup>. When these ideas are coupled with a framework explicitly separating learning about environmental states from associating them with their features and assigning value to them<sup>31</sup>, our model demonstrated both one-trial learning and flexibility under changes of environmental and internal state variables. Replay during offline periods further increased the scope and generalizability of learning without additional experience and mitigating catastrophic forgetting<sup>42,43</sup>. In the T-maze simulation, for example, the network simulated an analog of the finding from a previous work<sup>28</sup> by traversing unexplored pathways, capturing the T-shaped structure and connectivity of the arms. This illustrates how the mechanisms we have explored, when taken together, can allow limited experiences to be integrated to fill structural gaps. Furthermore, replay in our model emerged spontaneously during offline phases, without supervisory input, and produced activity that strengthened previously acquired knowledge. This suggests a path toward AI systems that autonomously improve their internal representations during "rest" periods, thereby achieving greater efficiency and

adaptability.

Taken together, we propose an overarching framework for the cognitive map in which BTSP-based online learning and offline replay jointly support adaptive prediction and decision making. We showed how this framework applies to different behavioral tasks, including T-maze navigation and shock avoidance. By connecting synaptic learning rules to network-level representations and behavior, our work provides a plausible, efficient, and adaptive account of how animals learn, generalize, and make decisions in dynamic environments.

## Methods

### Neural network model

#### *The hippocampal network*

We modeled the hippocampal network as a two-layer spiking neural network, corresponding to the CA3 and CA1 subregions of the hippocampus. Consistent with experimental findings, the CA3 layer included recurrent connectivity ( $W_{CA3}$ ) and projected feedforward output to CA1 ( $W_{CA1}$ ). Synaptic weights in both pathways were constrained between 0 and 10, initialized from a Gaussian distribution (mean = 0.01, standard deviation =  $10^{-4}$ ), and formed with a fixed connection probability of 0.1.

Both CA3 and CA1 layers contained the same number of excitatory neurons, determined by the number of states in the simulated maze. Each state was represented by 400 neurons, resulting in 3,200 excitatory neurons per layer for the linear track (8 states) and 4,000 neurons per layer for the T-maze (10 states). CA3 excitatory neurons were designated as place cells, with uniformly distributed place fields and spatial inputs reflecting the animal's position during online exploration. CA1 excitatory neurons did not receive direct spatial input; instead, their place-selective responses emerged through learned  $W_{CA1}$ .

Inhibitory modulation differed between online and offline phases. During online exploration, a uniform feedback inhibition signal was applied to all excitatory neurons within each layer to modulate the probability of plateau potential (PP) events. During the offline replay phase, 150 inhibitory interneurons were explicitly included in each layer to support spontaneous replay dynamics.

#### *Feature prediction network*

The feature prediction network aimed to predict the physical presence of environmental features, including background, food, water, and shock. It comprised units representing the predicted presence ( $\psi$ ) of each feature. These units received fully connected input from CA1 neurons via  $W_{pred}$  and sent a novelty signal ( $N$ ) back to CA1, modulating the likelihood of plateau potentials in CA1 cells. We interpret this network as a simplified representation of brain circuits involved in feature prediction and value computation, such as the nucleus accumbens (NAc), which forms a recurrent loop with CA1 and relays dopaminergic signals.

### Online exploration

We simulated rodent's maze learning task which is commonly studied in experimental literature. On each lap, the animal followed a sequence of movements, and the hippocampal and feature prediction networks were trained accordingly.

#### *Maze exploration scenarios*

We simulated three different maze learning scenarios:

- **An animal unidirectionally navigating a linear-track treadmill<sup>18</sup>:** Each complete rotation of the belt corresponded to a 2.4 m lap. For convenience, the track was discretized into 8 states (state 1 to 8, from left to right), each 30 cm in length. In each lap, the animal began at the leftmost state (state 1), traversed to the rightmost state (state 8), reached the end of the belt, and was returned to the starting point. Formally,

$$\mathcal{S}_{lap} = \langle 1, 2, \dots, 8 \rangle.$$

The next lap began immediately after the animal returned to state 1.

- **An animal navigating a T-maze<sup>10</sup>:** The maze consisted of a 0.9 m vertical stem (state 1 to 3, from bottom to top), connected at its midpoint to a 2.1 m horizontal arm (state 4 to 10, from left to right, with state 7 as the intersection). In each lap, the animal began at the base of the stem (state 1), moved up to the center of the horizontal arm (state 7), and then turned left in odd-numbered laps (ending at state 4) or right in even-numbered laps (ending at state 10).

Formally,

$$\mathcal{S}_{lapN} = \begin{cases} \langle 1, 2, 3, 7, 6, 5, 4 \rangle, & \text{if } N \text{ is odd,} \\ \langle 1, 2, 3, 7, 8, 9, 10 \rangle, & \text{if } N \text{ is even.} \end{cases} \quad (7)$$

- **An animal round-tripping a linear maze<sup>9</sup>:** This scenario replicates an animal exploring a confined, 2.4 m segment of linear maze. The track was discretized into 8 states (state 1 to 8, from left to right), but unlike the linear treadmill, state 8 and state 1 are not connected. A single lap began with the animal at the left end (state 1), moving sequentially to the right end (state 8), and then returning back to the start (state 1). Formally,

$$\mathcal{S}_{lap} = \langle 1, 2, \dots, 7, 8, 7, \dots, 1 \rangle.$$

#### *Environmental features*

In our simulation, states were associated with distinct environmental features,  $f$ . Environmental features are accompanied with their physical intensity  $I_f$  as well as motivational importance  $MI_f$ . The presence and absence of the features at a certain state  $s$  were encoded as a presence vector  $P(s)$ . Although real maze exploration environments contain many features, we simplified this to capture essential elements relevant to our work: background, food, water and shock.

- **Background:** Each state involves state-specific background feature, which we thought of as sensory cues that enables the animal to distinguish the state from other ones. Since this feature is not attached to any affective value, we set the *MI* of this feature to be 1. When the state only has background but not other significant environmental features, the animal moves at the baseline speed, taking  $T = 1$  second.
- **Food and water:** These features serve as a positive outcome in our simulation, and has *MI* of 3. When the animal was passing the state with these features, it slowed to half the baseline speed, simulating an extended stay during reward consumption ( $T = 2$  s).
- **Shock:** This feature serves as a negative outcome, with *MI* of 10. When the animal passes a state involved with shock, it sped up to twice the baseline speed, reflecting an aversive response ( $T = 0.5$  s).

In our simulation, we didn't vary the physical intensity much. Simply,  $I_f$  was set to be 2 if  $f$  is present, and 1 if it is absent. i.e., in our simulation,  $I_f = P_f + 1$ .

Another component attached to environmental features is novelty, denoted as  $N_f(t)$ . This quantity tracks the animal's subjective surprise associated with feature  $f$ , and defined as the difference between the predicted presence and actual presence of the feature. The novelty values were updated at each timestep based on the absolute prediction error computed by the feature prediction network,  $\epsilon$ , as described in the following subsection.

$$N_f(t) = |\epsilon_f(t)| \quad (8)$$

#### Computation of perceived salience

The perceived salience (PS) of an environmental feature  $f$ , denoted as  $PS_f$ , represents the animal's subjective evaluation of how noticeable or memorable  $f$  is. In our simulations, PS was modulated by three factors: motivational importance *MI*, intensity  $I$ , and the novelty  $N$  of the environmental feature. Specifically, the perceived salience of feature  $f$ ,  $PS_f$ , was then computed as:

$$PS_f(t) = I_f \cdot MI_f \cdot N_f(t) \quad (9)$$

The perceived salience of a state at time  $t$ ,  $PS(t)$ , was computed as the sum of PS across all environmental features.

#### Spike train generation and activity level estimation

CA3 place fields were uniformly distributed along the  $x$ - and  $y$ -axes, with each field centered at a distinct position. Place fields were circular with a radius of 30 cm.

Spiking activity in CA3 excitatory neurons arose from two sources: external spatial input from the entorhinal cortex and recurrent input from other CA3 neurons. Firing driven by spatial input ( $FR_{place}$ ) depended on the animal's Euclidean distance from the center of a neuron's place field and followed

a Gaussian profile. The maximum value of  $FR_{place}$  was set to 20 Hz and the standard deviation of the Gaussian was chosen to be the radius of the field.

Upstream input contributed an additional firing component ( $FR_{recurrent}$ ), calculated as the sum of activities from all connected presynaptic neurons, weighted by  $W_{CA3}$ . This input was then passed through a sigmoid nonlinearity to yield the final recurrent input-driven firing rate for each neuron. The sigmoid function was defined as:

$$\sigma(x, a, b) = \frac{1}{1 + \exp(-b(x - a))}, \quad (10)$$

where  $a = 4$  and  $b = 3$ .

Finally, the total firing rate of a CA3 neuron  $i$  was given by:

$$FR_i = FR_{place,i} + FR_{recurrent,i}. \quad (11)$$

CA1 excitatory neurons were modeled similarly but did not receive direct spatial input. Their firing rates were determined exclusively by feedforward input from CA3: the activity of CA3 presynaptic neurons was multiplied by  $W_{CA1}$  and passed through the same sigmoid function with  $a = 4$  and  $b = 2$  to compute  $FR_{feedforward}$ . Therefore, the total firing rate of a CA1 neuron  $j$  is:

$$FR_j = FR_{feedforward,j}. \quad (12)$$

After calculating  $FR$ s of all neurons in both CA3 and CA1, we generated spike trains for each neuron using inhomogeneous poisson processes. These spike trains were then filtered to remove spikes that violated the 5 ms refractory period.

Spike trains for both CA3 and CA1 were generated at every  $\Delta t_{spike}$  interval, which was set to 100 ms. For CA3 spike train generation, the input activity level was determined from the CA3 spike train generated in the previous time block.

To quantify neuronal activity, we calculated the actual average firing rate of each neuron by dividing the total number of generated spikes by  $\Delta t_{spike}$ . Note that this differs from the ideal  $FR$ , as it reflects post-processing by both the inhomogeneous Poisson process and the refractory period filter.

#### Training the feature prediction network using the delta rule

The feature prediction network was trained during the online exploration phase using a delta rule. Specifically, a synaptic weight  $w_{fi}$  connecting a CA1 neuron  $i$  and a feature unit  $f$ , was updated as:

$$\Delta w_{fi}(t) = \mathcal{A}_i(t) \cdot \epsilon_f(t) \quad (13)$$

where  $\mathcal{A}_i(t)$  is the activity of neuron  $i$  and  $\epsilon_f(t)$  is the prediction error of feature  $f$  at time  $t$ <sup>8</sup>. The prediction error was defined as the error between predicted presence ( $\psi$ ) and the physical presence ( $\phi$ ) for that feature:

$$\epsilon_f(t) = \psi_f(t) - \phi_f, \text{ where} \quad (14)$$

where

$$\psi_f(t) = \sum_i \mathcal{A}_i(t) \cdot w_{ji}(t) \quad (15)$$

Note that the absolute value of the prediction error of a feature is the novelty  $N$  associated with the feature.

### **Training the hippocampal network using BTSP**

Synaptic weight updates in the hippocampal network followed a BTSP rule, which depends on the interaction between presynaptic eligibility traces (ETs) and postsynaptic instructive signals (ISs). ETs were computed based on the spiking activity of presynaptic neurons, while ISs were driven by the postsynaptic neuron's dendritic plateau potentials (PPs).

ETs were calculated only for CA3 neurons, as these served as the sole presynaptic population in our model. When a presynaptic spike occurred, the neuron's ET increased by a fixed amplitude,  $A_{ET}$ , and then decayed exponentially with a time constant  $\tau_{ET}$ .

ISs were evoked by plateau potentials (PPs), modeled as stochastic events occurring in both CA3 and CA1 neurons. Each neuron  $k$  was assigned a PP probability  $p_{plateau,k}$ , initialized to a basal value  $p_{basal}$  and updated every  $\Delta t_{spike}$ . In our simulations,  $p_{basal}$  was set to 0.05. At each 1 ms time step, a Bernoulli trial was performed using  $p_{plateau,k}$  to determine whether neuron  $k$  initiated a PP. The probability of PP occurrence,  $p_{plateau,k}$ , was enhanced by the neuron's recent activity and the perceived salience ( $PS$ ) of the current state, but decreased with normalized population activity  $\mathcal{A}_{norm}$ , defined as the mean activity across neurons within the layer<sup>21,22</sup> (Supplementary Fig.1h). Formally,

$$p_{plateau,k} = \mathcal{A}_k \cdot c_{norm}, \quad (16)$$

where

$$c_{norm} = \begin{cases} p_{min} + 2p_{basal} \cdot \sigma(\mathcal{A}_{target} - \mathcal{A}_{norm}, 0, 5), & \text{if } \mathcal{A}_{target} > \\ & \text{otherwise.} \\ 0, & \end{cases} \quad (17)$$

When a PP occurred, the postsynaptic neuron's instructive signal increased by  $A_{IS}$  and decayed exponentially with a time constant  $\tau_{IS}$ . BTSP-mediated synaptic updates occurred only during PP events. The BTSP update of the weights is a function of  $ET \times IS$  as well as the initial value of the weight<sup>21</sup>. Initially weak weights undergo stronger potentiation even with smaller  $ET \times IS$ , whereas the weights that are already strong enough undergo depression even with stronger  $ET$  and  $IS$  signal. Formally,

$$\Delta w_{ji} = (w_{max} - w_{ji}) k_{pos} c_{pos} - w_{ji} k_{neg} c_{neg}, \quad (18)$$

$$c_{pos} = \frac{\sigma(ET_i \cdot IS_j; a_{pos}, b_{pos}) - \sigma(0; a_{pos}, b_{pos})}{\sigma(1; a_{pos}, b_{pos}) - \sigma(0; a_{pos}, b_{pos})}, \quad (19)$$

$$c_{neg} = \frac{\sigma(ET_i \cdot IS_j; a_{neg}, b_{neg}) - \sigma(0; a_{neg}, b_{neg})}{\sigma(1; a_{neg}, b_{neg}) - \sigma(0; a_{neg}, b_{neg})}. \quad (20)$$

In our simulations, we set  $k_{pos} = 0.085$ ,  $k_{neg} = 0.01$ ,  $a_{pos} = 0.8$ ,  $b_{pos} = 6$ ,  $a_{neg} = 0.05$ , and  $b_{neg} = 44.44$ .

### **Offline phase**

The offline simulation framework was adapted from Ecker et al.<sup>33</sup>. Here, neurons were modeled using adaptive exponential integrate-and-fire dynamics. Hippocampal weights learned during online exploration ( $W_{CA3}$  and  $W_{CA1}$ ) were carried over into the offline phase, with additional connectivity to inhibitory interneurons introduced. Spontaneous activity emerging in the CA3 and CA1 layers during this phase was then used to update  $W_{CA3}$  and  $W_{CA1}$  according to a symmetric STDP rule.

### **Connectivity optimization**

Additional connectivity was introduced and optimized using an evolutionary algorithm, including: entorhinal cortex input to CA3 excitatory cells ( $EC \rightarrow Ext_{CA3}$ ); CA3 excitatory cell to CA3 inhibitory interneuron ( $Ext_{CA3} \rightarrow Inh_{CA3}$ ); interneuron–interneuron ( $Inh_{CA3} \rightarrow Inh_{CA3}$ ); and inhibitory interneuron to excitatory cell ( $Inh_{CA3} \rightarrow Ext_{CA3}$ ). The same connectivity types were established in CA1 ( $Ext_{CA1} \rightarrow Inh_{CA1}$ ,  $Inh_{CA1} \rightarrow Inh_{CA1}$ ,  $Inh_{CA1} \rightarrow Ext_{CA1}$ ).

The optimization algorithm minimized a loss function designed to match network-level physiological targets: maintaining realistic population firing rates in excitatory cells, suppressing gamma-band (30–100 Hz) oscillations, and enhancing ripple-band oscillations in excitatory cell populations.

During the offline phase, spontaneous activity was simulated in both CA3 and CA1, with low-rate entorhinal cortex input to CA3.  $W_{CA3}$  and  $W_{CA1}$  were updated using a symmetric STDP rule throughout this period.

The detailed description of spiking neural network model along with its parameters are described in supplementary materials.

### **Training the hippocampal network using symmetric STDP**

Symmetric STDP was used to update network using the spike train generated during the offline phase. It only involves potentiating the connections which connect neurons that fire together. Precisely, the update rule we used is described as follows, with  $A_{STDP} = 1 \times 10^{-3}$  and  $\tau_{STDP} = 20ms$ .

$$\Delta w = A_{STDP} \cdot e^{|t_{post} - t_{pre}| / \tau_{STDP}} \quad (21)$$

### **Behavioral simulation**

After the network learned the environment including its spatial structure and associated features, we evaluated its performance

in behavioral decision-making tasks. This phase corresponds to the exploitation stage in reinforcement learning, during which the agent uses previously acquired knowledge to make optimal choices.

### **Decision-making scenarios**

We simulated two behavioral decision-making scenarios: one involving a positive outcome in a T-maze and another involving an aversive outcome in a linear track.

- **T-maze with two goal locations and distinct rewards<sup>10</sup>:** Each trial began with the animal positioned at the stem of the maze (state 1) and ended when it reached either goal state (state 4 or 10). Food was located at the left end (state 4) and water at the right end (state 10), and the animal's motivational needs for these two rewards were systematically varied.

To test the network's adaptability to different starting conditions, additional simulations were conducted in which the animal started from the right end (state 10) and traversed either toward the left end (state 4) or the stem (state 1), with food placed at state 4.

In all cases, the animal was allowed a maximum of 30 steps per trial. Trials were terminated if the animal failed to reach a goal within this limit, and such cases were labeled as "failure to find the goal."

- **Linear track with an aversive outcome<sup>9</sup>:** Each trial began with the animal positioned at the left end of the track (state 1) and allowed to move freely for 600 steps. A shock stimulus was positioned at the right end of the track (state 8).

### **Value computation and its usage in decision-making**

At each time step, the animal evaluated the value of neighboring states and selected its next action according to a softmax policy. The value of a given state  $s$ , denoted  $V(s)$ , was computed as:

$$V(s) = \sum_f \psi_f(s) \cdot v_f, \quad (22)$$

where  $\psi_f(s)$  represents the predicted presence of feature  $f$  at state  $s$ , and  $v_f$  denotes the animal's subjective value of that feature. The feature values  $v_f$  were dynamically modulated by the animal's transient motivational state. Given these computed state values, transition probabilities were derived using a softmax function, and the environment was treated as a Markov chain with known state transitions.

### **Acknowledgement**

The first author is supported by the Emergent Cognitive Functions Fund, Department of Psychology, Stanford University.

## **References**

1. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208, DOI: [10.1037/h0061626](https://doi.org/10.1037/h0061626) (1948).
2. Behrens, T. E. J. *et al.* What is a cognitive map? organizing knowledge for flexible behavior. *Neuron* **100**, 490–509, DOI: <https://doi.org/10.1016/j.neuron.2018.10.002> (2018).
3. Rosenberg, M., Zhang, T., Perona, P. & Meister, M. Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *eLife* **10**, e66175, DOI: [10.7554/eLife.66175](https://doi.org/10.7554/eLife.66175) (2021).
4. Dayan, P. Improving generalization for temporal difference learning: The successor representation. *Tech. Rep.* (1993).
5. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653, DOI: [10.1038/nn.4650](https://doi.org/10.1038/nn.4650) (2017).
6. Mehta, M. R., Quirk, M. C. & Wilson, M. A. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* **25**, 707–715, DOI: [10.1016/S0896-6273\(00\)81072-7](https://doi.org/10.1016/S0896-6273(00)81072-7) (2000).
7. Gustafson, N. J. & Daw, N. D. Grid cells, place cells, and geodesic generalization for spatial reinforcement learning. *PLOS Comput. Biol.* **7**, e1002235, DOI: [10.1371/journal.pcbi.1002235](https://doi.org/10.1371/journal.pcbi.1002235) (2011).
8. Gershman, S. J. *et al.* Explaining dopamine through prediction errors and beyond. *Nat. Neurosci.* **27**, 1645–1655, DOI: [10.1038/s41593-024-01705-4](https://doi.org/10.1038/s41593-024-01705-4) (2024).
9. Wu, C.-T., Haggerty, D., Kemere, C. & Ji, D. Hippocampal awake replay in fear memory retrieval. *Nat. Neurosci.* **20**, 571–580, DOI: [10.1038/nn.4507](https://doi.org/10.1038/nn.4507) (2017).
10. Carey, A. A., Tanaka, Y. & van der Meer, M. A. A. Reward revaluation biases hippocampal replay content away from the preferred outcome. *Nat. Neurosci.* **22**, 1450–1459, DOI: [10.1038/s41593-019-0464-6](https://doi.org/10.1038/s41593-019-0464-6) (2019).
11. Krishnan, S., Heer, C., Cherian, C. & Sheffield, M. E. Reward expectation extinction restructures and degrades ca1 spatial maps through loss of a dopaminergic reward proximity signal. *Nat. Commun.* **13**, DOI: [10.1038/s41467-022-34465-5](https://doi.org/10.1038/s41467-022-34465-5) (2022).

12. Kutlu, M. G. *et al.* Dopamine release in the nucleus accumbens core signals perceived saliency. *Curr. Biol.* **31**, 4748–4761.e8, DOI: [10.1016/j.cub.2021.08.052](https://doi.org/10.1016/j.cub.2021.08.052) (2021). Doi: 10.1016/j.cub.2021.08.052.
13. 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. *The J. Neurosci.* **21**, 1635, DOI: [10.1523/JNEUROSCI.21-05-01635.2001](https://doi.org/10.1523/JNEUROSCI.21-05-01635.2001) (2001).
14. Gauthier, J. L. & Tank, D. W. A dedicated population for reward coding in the hippocampus. *Neuron* **99**, 179–193.e7, DOI: <https://doi.org/10.1016/j.neuron.2018.06.008> (2018).
15. Mamad, O. *et al.* Place field assembly distribution encodes preferred locations. *PLOS Biol.* **15**, e2002365– (2017).
16. Nikbakht, N. *et al.* Efficient encoding of aversive location by ca3 long-range projections. *Cell Reports* **43**, DOI: [10.1016/j.celrep.2024.113957](https://doi.org/10.1016/j.celrep.2024.113957) (2024). Doi: 10.1016/j.celrep.2024.113957.
17. Dong, C., Madar, A. D. & Sheffield, M. E. Distinct place cell dynamics in ca1 and ca3 encode experience in new environments. *Nat. Commun.* **12**, DOI: [10.1038/s41467-021-23260-3](https://doi.org/10.1038/s41467-021-23260-3) (2021).
18. Bittner, K. C., Milstein, A. D., Grienberger, C., Romani, S. & Magee, J. C. Behavioral time scale synaptic plasticity underlies ca1 place fields. *Tech. Rep.* (2017).
19. Bittner, K. C. *et al.* Conjunctive input processing drives feature selectivity in hippocampal ca1 neurons. *Nat. Neurosci.* **18**, 1133–1142, DOI: [10.1038/nn.4062](https://doi.org/10.1038/nn.4062) (2015).
20. Balind, S. R. *et al.* Diverse synaptic and dendritic mechanisms of complex spike burst generation in hippocampal ca3 pyramidal cells. *Nat. Commun.* **10**, 1859, DOI: [10.1038/s41467-019-09767-w](https://doi.org/10.1038/s41467-019-09767-w) (2019).
21. Milstein, A. D. *et al.* Bidirectional synaptic plasticity rapidly modifies hippocampal representations. *eLife* **10**, e73046, DOI: [10.7554/eLife.73046](https://doi.org/10.7554/eLife.73046) (2021).
22. Grienberger, C. & Magee, J. C. Entorhinal cortex directs learning-related changes in ca1 representations. *Nature* **611**, 554–562, DOI: [10.1038/s41586-022-05378-6](https://doi.org/10.1038/s41586-022-05378-6) (2022).
23. Galloni, A. R. *et al.* Neuromorphic one-shot learning utilizing a phase-transition material. *Proc. Natl. Acad. Sci.* **121**, e2318362121, DOI: [10.1073/pnas.2318362121](https://doi.org/10.1073/pnas.2318362121) (2024). <https://www.pnas.org/doi/pdf/10.1073/pnas.2318362121>.
24. Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679, DOI: [10.1126/science.8036517](https://doi.org/10.1126/science.8036517) (1994). Doi: 10.1126/science.8036517.
25. Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G. & Zugaro, M. B. Selective suppression of hippocampal ripples impairs spatial memory. *Nat. Neurosci.* **12**, 1222–1223, DOI: [10.1038/nn.2384](https://doi.org/10.1038/nn.2384) (2009).
26. McNaughton, B. L. Cortical hierarchies, sleep, and the extraction of knowledge from memory, DOI: [10.1016/j.artint.2009.11.013](https://doi.org/10.1016/j.artint.2009.11.013) (2010).
27. Goto, A. & Hayashi, Y. Offline neuronal activity and synaptic plasticity during sleep and memory consolidation. *Neurosci. Res.* **189**, 29–36 (2023).
28. Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S. & Redish, A. D. Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705, DOI: <https://doi.org/10.1016/j.neuron.2010.01.034> (2010).
29. Frank, L. M., Stanley, G. B. & Brown, E. N. Hippocampal plasticity across multiple days of exposure to novel environments. *The J. Neurosci.* **24**, 7681, DOI: [10.1523/JNEUROSCI.1958-04.2004](https://doi.org/10.1523/JNEUROSCI.1958-04.2004) (2004).
30. Bakermans, J. J. W., Warren, J., Whittington, J. C. R. & Behrens, T. E. J. Constructing future behavior in the hippocampal formation through composition and replay. *Nat. Neurosci.* **28**, 1061–1072, DOI: [10.1038/s41593-025-01908-3](https://doi.org/10.1038/s41593-025-01908-3) (2025).
31. Barreto, A. *et al.* Successor features for transfer in reinforcement learning. In Guyon, I. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 30 (Curran Associates, Inc., 2017).
32. Li, Y., Briguglio, J. J., Romani, S. & Magee, J. C. Mechanisms of memory-supporting neuronal dynamics in hippocampal area ca3. *Cell* **187**, 6804–6819.e21, DOI: [10.1016/j.cell.2024.09.041](https://doi.org/10.1016/j.cell.2024.09.041) (2024). Doi: 10.1016/j.cell.2024.09.041.
33. Ecker, A. *et al.* Hippocampal sharp wave-ripples and the associated sequence replay emerge from structured synaptic interactions in a network model of area ca3. *eLife* **11**, DOI: [10.7554/eLife.71850](https://doi.org/10.7554/eLife.71850) (2022).
34. Davidson, T. J., Kloosterman, F. & Wilson, M. A. Hippocampal replay of extended experience. *Neuron* **63**, 497–507, DOI: [10.1016/j.neuron.2009.07.027](https://doi.org/10.1016/j.neuron.2009.07.027) (2009). Doi: 10.1016/j.neuron.2009.07.027.
35. Foster, D. J. & Wilson, M. A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683 (2006).

- 36.** Cheng, H. & Brown, J. W. Replay as a basis for backpropagation through time in the brain. DOI: [10.1101/2023.02.23.529770](https://doi.org/10.1101/2023.02.23.529770).
- 37.** Ambrose, R. E., Pfeiffer, B. E. & Foster, D. J. Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron* **91**, 1124–1136, DOI: [10.1016/j.neuron.2016.07.047](https://doi.org/10.1016/j.neuron.2016.07.047) (2016). doi: 10.1016/j.neuron.2016.07.047.
- 38.** Leutgeb, S., Leutgeb, J., Treves, A., Moser, M.-B. & Moser, E. Leutgeb & moser, science, 2004, distinct ensemble codes in hippocampal areas ca3 and ca1. *Science* **305**, 1295–1298 (2004).
- 39.** Milstein, A. D., Tran, S., Ng, G. & Soltesz, I. Offline memory replay in recurrent neuronal networks emerges from constraints on online dynamics. *The J. Physiol.* **601**, 3241–3264, DOI: <https://doi.org/10.1111/JP283216> (2023).
- 40.** Johnson, A. & Redish, A. D. Neural ensembles in ca3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12176–12189, DOI: [10.1523/JNEUROSCI.3761-07.2007](https://doi.org/10.1523/JNEUROSCI.3761-07.2007) (2007).
- 41.** Pfeiffer, B. E. & Foster, D. J. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79, DOI: [10.1038/nature12112](https://doi.org/10.1038/nature12112) (2013).
- 42.** van de Ven, G. M., Siegelmann, H. T. & Tolias, A. S. Brain-inspired replay for continual learning with artificial neural networks. *Nat. Commun.* **11**, DOI: [10.1038/s41467-020-17866-2](https://doi.org/10.1038/s41467-020-17866-2) (2020).
- 43.** Saxena, R., Shobe, J. L. & Mcnaughton, B. L. Learning in deep neural networks and brains with similarity-weighted interleaved learning. DOI: [10.1073/pnas](https://doi.org/10.1073/pnas) (2025).
- 44.** Gillespie, A. K. *et al.* Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *Neuron* **109**, 3149–3163.e6, DOI: <https://doi.org/10.1016/j.neuron.2021.07.029> (2021).
- 45.** Cheng, H. & Brown, J. W. Replay as a basis for backpropagation through time in the brain, DOI: [10.1101/2023.02.23.529770](https://doi.org/10.1101/2023.02.23.529770) (2023).
- 46.** Sučević, J. & Schapiro, A. C. A neural network model of hippocampal contributions to category learning. *eLife* **12**, DOI: [10.7554/eLife.77185](https://doi.org/10.7554/eLife.77185) (2023).

[11pt]article

graphicx amsmath amssymb mathptmx mathrsfs array OMScmsymn [ruled]algorithm2e

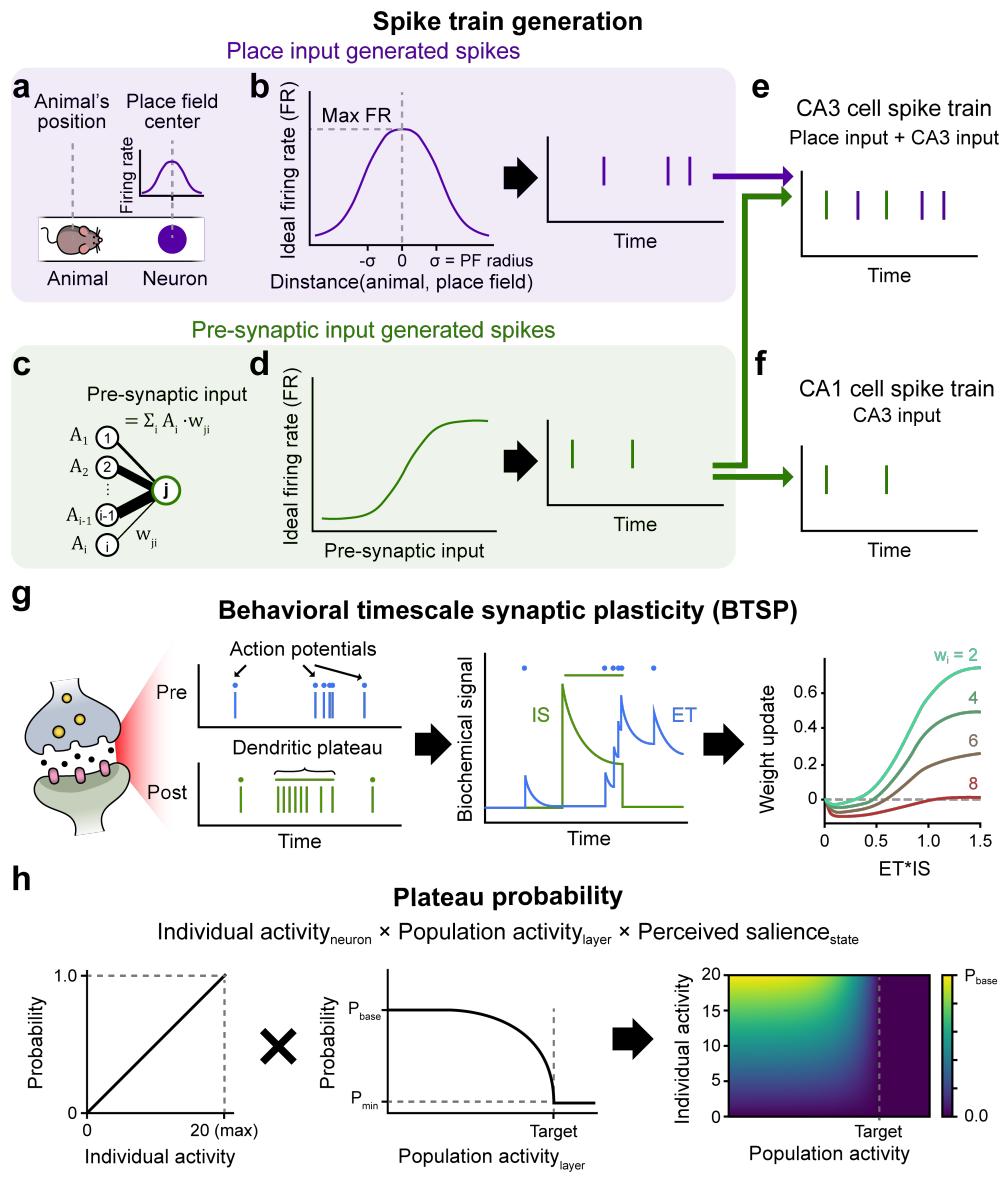
newfloat [name=Supplementary Figure]suppfigure

Supplementary materials

## Contents

<b>References</b>	<b>14</b>
<b>1 Details of synaptic learning rule during the online phase</b>	<b>18</b>
<b>2 Pseudo code of the online learning</b>	<b>19</b>
<b>3 CA1 place cell activity of individual seeds</b>	<b>20</b>
<b>4 Online learning using symmetric STDP</b>	<b>21</b>
<b>5 Equilibrium BTSP weight</b>	<b>22</b>
<b>6 Offline phase simulation details</b>	<b>23</b>
6.1 Single cell model . . . . .	23
6.2 Synapse model . . . . .	23
6.3 Connection weight optimization . . . . .	23
<b>7 Replay frequency modulation based on perceived salience</b>	<b>25</b>
<b>References</b>	<b>26</b>

## 1 Details of synaptic learning rule during the online phase



**Supplementary Figure 1.** This is a supplementary figure.

## 2 Pseudo code of the online learning

---

**Algorithm 1:** Training during online maze exploration

---

**Input :**  $\mathcal{S}_{\text{lap}}, I_s, P_s, \text{MI}$   
**Output :**  $W_{\text{CA3}}, W_{\text{CA1}}, W_{\text{pred}}$

**Initialize:**

```

# Initialize weights and signals.
 $W_{\text{CA3}} \sim \mathcal{N}(10^{-2}, 10^{-8}); W_{\text{CA1}} \sim \mathcal{N}(10^{-2}, 10^{-8}); W_{\text{pred}} \leftarrow 0$ 
 $A_{\text{CA3}} \leftarrow 0; ET_{\text{CA3}} \leftarrow 0; IS_{\text{CA3}} \leftarrow 0; IS_{\text{CA1}} \leftarrow 0$ 

for lap = 1 to Totallaps do
    for  $s \in \mathcal{S}_{\text{lap}}$  do
        for  $t = 1$  to  $T_s$  do
            # Generate new spike trains and calculate the activities at every
            #  $T_{\text{SpikeUpdate}}$ .
            if mod( $t, T_{\text{SpikeUpdate}}$ ) = 0 then
                SpikeTrainCA3  $\leftarrow$  GenerateSpikeTrain( $s, A_{\text{CA3}}, W_{\text{CA3}}$ )
                 $A_{\text{CA3}} \leftarrow \text{Activity}(\text{SpikeTrain}_{\text{CA3}})$ 

                SpikeTrainCA1  $\leftarrow$  GenerateSpikeTrain( $\cdot, A_{\text{CA3}}, W_{\text{CA1}}$ )
                 $A_{\text{CA1}} \leftarrow \text{Activity}(\text{SpikeTrain}_{\text{CA1}})$ 
            end

            # Compute novelty and train the feature prediction network.
             $(N(t), W_{\text{pred}}) \leftarrow \text{DeltaRule}(P_s, A_{\text{CA1}}, W_{\text{pred}})$ 
            PS( $t$ )  $\leftarrow$  PerceivedSalience( $I_s, N(t), \text{MI}$ )

            # Calculate the plateau probability of hippocampal neurons.
             $P_{\text{plateau}, \text{CA3}} \leftarrow \text{PlateauProbability}(A_{\text{CA3}}, \cdot)$ 
             $P_{\text{plateau}, \text{CA1}} \leftarrow \text{PlateauProbability}(A_{\text{CA1}}, \text{PS}(t))$ 

            # Update ETs and ISs.
             $ET_{\text{CA3}} += \text{UpdateET}(\text{SpikeTrain}_{\text{CA3}})$ 
             $IS_{\text{CA3}} += \text{UpdateIS}(P_{\text{plateau}, \text{CA3}})$ 
             $IS_{\text{CA1}} += \text{UpdateIS}(P_{\text{plateau}, \text{CA1}})$ 

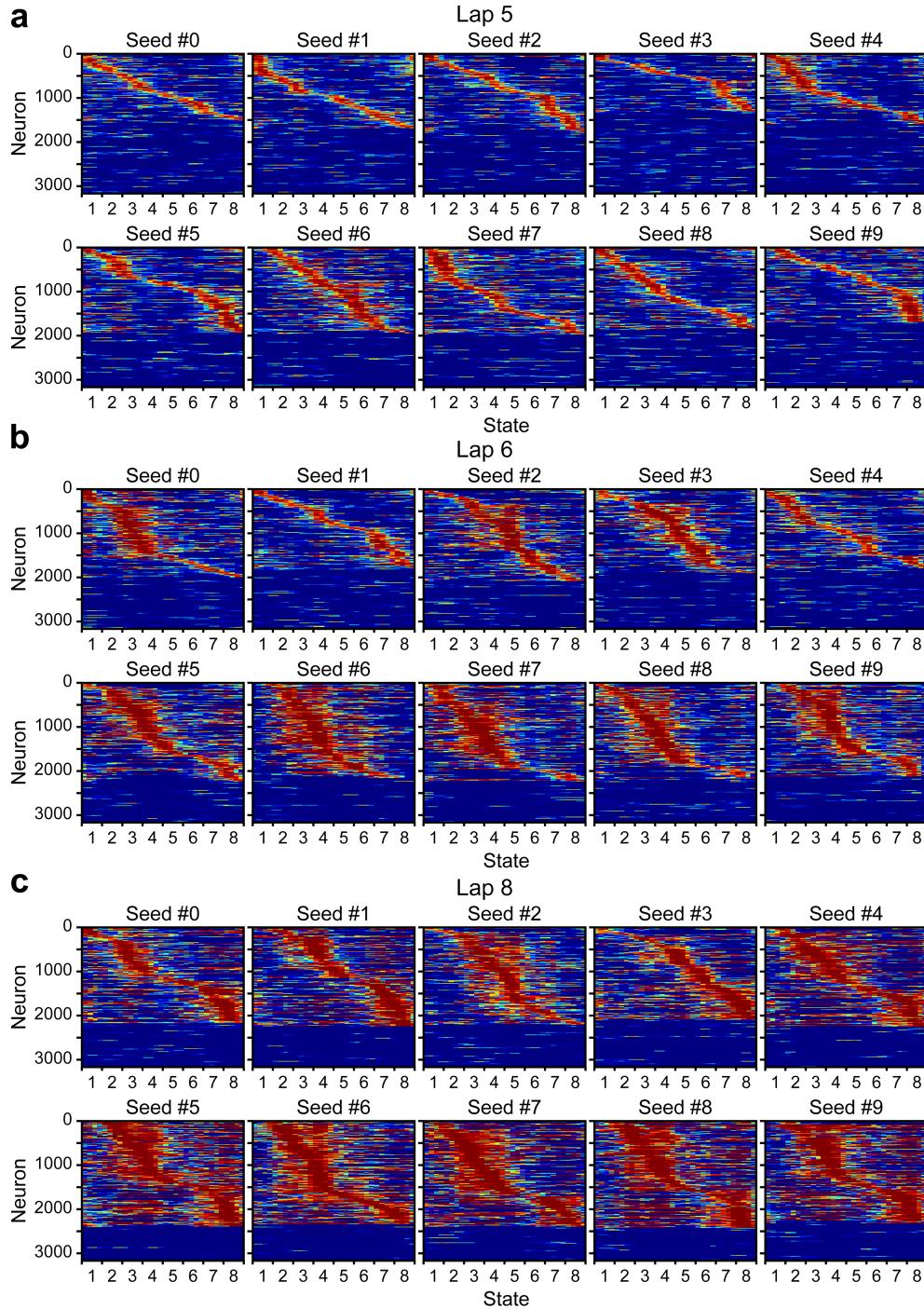
            # Train the hippocampal network.
             $W_{\text{CA3}} \leftarrow \text{BTSP}(ET_{\text{CA3}}, IS_{\text{CA3}}, W_{\text{CA3}})$ 
             $W_{\text{CA1}} \leftarrow \text{BTSP}(ET_{\text{CA3}}, IS_{\text{CA1}}, W_{\text{CA1}})$ 

            # Decay eligibility traces and instructive signals.
             $ET_{\text{CA3}} -= ET_{\text{CA3}}/\tau_{\text{CA3}}$ 
             $IS_{\text{CA3}} -= IS_{\text{CA3}}/\tau_{\text{CA3}}$ 
             $IS_{\text{CA1}} -= IS_{\text{CA1}}/\tau_{\text{CA1}}$ 
        end
    end

```

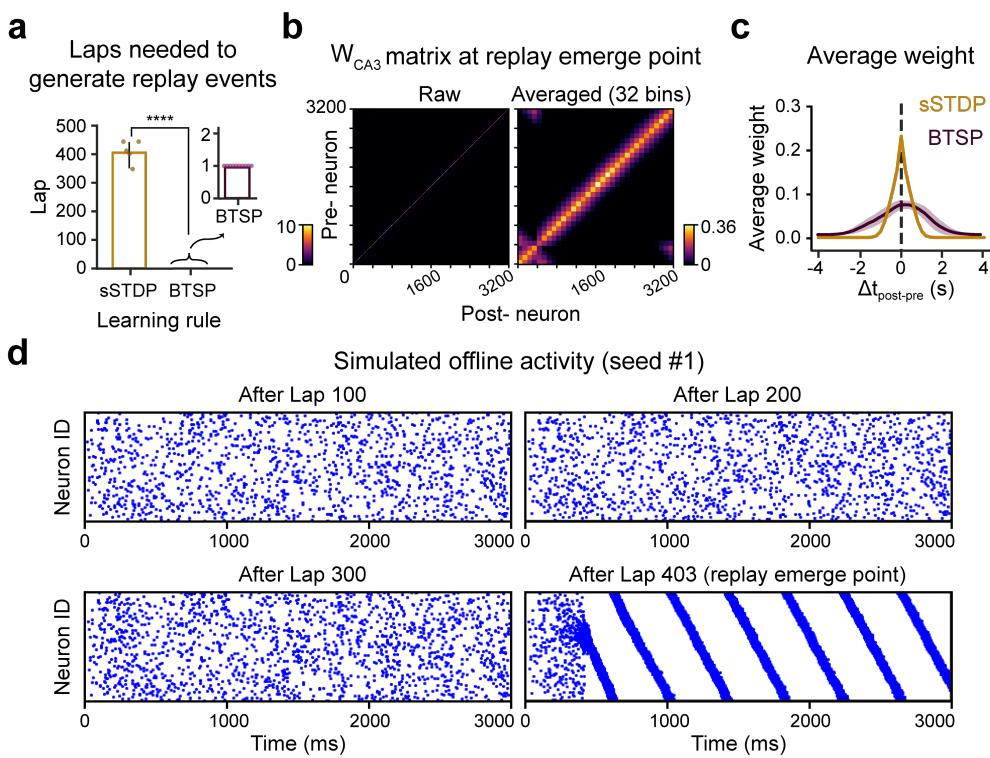
---

### 3 CA1 place cell activity of individual seeds



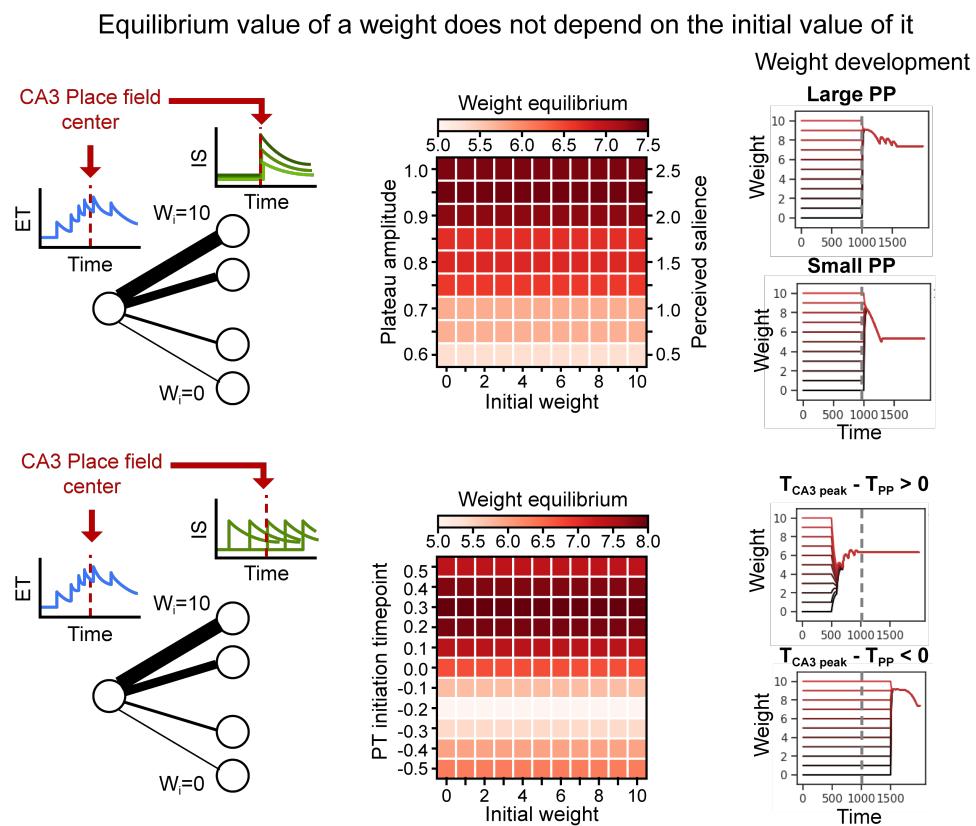
**Supplementary Figure 2.** This is a supplementary figure.

## 4 Online learning using symmetric STDP



**Supplementary Figure 3.** This is a supplementary figure.

## 5 Equilibrium BTSP weight



**Supplementary Figure 4.** This is a supplementary figure.

## 6 Offline phase simulation details

Offline-phase simulations were performed using the synaptic weight matrices  $W_{\text{CA}3}$  and  $W_{\text{CA}1}$  obtained from the online learning phase. Starting from these learned weights, we adopted the offline simulation framework of Ecker et al.<sup>33</sup>, using their publicly available code as the basis for our implementation.

### 6.1 Single cell model

We used a neuron model with the adaptive exponential integrate-and-fire (AdExIF) equations<sup>2,2</sup>, inspired by the previous work<sup>33</sup>. Here, the membrane potential  $V(t)$  and adaptation current  $w(t)$  interact, following the equations:

$$C_m \frac{dV(t)}{dt} = -g_L(V(t) - V_{\text{rest}}) + g_L \Delta_T \exp\left(\frac{V(t) - \vartheta}{\Delta_T}\right) - w(t) + I_{\text{syn}}(t), \quad (23)$$

$$\tau_w \frac{dw(t)}{dt} = a(V(t) - V_{\text{rest}}) - w(t). \quad (24)$$

When  $V(t)$  crosses the threshold  $\vartheta$ , it is reset to  $V_{\text{reset}}$  and held for a refractory period  $t_{\text{ref}}$ . The model incorporates adaptation process, which is important in replay generation<sup>33</sup>. The adaptation current  $w(t)$  is incremented by  $b$  at each spike, where  $a$  and  $b$  control subthreshold and spike-triggered adaptation, respectively.

The actual parameter values used for the simulations are summarized in Table 1, largely adapted from the previous work<sup>33</sup>. Physical units are as follows:  $C_m$  (pF),  $g_L$  and  $a$  (nS),  $V_{\text{rest}}$ ,  $\Delta_T$ ,  $\vartheta$ , and  $V_{\text{reset}}$  (mV),  $t_{\text{ref}}$  and  $\tau_w$  (ms), and  $b$  (pA).

**Table 1.** Neuron model parameters.

Parameter	Symbol	Value
Membrane capacitance	$C_m$	180.13
Leak conductance	$g_L$	5
Leak reversal (resting) potential	$V_{\text{rest}}$	-75.19
Slope of exponential term	$\Delta_T$	4.23
Firing threshold	$\vartheta$	-24.42
Adaptation time constant	$\tau_w$	84.93
Subthreshold adaptation strength	$a$	-0.01
Spike-triggered adaptation increment	$b$	500
Reset potential	$V_{\text{reset}}$	-29.74
Absolute refractory period	$t_{\text{ref}}$	5.96

### 6.2 Synapse model

Synapses were modeled as conductances with biexponential kinetics:

$$g(t) = \hat{g}A \left[ \exp\left(-\frac{t}{\tau_d}\right) - \exp\left(-\frac{t}{\tau_r}\right) \right], \quad (25)$$

where  $\hat{g}$  is the peak conductance, and  $\tau_r$  and  $\tau_d$  are the rise and decay time constants, respectively. The normalization constant  $A = \exp(-t_p/\tau_d) - \exp(-t_p/\tau_r)$  ensures that the conductance peaks at time  $t_p = \tau_r \tau_d / (\tau_d - \tau_r) \ln(\tau_d/\tau_r)$ . Synaptic currents were computed as

$$I_{\text{syn}}(t) = g_{\text{AMPA}}(t)(V(t) - E_{\text{exc}}) + g_{\text{GABA}}(t)(V(t) - E_{\text{inh}}), \quad (26)$$

where  $E_{\text{exc}} = 0$  mV and  $E_{\text{inh}} = -70$  mV are the reversal potentials of excitatory and inhibitory synapses, respectively.

The parameters used for the synapses are summarized in Table 2. Ext stands for excitatory cells, Inh stands for inhibitory cells, and EC stands for entorhinal cortex populations. Physical dimensions are as follows:  $\hat{g}$  (nS),  $\tau_r$ ,  $\tau_d$ , and  $t_d$  (ms), and connection probability  $p_{\text{conn}}$  is dimensionless.

### 6.3 Connection weight optimization

Synaptic weights of the network were optimized using the publicly available code from the previous study<sup>33</sup> after being modified to incorporate  $W_{\text{CA}1}$  network. We sought to determine an optimal set of single-valued parameters determining the strength of excitatory and inhibitory connections within and between hippocampal subfields. Specifically, the parameters optimized are listed below:

**Table 2.** Synaptic parameters (adapted from<sup>2,3,4,5,6</sup>).

Connection	$\hat{g}$ (nS)	$\tau_r$ (ms)	$\tau_d$ (ms)	$t_d$ (ms)	$p_{conn}$
Ext → Ext	0.1-6.3	0-15	1.3	9.5	0.1
Ext → Inh	0.85	1.0	4.1	0.9	0.1
Inh → Ext	0.65	0.3	3.3	1.1	0.25
Inh → Inh	5.0	0.25	1.2	0.6	0.25
EC → Ext	19.15	21.5	5.4	—	—

- Entorhinal cortex (EC) input rate
- EC → CA3 excitatory connection weight
- CA3 excitatory → CA3 inhibitory connection weight
- CA3 inhibitory → CA3 excitatory connection weight
- CA3 inhibitory → CA3 inhibitory connection weight
- Weight scaling factor for  $W_{CA3}$
- CA1 excitatory → CA1 inhibitory connection weight
- CA1 inhibitory → CA1 excitatory connection weight
- CA1 inhibitory → CA1 inhibitory connection weight

Optimization followed the evolutionary multi-objective framework<sup>33</sup>, with customized upper and lower bounds for each parameter listed in Table 3. The final optimized weights used in the three task environments are summarized in Table 4.

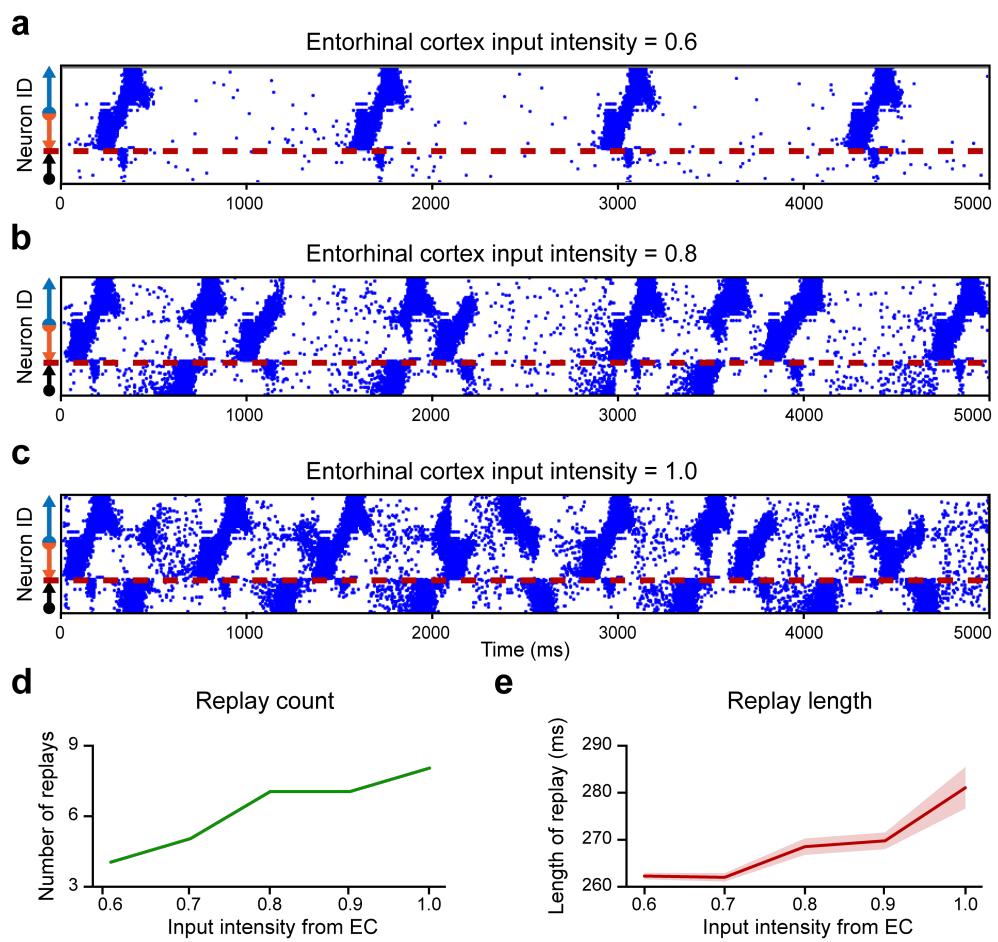
**Table 3.** Upper bound and lower bound values used for the optimization process.

Parameter	Lower bound	Upper bound
EC input rate	5.0	15.0
EC → CA3 Ext	0.5	40.0
CA3 Exc → Inh	0.2	5.0
CA3 Inh → Exc	0.5	20.0
CA3 Inh → Inh	1.0	15.0
CA3 weight scale ( $W_{CA3}$ )	1.0	2.5
CA1 Exc → Inh	0.2	5.0
CA1 Inh → Exc	1.0	10.0
CA1 Inh → Inh	1.0	15.0

**Table 4.** Optimized connection weights. Each value represents the final parameter obtained after evolutionary optimization.

Parameter	Linear treadmill	T-maze	Closed linear track
EC input rate	10.084	7.322	13.038
EC → CA3 Ext	23.957	34.481	15.843
CA3 Exc → Inh	0.349	1.765	1.715
CA3 Inh → Exc	1.941	1.381	1.646
CA3 Inh → Inh	2.971	11.110	14.732
CA3 weight scale ( $W_{CA3}$ )	2.210	1.886	2.306
CA1 Exc → Inh	1.000	1.765	0.337
CA1 Inh → Exc	1.198	1.877	1.052
CA1 Inh → Inh	14.725	13.819	12.671

## 7 Replay frequency modulation based on perceived salience



**Supplementary Figure 5.** This is a supplementary figure.

## References

1. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208, DOI: [10.1037/h0061626](https://doi.org/10.1037/h0061626) (1948).
2. Behrens, T. E. J. *et al.* What is a cognitive map? organizing knowledge for flexible behavior. *Neuron* **100**, 490–509, DOI: <https://doi.org/10.1016/j.neuron.2018.10.002> (2018).
3. Rosenberg, M., Zhang, T., Perona, P. & Meister, M. Mice in a labyrinth show rapid learning, sudden insight, and efficient exploration. *eLife* **10**, e66175, DOI: [10.7554/eLife.66175](https://doi.org/10.7554/eLife.66175) (2021).
4. Dayan, P. Improving generalization for temporal difference learning: The successor representation. *Tech. Rep.* (1993).
5. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653, DOI: [10.1038/nn.4650](https://doi.org/10.1038/nn.4650) (2017).
6. Mehta, M. R., Quirk, M. C. & Wilson, M. A. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* **25**, 707–715, DOI: [10.1016/S0896-6273\(00\)81072-7](https://doi.org/10.1016/S0896-6273(00)81072-7) (2000).
7. Gustafson, N. J. & Daw, N. D. Grid cells, place cells, and geodesic generalization for spatial reinforcement learning. *PLOS Comput. Biol.* **7**, e1002235, DOI: [10.1371/journal.pcbi.1002235](https://doi.org/10.1371/journal.pcbi.1002235) (2011).
8. Gershman, S. J. *et al.* Explaining dopamine through prediction errors and beyond. *Nat. Neurosci.* **27**, 1645–1655, DOI: [10.1038/s41593-024-01705-4](https://doi.org/10.1038/s41593-024-01705-4) (2024).
9. Wu, C.-T., Haggerty, D., Kemere, C. & Ji, D. Hippocampal awake replay in fear memory retrieval. *Nat. Neurosci.* **20**, 571–580, DOI: [10.1038/nn.4507](https://doi.org/10.1038/nn.4507) (2017).
10. Carey, A. A., Tanaka, Y. & van der Meer, M. A. A. Reward revaluation biases hippocampal replay content away from the preferred outcome. *Nat. Neurosci.* **22**, 1450–1459, DOI: [10.1038/s41593-019-0464-6](https://doi.org/10.1038/s41593-019-0464-6) (2019).
11. Krishnan, S., Heer, C., Cherian, C. & Sheffield, M. E. Reward expectation extinction restructures and degrades ca1 spatial maps through loss of a dopaminergic reward proximity signal. *Nat. Commun.* **13**, DOI: [10.1038/s41467-022-34465-5](https://doi.org/10.1038/s41467-022-34465-5) (2022).
12. Kutlu, M. G. *et al.* Dopamine release in the nucleus accumbens core signals perceived saliency. *Curr. Biol.* **31**, 4748–4761.e8, DOI: [10.1016/j.cub.2021.08.052](https://doi.org/10.1016/j.cub.2021.08.052) (2021). Doi: 10.1016/j.cub.2021.08.052.
13. 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. *The J. Neurosci.* **21**, 1635, DOI: [10.1523/JNEUROSCI.21-05-01635.2001](https://doi.org/10.1523/JNEUROSCI.21-05-01635.2001) (2001).
14. Gauthier, J. L. & Tank, D. W. A dedicated population for reward coding in the hippocampus. *Neuron* **99**, 179–193.e7, DOI: <https://doi.org/10.1016/j.neuron.2018.06.008> (2018).
15. Mamad, O. *et al.* Place field assembly distribution encodes preferred locations. *PLOS Biol.* **15**, e2002365– (2017).
16. Nikbakht, N. *et al.* Efficient encoding of aversive location by ca3 long-range projections. *Cell Reports* **43**, DOI: [10.1016/j.celrep.2024.113957](https://doi.org/10.1016/j.celrep.2024.113957) (2024). Doi: 10.1016/j.celrep.2024.113957.
17. Dong, C., Madar, A. D. & Sheffield, M. E. Distinct place cell dynamics in ca1 and ca3 encode experience in new environments. *Nat. Commun.* **12**, DOI: [10.1038/s41467-021-23260-3](https://doi.org/10.1038/s41467-021-23260-3) (2021).
18. Bittner, K. C., Milstein, A. D., Grienberger, C., Romani, S. & Magee, J. C. Behavioral time scale synaptic plasticity underlies ca1 place fields. *Tech. Rep.* (2017).
19. Bittner, K. C. *et al.* Conjunctive input processing drives feature selectivity in hippocampal ca1 neurons. *Nat. Neurosci.* **18**, 1133–1142, DOI: [10.1038/nn.4062](https://doi.org/10.1038/nn.4062) (2015).
20. Balind, S. R. *et al.* Diverse synaptic and dendritic mechanisms of complex spike burst generation in hippocampal ca3 pyramidal cells. *Nat. Commun.* **10**, 1859, DOI: [10.1038/s41467-019-09767-w](https://doi.org/10.1038/s41467-019-09767-w) (2019).
21. Milstein, A. D. *et al.* Bidirectional synaptic plasticity rapidly modifies hippocampal representations. *eLife* **10**, e73046, DOI: [10.7554/eLife.73046](https://doi.org/10.7554/eLife.73046) (2021).
22. Grienberger, C. & Magee, J. C. Entorhinal cortex directs learning-related changes in ca1 representations. *Nature* **611**, 554–562, DOI: [10.1038/s41586-022-05378-6](https://doi.org/10.1038/s41586-022-05378-6) (2022).
23. Galloni, A. R. *et al.* Neuromorphic one-shot learning utilizing a phase-transition material. *Proc. Natl. Acad. Sci.* **121**, e2318362121, DOI: [10.1073/pnas.2318362121](https://doi.org/10.1073/pnas.2318362121) (2024). <https://www.pnas.org/doi/pdf/10.1073/pnas.2318362121>.

24. Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679, DOI: [10.1126/science.8036517](https://doi.org/10.1126/science.8036517) (1994). Doi: 10.1126/science.8036517.
25. Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G. & Zugaro, M. B. Selective suppression of hippocampal ripples impairs spatial memory. *Nat. Neurosci.* **12**, 1222–1223, DOI: [10.1038/nn.2384](https://doi.org/10.1038/nn.2384) (2009).
26. McNaughton, B. L. Cortical hierarchies, sleep, and the extraction of knowledge from memory, DOI: [10.1016/j.artint.2009.11.013](https://doi.org/10.1016/j.artint.2009.11.013) (2010).
27. Goto, A. & Hayashi, Y. Offline neuronal activity and synaptic plasticity during sleep and memory consolidation. *Neurosci. Res.* **189**, 29–36 (2023).
28. Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S. & Redish, A. D. Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705, DOI: <https://doi.org/10.1016/j.neuron.2010.01.034> (2010).
29. Frank, L. M., Stanley, G. B. & Brown, E. N. Hippocampal plasticity across multiple days of exposure to novel environments. *The J. Neurosci.* **24**, 7681, DOI: [10.1523/JNEUROSCI.1958-04.2004](https://doi.org/10.1523/JNEUROSCI.1958-04.2004) (2004).
30. Bakermans, J. J. W., Warren, J., Whittington, J. C. R. & Behrens, T. E. J. Constructing future behavior in the hippocampal formation through composition and replay. *Nat. Neurosci.* **28**, 1061–1072, DOI: [10.1038/s41593-025-01908-3](https://doi.org/10.1038/s41593-025-01908-3) (2025).
31. Barreto, A. *et al.* Successor features for transfer in reinforcement learning. In Guyon, I. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 30 (Curran Associates, Inc., 2017).
32. Li, Y., Briguglio, J. J., Romani, S. & Magee, J. C. Mechanisms of memory-supporting neuronal dynamics in hippocampal area ca3. *Cell* **187**, 6804–6819.e21, DOI: [10.1016/j.cell.2024.09.041](https://doi.org/10.1016/j.cell.2024.09.041) (2024). Doi: 10.1016/j.cell.2024.09.041.
33. Ecker, A. *et al.* Hippocampal sharp wave-ripples and the associated sequence replay emerge from structured synaptic interactions in a network model of area ca3. *eLife* **11**, DOI: [10.7554/eLife.71850](https://doi.org/10.7554/eLife.71850) (2022).
34. Davidson, T. J., Kloosterman, F. & Wilson, M. A. Hippocampal replay of extended experience. *Neuron* **63**, 497–507, DOI: [10.1016/j.neuron.2009.07.027](https://doi.org/10.1016/j.neuron.2009.07.027) (2009). Doi: 10.1016/j.neuron.2009.07.027.
35. Foster, D. J. & Wilson, M. A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683 (2006).
36. Cheng, H. & Brown, J. W. Replay as a basis for backpropagation through time in the brain. DOI: [10.1101/2023.02.23.529770](https://doi.org/10.1101/2023.02.23.529770).
37. Ambrose, R. E., Pfeiffer, B. E. & Foster, D. J. Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron* **91**, 1124–1136, DOI: [10.1016/j.neuron.2016.07.047](https://doi.org/10.1016/j.neuron.2016.07.047) (2016). Doi: 10.1016/j.neuron.2016.07.047.
38. Leutgeb, S., Leutgeb, J., Treves, A., Moser, M.-B. & Moser, E. Leutgeb & moser, science, 2004, distinct ensemble codes in hippocampal areas ca3 and ca1. *Science* **305**, 1295–1298 (2004).
39. Milstein, A. D., Tran, S., Ng, G. & Soltesz, I. Offline memory replay in recurrent neuronal networks emerges from constraints on online dynamics. *The J. Physiol.* **601**, 3241–3264, DOI: <https://doi.org/10.1113/JP283216> (2023).
40. Johnson, A. & Redish, A. D. Neural ensembles in ca3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.* **27**, 12176–12189, DOI: [10.1523/JNEUROSCI.3761-07.2007](https://doi.org/10.1523/JNEUROSCI.3761-07.2007) (2007).
41. Pfeiffer, B. E. & Foster, D. J. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79, DOI: [10.1038/nature12112](https://doi.org/10.1038/nature12112) (2013).
42. van de Ven, G. M., Siegelmann, H. T. & Tolias, A. S. Brain-inspired replay for continual learning with artificial neural networks. *Nat. Commun.* **11**, DOI: [10.1038/s41467-020-17866-2](https://doi.org/10.1038/s41467-020-17866-2) (2020).
43. Saxena, R., Shobe, J. L. & McNaughton, B. L. Learning in deep neural networks and brains with similarity-weighted interleaved learning. DOI: [10.1073/pnas](https://doi.org/10.1073/pnas) (2025).
44. Gillespie, A. K. *et al.* Hippocampal replay reflects specific past experiences rather than a plan for subsequent choice. *Neuron* **109**, 3149–3163.e6, DOI: <https://doi.org/10.1016/j.neuron.2021.07.029> (2021).
45. Cheng, H. & Brown, J. W. Replay as a basis for backpropagation through time in the brain, DOI: [10.1101/2023.02.23.529770](https://doi.org/10.1101/2023.02.23.529770) (2023).
46. Sučević, J. & Schapiro, A. C. A neural network model of hippocampal contributions to category learning. *eLife* **12**, DOI: [10.7554/eLife.77185](https://doi.org/10.7554/eLife.77185) (2023).