The hippocampal formation as a hierarchical generative model supporting generative replay and continual learning

Ivilin Stoianov[1], Domenico Maisto[2], Giovanni Pezzulo[1]

1. Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy

2. Institute for High Performance Computing and Networking, National Research Council, Naples, Italy

Corresponding author:

Giovanni Pezzulo

Institute of Cognitive Sciences and Technologies, National Research Council

Via S. Martino della Battaglia, 44

00185 Rome, Italy

Email: giovanni.pezzulo@istc.cnr.it

Phone: +39 06 4459 5206

Abstract

We advance a novel computational theory of the hippocampal formation as a hierarchical generative model that organizes sequential experiences, such as rodent trajectories during spatial navigation, into coherent spatiotemporal contexts. We propose that the hippocampal generative model is endowed with inductive biases to pattern-separate individual items of experience (first hierarchical layer), organize them into sequences (second layer) and cluster them into maps (third layer). This theory entails a novel characterization of hippocampal reactivations as *generative replay*: the offline resampling of fictive sequences from the generative model, which supports the continual learning of multiple sequential experiences. We show that the model learns and efficiently retains multiple spatial navigation trajectories, by organizing them into spatial maps. Furthermore, the hierarchical model reproduces flexible and prospective aspects of hippocampal dynamics that are challenging to explain within existing frameworks. This theory reconciles multiple roles of the hippocampal formation in map-based navigation, episodic memory and imagination.

Keywords: hippocampus; generative model; generative replay; cognitive map; sequence generation, continual learning

Introduction

During our lives, we continuously learn new skills and accumulate novel memories. A fundamental problem in both neuroscience and machine learning is preventing this novel acquired knowledge from interfering with past memories and causing *catastrophic forgetting* [1]. The hippocampus has long been considered important in solving this biological problem. According to the influential *Complementary Learning Systems* theory, it works as an episodic memory buffer that stores novel experiences rapidly and successively reactivates them offline (e.g., during sleep), in order to train a separate, semantic memory system located in the cortex, which gradually integrates and consolidates memories, preventing catastrophic forgetting [2,3].

The idea that the hippocampus stores and reactivates episodic memories offline has received empirical support from rodent studies during spatial navigation. In phases of wakeful rest during spatial navigation as well as subsequent sleep, the rodent hippocampus spontaneously generates (time-compressed) sequences of neural activations resembling those observed during behaviour, such as sequences of place cells corresponding to spatial trajectories experienced during navigation [4,5], sometimes in reverse order [6]. These *internally generated hippocampal sequences* have biological significance as their disruption has been shown to impair memory-based navigation [7]. These sequences have been called "replays" to highlight the fact that they might be interpreted as reactivations of previous experiences stored in an episodic memory buffer located in the hippocampus.

The idea of using an episodic memory buffer to store and replay experiences during learning is widespread in machine learning, as well. This “experience replay” method is especially effective for *continual learning*: when an artificial system has to learn multiple tasks sequentially. In most machine learning methods (e.g. those using neural networks), learning a novel task (task C) usually implies forgetting previously learned tasks (tasks A and B), because the memories of C interfere with those of A and B. Extending the neural net with experience replay mitigates this problem. Essentially, when the extended neural net learns tasks A and B, it stores a history of past episodes in a memory buffer. Then, while learning a novel task C, past episodes experienced during tasks A and B are intermittently sampled from the replay buffer and replayed thereby mitigating the risk of catastrophic forgetting. The functioning of this extended model has been often associated to the functioning of hippocampal “replay” during sleep [8,9].

While this view of “replay” is common in both neuroscience and machine learning, several studies challenge it. In rodents, internally generated hippocampal sequences during sleep or wakeful rest can depict paths to future goal locations rather than only past trajectories [10]; random trajectories, resembling Brownian diffusion [11]; trajectories that have never been directly explored but reflect prior spatial knowledge [12,13], such as shortcuts [14] and future event locations [15]; and *preplay* sequences of cells before a navigation episode, which are subsequently expressed during navigation [16]. In humans, internally generated sequences can re-order (non-spatial) item representations to reflect learned rules, rather than replaying the order in which they were actually experienced [17].

In sum, internally generated hippocampal sequences during sleep or wakeful rest extend well beyond previous experiences and demonstrate prospective aspects that are difficult to reconcile with simple replay from a memory buffer [5,18–20]. However, a robust theoretical framework unifying these findings has yet to emerge. Here we advance a novel theoretical proposal to explain flexible and prospective aspects of internally generated hippocampal sequences, within a statistical learning framework. We propose that the hippocampal formation functions as a *hierarchical generative model* that organizes sequential experiences

(e.g., spatial navigation trajectories) into a set of mutually exclusive spatiotemporal contexts (e.g., spatial maps). Furthermore, we propose that internally generated hippocampal sequences stem from *generative replay*, or the resampling of fictive experiences from the generative model, rather than from a verbatim replay of memories from a buffer.

This new theory is based on three main assumptions. First, that the hippocampal formation encodes experiences by *learning a hierarchical generative model* of observed data. The proposed hierarchical model has three layers, see Figure 1A. The first hierarchical layer forms latent codes for individual *items* of experience (e.g., the animal's spatial location) inferred from sensory observations. These item codes are pattern-separated and hence deprived of context. However, the second and third layers organize items into coherent spatiotemporal contexts. Specifically, the second layer forms latent codes for *sequences* within which the lower-level items are embedded (e.g., the animal's spatial trajectory), supporting their sequential processing – hence establishing the hippocampus a "sequence generator" [21]. The third layer forms latent codes for structured *maps* that cluster experiences into mutually exclusive spatial or relational contexts (e.g., the maze the animal is currently in versus alternative mazes). The hierarchical structure is key for the functioning of the hippocampal model, as it provides guiding principles (or "inductive biases") to organize individual *items* of experience in meaningful ways: namely, as *sequences* (temporally ordered items) and then *maps* (spatially clustered sequences).

The second assumption is that the hierarchical generative model serves a fundamental function: binding experiences into coherent spatiotemporal contexts [22,23]. This notion stems directly from the fact that in the hierarchical model, sets of item and sequence codes at the lower level (which correspond to individual experiences) are organized into separate maps (which correspond to mutually exclusive relational contexts). Hence, experiences that share surface similarities, such as running in two similar corridors, can engage different map codes in the generative model - and sometimes, different sets of neurons. The possibility to bind sets of individual experiences to separate relational contexts is a key feature of episodic and autobiographical memory [24]. Furthermore, a large body of literature illustrates that (statistical) context learning improves robustness and generalization [25–30], hence suggesting that the capability to group experiences into separate contexts might help the hippocampus alleviate catastrophic forgetting.

The third assumption is that internally generated hippocampal sequences are manifestations of the inferential dynamics of the hierarchical generative model rather than "replays" from a memory buffer. In other words, internally generated hippocampal sequences are the re-enactment of an experienced event (and its context) or the imagination of a novel, never experienced event having a similar statistical structure as the events learned by the generative model. Furthermore, and importantly, we assume that a key function of internally generated hippocampal sequences is learning. Specifically, we posit that the hippocampal generative model supports *generative replay*: a method for continual learning that uses fictive experiences resampled from a generative model to self-train the model that generates them (as well as to train separate models or controllers). Previous research has established that generative replay is appealing from a computational perspective, since it can prevent catastrophic forgetting and (unlike experience replay) does not require unbounded memory resources to store previous experiences [31–33]. Here, instead, we use the concept of generative replay to explain flexible and prospective aspects of internally generated hippocampal sequences, which are difficult to reconcile with previous theories.

In this article, we present a series of simulations that illustrate the computational efficacy ("Continual learning of multiple spatial maps with generative replay") and biological relevance ("Neurobiological underpinnings of generative replay") of the novel hypothesis.

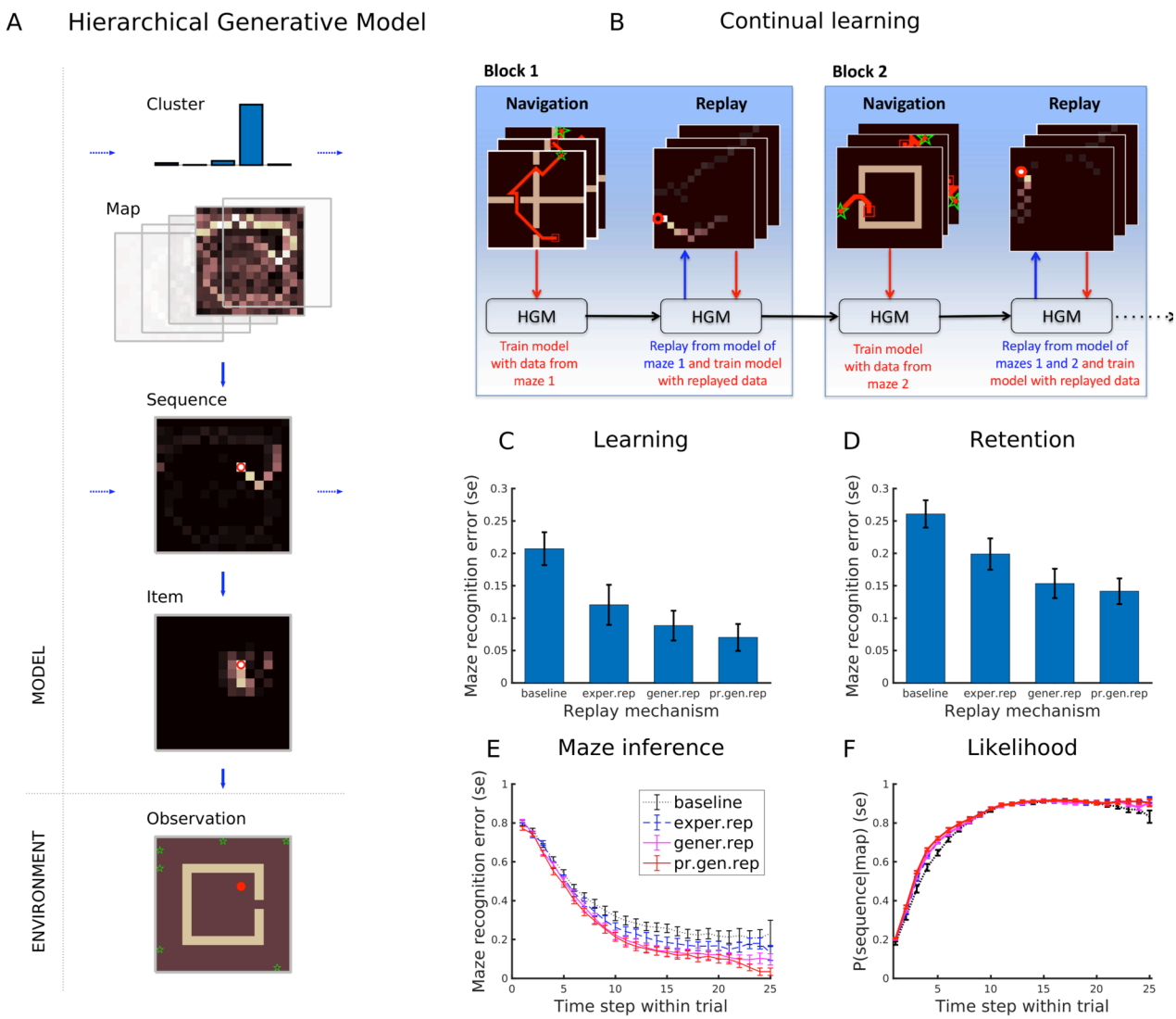


Figure 1. Hippocampal generative model. A: Components of the hierarchical generative model, from top to bottom: clusters / maps, sequences, items and observations. The red circle denotes the agent's current position. Colour codes correspond to probabilities, from high (white) to low (black). See the main text for explanation. The current implementation focuses on spatial navigation, where items, sequences, and maps may correspond to place cells, place cell sequences and 2D spatial maps, respectively. However, the same model could also apply to other domains, where the structure of maps could be different and even include factorizations (e.g., separate maps for objects or locations). B: Structure of the continual learning experiment. We compare four learning agents having the same hierarchical generative model but different replay mechanisms: no replay, experience replay, generative replay and generative prioritized replay. The agents' task consists in learning five mazes, each presented in a different block. For each block, the learning agents receive 20 trajectory data from one single maze, and use them to train their hierarchical generative model. Between blocks, the learning agents can replay experiences in different ways; see the main text. C-F: experimental results (means and standard errors of n=16 replicas of each agent). C: Performance of the four agents during learning (i.e., inference of the correct maze), measured after training on 5 trajectories of each block. D: Same as C, but measured after learning, on 20 unseen trials per maze. (E,F): Dynamics of inference. E: Same as D, but as a function of time step within trials (for 25 time steps). F: Accuracy of the reconstruction of sequences, as indexed by the likelihood of the sequence given the correct map.

Results

Continual learning of multiple spatial maps with generative replay

To assess the computational efficacy of generative replay, we designed a continual learning task consisting of learning multiple spatial mazes from (simulated) navigation trajectories. We compare four learning agents using the same hierarchical generative model (Figure 1A) but different replay methods: "baseline" (i.e., no replay), "experience replay", "generative replay", and a novel variant of generative replay: "prioritized generative replay".

In this continual learning task, each agent must learn five 15x15 mazes, using a single generative model, see Figure 1B. To render the learning experiment challenging, we selected five mazes that have large portions of spatial structure in common.

The agents learn the five mazes in different blocks (mazes were selected randomly without repetitions for each agent and block). During a block, each of the four learning agents receives the same batch of 20 spatial trajectories from the current maze, and uses these data to update their own hierarchical generative model. The trajectory data are generated by a separate model-based Bayesian controller that, for each trial and maze, moves between a randomly selected origin and goal location, by following a near-optimal trajectory [34].

Below we introduce the agents' generative model (which is the same for all agents) and their training procedures (which is different for each agent).

Hierarchical generative model

All agents are endowed with the same hierarchical generative model, shown in Figure 1A (see the Methods section for details). The generative model is composed of three layers of hidden states, encoding *items*, *sequences* and *maps*, respectively, as well as a bottom layer encoding the present *observation*. We describe the model from the bottom-up, beginning with observations.

At each moment, the agent receives an *observation* about its current spatial location in the maze (or an item in a spatial location). Note that the agent never observes full trajectories, but instead receives one data point at a time.

The first hidden layer of the hierarchical model encodes *item* codes, putatively corresponding to place cells. Note that while (for illustrative purposes) *observations* and *items* are plotted similarly in single grid coordinates in Figure 1A, these are distinct elements: the former are model inputs, whereas the latter are hidden (or latent) states of the generative model. Item codes are inferred probabilistically, using a scheme analogous to predictive coding [35,36], which integrates four sources of information: the current (bottom) up observation, top-down predictions from the higher hierarchical layers, lateral information from the previous sequence of items, and a (fixed) model of movement dynamics (see the Methods section).

The second hidden layer of the hierarchical model encodes *sequences* of items, which are inferred by considering three sources of information: lateral information from the sequence inferred at the previous time step, bottom-up information from the last inferred item, and top-down information from the currently inferred map.

The third hidden layer of the hierarchical model encodes *maps* (i.e., clusters of items) using a probabilistic mixture model: a way to represent the fact that different sequences of items can belong to different, mutually exclusive clusters (i.e., different spatial trajectories can belong to different spatial maps). The "content" of each map is a probability distribution over an

arbitrary structure. In our simulations, the structure is simply a vector of items, which therefore has no fixed semantics. We plot it as a 2D grid to facilitate convenient interpretation as a hippocampal spatial map (see Figure 1). During learning, the maps acquire information about the relations between its elements (items), based on two factors: *input statistics* and *structural priors*. The former (input statistics) correspond to the sequences of observations about the agent's locations in the maze that the model receives during learning. The latter (structural priors) depend on the way the model infers the next item. During inference of the next item, the model makes two assumptions: first, that the locations of successive items change gradually (and hence there are only "small gaps" between consecutive locations) and seconds, that there is no preferred directionality between successive items; see equation 8 of the Methods section for a formal description of these assumptions. This implies that the model is preconfigured (i.e., has *inductive biases*) with spatial-topological constraints, such as the fact that sequences of items (trajectories) have few spatial discontinuities and can be bidirectional. Both input statistics and structural priors contribute to gradually turn naive vectors of items into maps that encodes a space's topological information and adjacency relations between locations, as commonly hypothesized for hippocampal maps [37].

Furthermore, maps do not simply represent items in spatial locations but also their *behavioural relevance*. This is because the items have associated probability distributions that are updated using visitation counts during both real and fictive experiences (explained in detail later) and come to reflect the future expected occupancy – or empirical priors in Bayesian parlance (but note that using other methods, such as algorithmic complexity [38–40] and successor representations [41], would produce similar results as visitation counts). This expectation is typically zero at locations that cannot be crossed (e.g., walls) and high in frequently visited locations, such as goal sites and other behaviourally relevant locations, such as junction points or bottlenecks [42]. The fact that maps encode behavioural relevance reflects the idea that some aspects of value are incorporated into the hippocampal map and influence replay statistics [43]. This is an aspect that we will discuss below, in the section on "Biological interpretation of generative replay and supporting evidence".

Finally, it is worth noting that the third layer of the hierarchical model continuously infers which map the agent is currently in. For this, the model maintains a (categorical) probability distribution over maps (or clusters) and updates it whenever it receives new observations, real or fictive (i.e., produced by generative replay). The map having the highest probability is the currently inferred one; for example, in Figure 1A, the currently inferred map is the fourth. Crucially, at each moment in time, only the currently inferred map is updated during learning and used for generative replay. This selective learning strategy permits acquiring well-differentiated maps.

Obviously, the quality of learning also depends on the number of clusters of the model. In our main simulation, the generative model is provided with five clusters. This is done to facilitate the analyses of the correspondences between maps and mazes, under the hypothesis that each cluster will "specialize" in one of the five mazes. However, we also generalize this approach to an arbitrary number of clusters, using a nonparametric method that automatically selects the best number of clusters, given the agent's observations (see the section on "Nonparametric model").

Training procedure

The four learning agents are trained in the same way *within* blocks: they receive 20 spatial trajectories from the maze they are currently in as input (experiencing each trajectory just once in pseudorandom order and one location at a time) and use these data to train their

generative models. Note that this is a remarkably small number of learning trials compared to alternative models of hippocampal function [29,44,45], which require several hundreds or thousands of examples.

What distinguishes the four agents is how they update their generative model *between* blocks, see Figure 1B. The first (baseline) agent does not update the generative model between blocks. The other three agents use replay to further train their generative models offline: they replay 20 trajectories (to make them commensurate with real experience; see also the control simulations in the Methods section) and use these fictive data to update their generative models, in exactly the same way they do with real trajectories. However, each agent uses a different replay mechanism. The second agent uses *experience replay*: it replays trajectories randomly sampled from a full experience history. The third agent uses *generative replay*: it resamples trajectories from its currently inferred map, i.e., the one corresponding to the cluster with maximum probability (note that as the probabilities are continuously updated, successive replays can be generated from different learned maps). The fourth agent uses *prioritized generative replay*: it samples trajectories from its generative model, but using *prioritized maps* to initialize a replay instead of the standard maps used by the third agent. Prioritized maps are auxiliary maps (one for each cluster) that encode the average *surprise* that the agent experienced at each location during navigation, rather than visitation counts. Average surprise accumulates (Shannon) surprise derived from the probabilities encoded in the maps, for each location. Using prioritized maps permits resampling trajectories corresponding to the portions of the maze that were sampled less frequently during navigation and hence might be coded imperfectly, similar to prioritized sweeping in reinforcement learning [46], hence potentially explaining why novel sequential activations were replayed more frequently after exposure [12,13].

Simulation results

The results of the continual learning experiment are shown in Figure 1C-F. Figure 1C shows the agents' learning error, measured as their ability to correctly infer what maze generated their observations (i.e., which maze they were in) – which in Bayesian terms, is a form of *latent state (map)* inference [25,47,48]. To compute learning error for a trial we first establish the agent's best matching cluster for each maze at the end of learning, which we consider to be a "ground truth," given it is the cluster the agent would select in the limit with infinite data. Note that we select the single highest probability cluster despite the fact that in some cases, and most notably when using the non-parametric model, it is possible that several clusters have non-negligible probabilities for a given maze. Next, we count for each step how many times the agent correctly inferred the "ground truth" cluster, after the first 10 observations of each trial (but we exclude the first 5 trials of each block from this analysis). All the agents showed a good degree of contextual learning and a good memory of the old mazes after learning novel mazes. This capability to counteract catastrophic forgetting stems directly from the possibility to organize experiences into mutually exclusive contexts or maps and is therefore common to all the agents. However, the two agents using generative replay and prioritized generative replay outperformed the baseline agent ($t=3.47$, s.d.$=0.10$, $df=30$, $p=0.002$ and $t=4.18$, s.d.$=0.09$, $df=30$, $p<0.001$, respectively) and marginally outperformed the experience replay agent ($t=0.83$, n.s. and $t=1.35$, n.s., respectively). Finally, we found no significant differences between generative replay and prioritized generative replay agents ($t=0.587$, n.s.).

Figure 1D shows the same learning error, but computed during a separate (retention) test session executed after the agents experienced all five blocks. During the retention test session,

we reset the cluster probabilities after each trial; hence the agent treats each new trajectory as independent. This makes the test retention particularly challenging, as the agents cannot accumulate trial-by-trial evidence about the maze they are currently in. Furthermore, learning novel mazes could lead to catastrophic forgetting of previous mazes. Again, the two agents using generative replay and prioritized generative replay outperformed the baseline agent ($t=3.47$, s.d.=0.09, $df=30$, $p=0.002$ and $t=4.13$, s.d.=0.08, $df=30$, $p<0.001$, respectively) and marginally outperformed the experience replay agent ($t=1.37$, n.s. and $t=1.84$, n.s., respectively). Finally, we found no significant differences between generative replay and prioritized generative replay agents ($t= 0.401$, n.s.).

Figures 1E and 1F characterize the dynamic effect of evidence accumulation *within* trials. They plot the performance of the agents during the test session during the first 25 time steps of each trial, during which agents gradually accumulate information about both the maze and the trajectories they observe (recall that agents receive one observation at a time and never observe entire trajectories). Our results show that as they accumulate more evidence within a trial, all agents demonstrate increasing maze inference accuracy (Figure 1E) as well as increasing likelihood of generated sequences (Figure 1F). The latter is equivalent to reconstruction accuracy, which is a performance measure widely used to train autoencoders and generative models in machine learning. However, the two agents using generative replay and prioritized generative replay outperform the baseline and experience replay agents, both in terms of speed and accuracy of inference.

The benefits of generative replay and prioritized generative replay can be further appreciated by analysing the clusters they form during learning (see Figure 2). As shown in Figure 2A, all the agents tend to select one cluster for each maze, but the two agents using generative replay and (especially) prioritized generative replay do this more systematically. The analysis of the cluster-maze "confusion matrices" of the four agents (Figure 2B) shows that the two agents using generative replay and prioritized generative replay disambiguate more clearly between the five mazes they experienced.

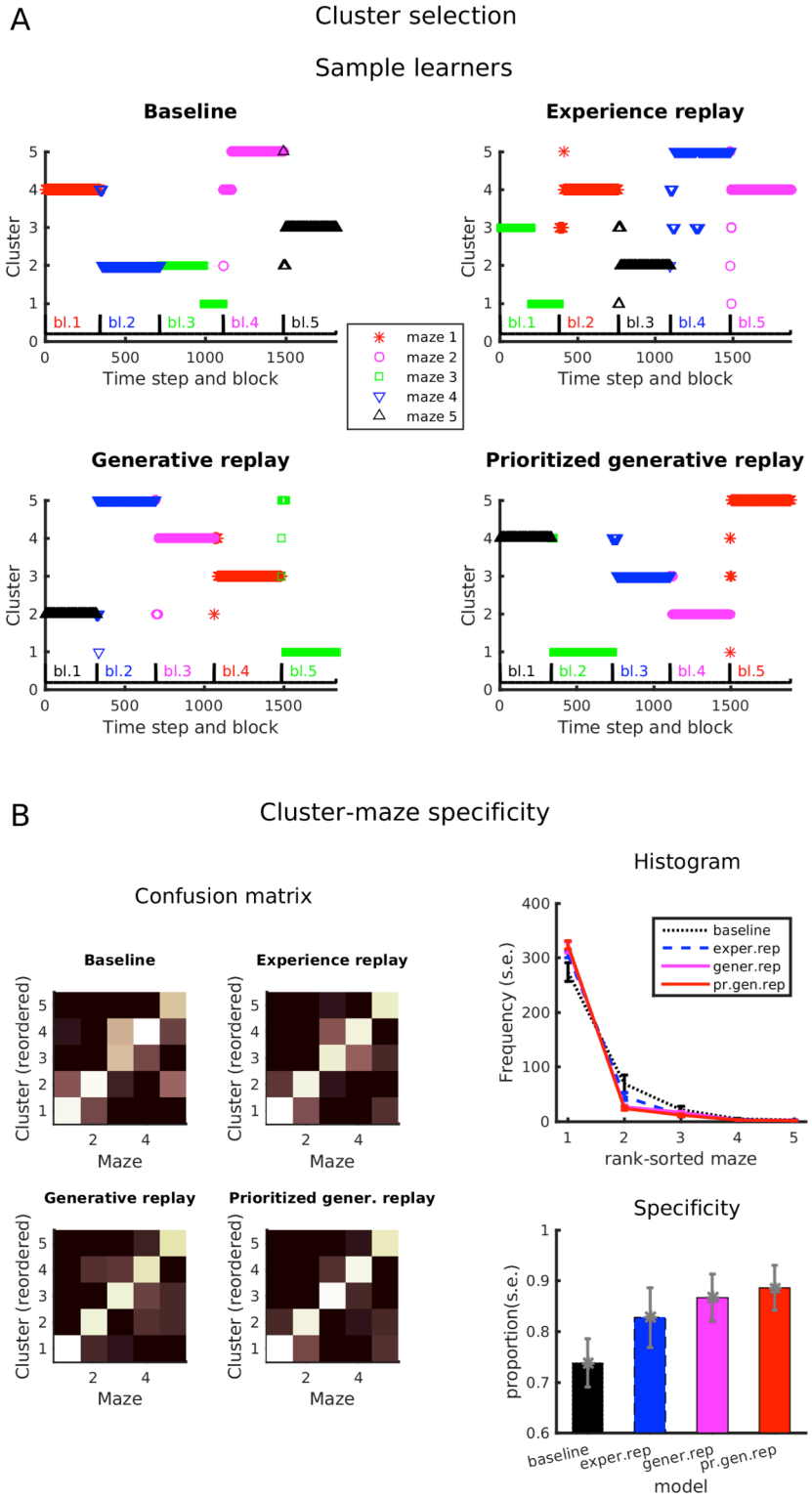


Figure 2. Cluster specificity during the continual learning experiment. (A) Cluster selection during navigation in sample replicas of each agent. The abscissa is double-labelled with time step (below) and colour coded blocks (above). The ordinate indexes the clusters (1 to 5) selected during the blocks. The five mazes correspond to different colours and symbols. In most cases, agents select different clusters for each block (especially after offline replay). (B) Analysis of cluster-maze specificity across all n=16 replicas of the four learning agents. Left: average cluster-maze confusion matrices. Right: Average frequency of cluster selection across clusters as a function of cluster-specific maze-preference rank (top) and proportion of the top-rank cluster selected during navigation (bottom).

Figure 3A provides examples of maps learned by generative replay and prioritized generative replay agents. These maps encode spatial-topological information consistent with the actual mazes and the *probability* of items at given locations; probabilities are colour coded, with white (black) corresponding to higher (lower) probabilities. This probabilistic information will become important later, when we discuss biological aspects of generative replay.

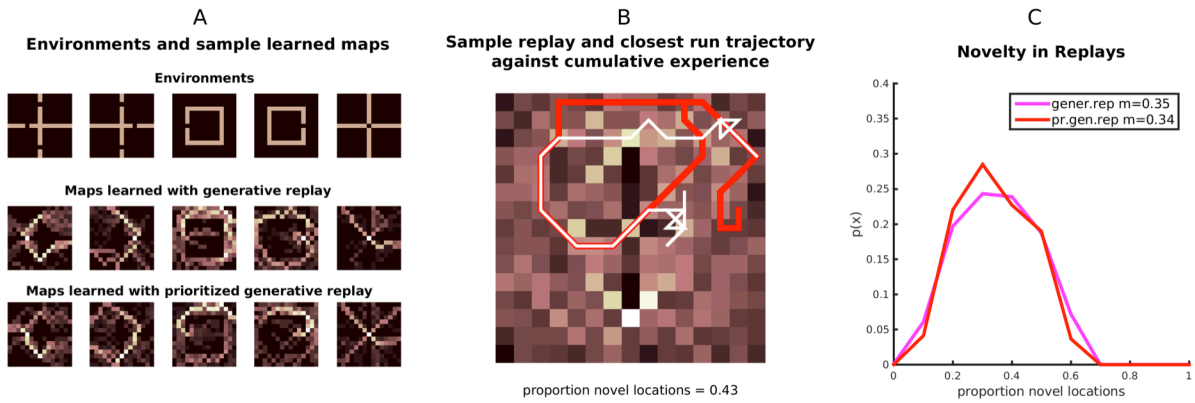


Figure 3. Examples of maps and replays of generative replay agents. (A) The five mazes used in the simulation (top) and some example maps learned by generative replay and prioritized generative replay agents. Colours represent probabilities, from white (high) to black (low). (B) Example novel trajectory produced during generative replay (red) superimposed on the closest experienced trajectory (white). (C) Quantification of the amount of novel map locations visited during generative replay.

The computational advantages of generative replay and prioritized generative replay come from the fact that (unlike baseline and experience replay agents) these agents can use their learned models to generate novel trajectories during replay. Figure 3B shows an example novel trajectory produced by generative replay, superimposed with an actually experienced trajectory. The percentage of novel locations virtually explored during replay is quantified in Figure 3C.

From a machine learning perspective, the possibility to generate novel trajectories might seem a limitation rather than a feature, given that such trajectories are sampled from an imperfect generative model and are therefore less faithful to underlying world dynamics compared to experience replay. Furthermore, generative replay might sometimes generate novel trajectories that are not semantically meaningful, e.g., trajectories that pass through walls. However, this is where the previous input statistics and the *structural priors* of the inference pay dividends. As discussed above, while the maps do not include any prior knowledge of the to-be-learned mazes, the model's inferential process is preconfigured with spatial-topological constraints, such as the fact that trajectories have only small gaps and can be bidirectional. By generating novel trajectories that reflect such spatial-topological constraints, generative replay and prioritized generative replay agents can effectively generalize their map knowledge beyond their actual experiences, for example, filling in gaps between adjacent locations or reversing their direction. This is similar to other models that permit the application of multiple episodic memories to improve generalization [49].

Nonparametric model

So far, we discussed a generative model with a fixed number of clusters (five). However, we conducted the same continual learning experiment using a nonparametric extension of the model, which permits an open-ended expansion of the number of clusters as the agent receives novel observations (see the Methods section for details). Unlike the previous simulations, here each maze was presented 4 times each, across 20 navigation blocks. This was done because nonparametric methods generally require more training data than parametric methods (as they need to infer the appropriate number of clusters). We trained 16 replicas of each agent.

The performance of the nonparametric method approximated the parametric method both during learning and retention; compare Figures 4C-D and 1C-D. During learning, the agents with generative replay and prioritized generative replay outperformed the baseline agent ($t=2.47$, s.d.=0.10, $df=30$, $p=0.02$ and $t=4.04$, s.d.=0.09, $df=30$, $p<0.001$, respectively) and the experience replay agent ($t=1.13$, s.d.=0.09, $df=30$, n.s. and $t=2.67$, s.d.=0.08, $df=30$, $p=0.01$, respectively). The analysis of the retention test showed an even stronger advantage of the generative approach: the generative replay and prioritized generative replay agents outperformed the baseline agent ($t=2.89$, s.d.=0.09, $df=30$, $p=0.007$ and $t=4.18$, s.d.=0.09, $df=30$, $p<0.001$, respectively) and the experience replay agent ($t=2.76$, s.d.=0.07, $df=30$, $p=0.01$ and $t=4.68$, s.d.=0.06, $df=30$, $p<0.001$, respectively). The two agents using generative replay and prioritized generative replay showed no significant differences in either learning ($t=1.437$, n.s.) or retention ($t=1.487$, n.s.).

As expected from a nonparametric method, all the agents showed some redundancy in their learned clusters (Figure 4A) but were still able to reliably map each maze to at least one cluster – as evident in the confusion matrices in Figure 4C. Similar to our parametric results, the two agents using generative replay and prioritized generative replay developed more specific clusters, which may explain their improved performance. These results illustrate that our approach to continual learning does not necessarily require a-priori information about the number of clusters.

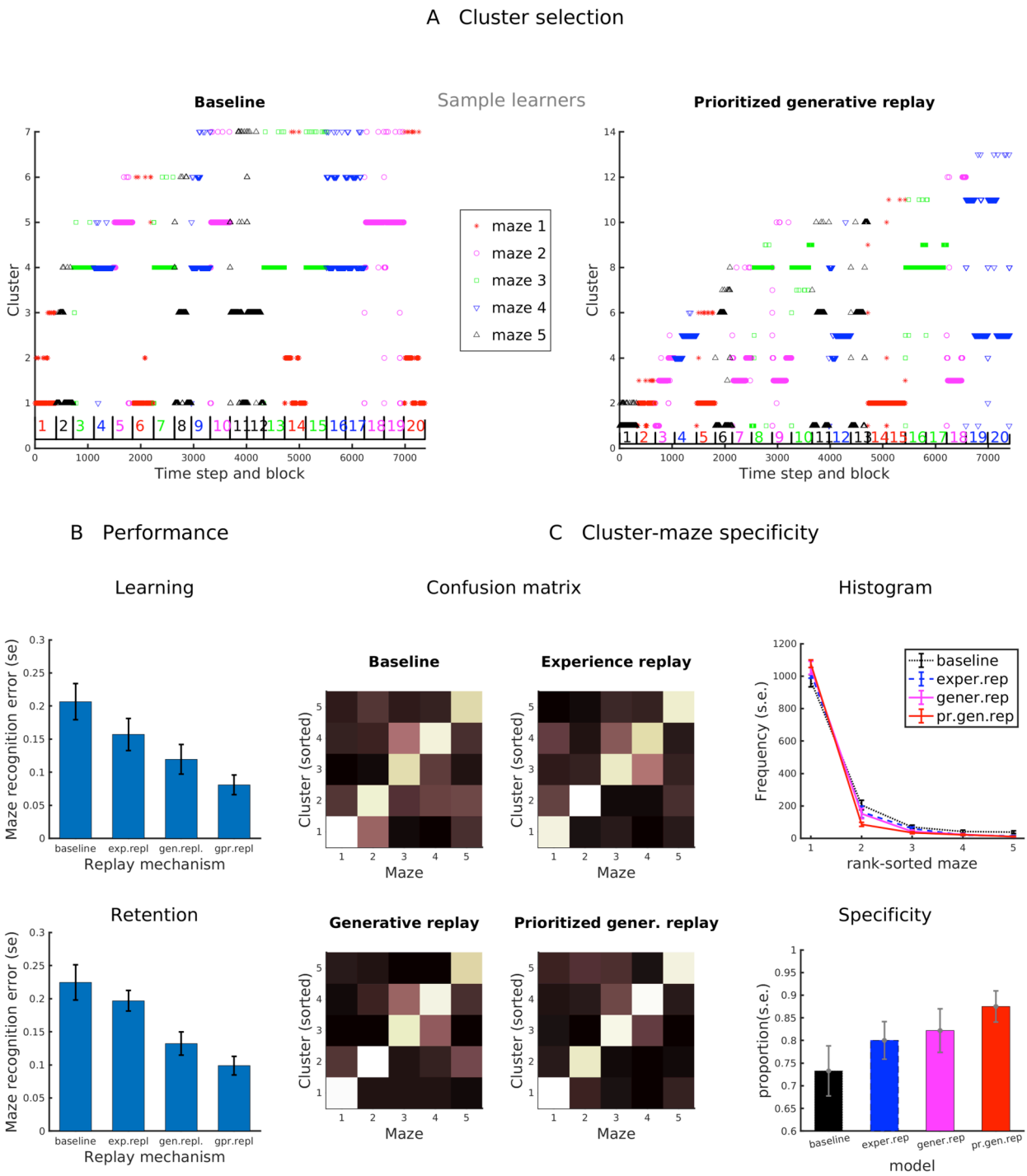


Figure 4. Clustering emerged during continuous nonparametric learning. (A) Cluster selection during navigation in sample baseline and prioritized generative replay agents. Note the greater maze-cluster consistency in the latter agent. (B) Performance of the four models during learning and retention (as in Fig 1C,D). (C) Analysis of the consistency of cluster-maze selection across all replicas of the four learning agents. Left: Average cluster-maze confusion of the clusters that are most frequently selected during the navigation in each maze (note that this analysis regards in total n=5 clusters per learner). Right: as in Figure 2, average frequency of cluster selection and proportion of top maze-specific clusters selected.

In sum, the continual learning experiment shows that agents endowed with a hierarchical generative model and generative replay mechanisms can efficiently learn and infer multiple

mazes using self-training, consistently outperforming the widely used experience replay method. Furthermore, our simulations show that the generative model supports spatial map formation, inference, and the generation of fictive experiences. Below we discuss how these inferential mechanisms map to key aspects of internally generated hippocampal sequences observed empirically.

Neurobiological underpinnings of generative replay

We have argued that generative replay provides a novel perspective on internally generated hippocampal sequences, suggesting that they may be better understood as *sequences resampled from a latent model (prior) distribution*, rather than the verbatim replay of previous experiences as commonly assumed. Here, we explain some empirical implications of this novel perspective.

Generative replay when external stimuli are absent or weak explain replay during sleep and their sensitivity to goal information

Generative replay functions both when stimuli are absent and when they are present, but in different manners. When external stimuli are absent or weak (e.g., when the animal sleeps), generative replay produces sequences entirely composed of fictive observations; this mechanism is sometimes called an *inputless decoder* in machine learning. Each episode is generated by a single map (i.e., the map that currently has the highest probability) and starts from an item / location sampled from this map. This is important because the maps used by the generative replay agent encode the probability of items at given locations (i.e., empirical priors in Bayesian terms), reflecting their behavioural relevance. Hence, in the absence of external stimuli, generative replay events will tend to start from behaviourally relevant locations, such as goal locations and frequently visited junction points of the problem (see Figure 5). This explains the finding that *hippocampal replay during sleep* tends to start form rewarded locations and proceed "backwards" to the trial's start position [50].

The maps used by the prioritized generative replay agent to initialize replays encode (Shannon) *surprise* rather than probabilities. However, after goal-directed navigation, when new rewards are encountered (as in [50]), these surprise maps tend to peak at similar, behaviourally relevant locations (e.g., goals), since the (prior) model does not yet reflect these goal locations. As a result, the prioritized generative replay agent behaves similar to the generative replay agent. Crucially, by selectively resampling sequences that include goals, both generative replay agents produce "biases" in their maps, leading them to over-represent goal and other behaviourally relevant locations (or items within them), above and beyond actual experience. This mechanism is not dissimilar from reinforcement learning models that assign (backward) replays a role in learning goal-directed policies, but proposes a putative mechanism (generative replay) rather than just a formal analysis [50,51].

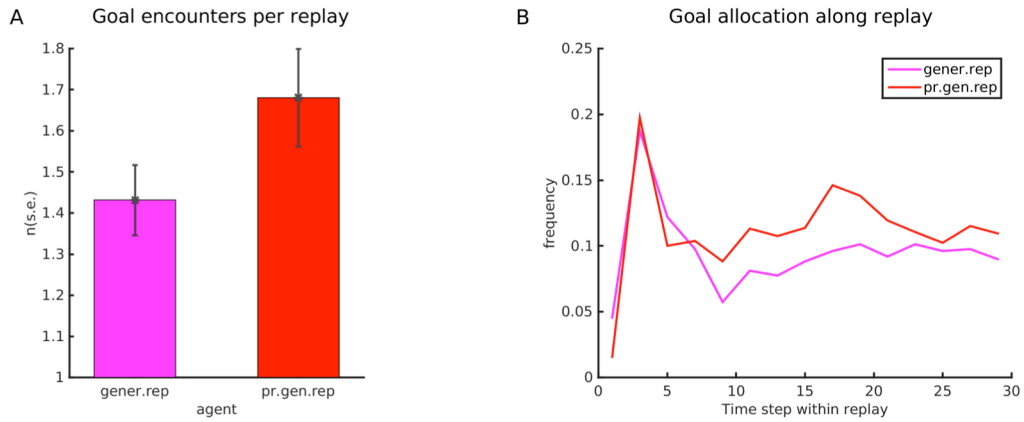


Figure 5. Analysis of goal-sensitivity during generative replays in the absence of external stimuli (akin to hippocampal replays during sleep). (A) Frequency of fictive visits to goal locations during generative replay. (B) Frequency of fictive visits to goal locations during the generative replays (lasting 30 time steps for simplicity). Generative replays tend to start close to goal locations, akin to what shown empirically; but show also goal sensitivity during the whole fictive trajectories.

Generative replay in the context of a behavioural task explains wakeful replay and theta sequences

When an animal is engaged in a behavioural task, it receives sequences of observations (e.g., previous and current locations), which can provide a context and a starting point for generative replay. In this case, generative replay may support forward prediction by producing fictive observations (conditioned on the currently inferred map) that start from the currently inferred location and proceed along a path of high probability.

Hence, generative replay in the context of a behavioural task can potentially explain two important hippocampal phenomena. First, it can explain *replay during the waking state*. As opposed to replay during sleep, replay during wakefulness often starts around the animal's position and is biased towards goal locations [10]. This can be explained by considering that while awake, the generative model may receive some weak observations (e.g., about the current location), which could be used to determine the starting point of a generative replay. Because the latent maps are biased, sequence predictions would proceed along high probability paths, potentially reaching behaviourally relevant items or locations. In this perspective, the differences between replay during wakefulness and sleep would only depend on the ways the generative episodes are initialized: by sampling from a map with or without the current observation as a seed.

Second, generative replay in the context of a behavioural task can explain hippocampal *theta sequences*. These are ordered sequences of place cell activation within each hippocampal theta cycle and occur during navigation [52]. Theta sequences have been implied in both memory encoding and the prediction of future spatial positions along the animal's current spatial trajectory [53–55]. Our computational model entails a novel perspective on theta sequences, suggesting that they interleave the (bottom-up) recognition of the current sequence and the (top-down) generation of predicted future locations; see [56] for the related suggestion that these two processes may be realized in early and late phases of each theta cycle, respectively. In other words, the first part of a theta cycle would play the role of a "filter" (requiring

bottom-up inputs) to represent the previous and current location, whereas the second half of the cycle would play the role of a "predictor" (requiring top-down inputs) to represent future locations. A speculative possibility is that the theta rhythm itself acts as a gain modulator for bottom-up streams [57], by increasing their precision during the first half of each theta cycle and then decreasing it during the second half – thus only allowing top-down predictions to occur in this second half. Interestingly, it has been reported that theta sequences are biased towards goal locations [58], which in our model would depend on the fact that the maps generating top-down predictions are biased.

In sum, our model suggests that replay during sleep, replay during wakefulness and theta sequences may all be manifestations of the dynamics of the same hierarchical generative model operating in different conditions. Replay during sleep are the "purest" form of (inputless) generative replay, whose starting point is sampled from the probabilistic maps. Note that during learning, our agents replay sequences from multiple maps and not just the most recently experienced one (they select a map for each replay, based on current cluster probabilities). This mechanism nicely explains why hippocampal replays can include distal experiences rather than just the last visited (portion of the) maze [14]. Replays during awake state take weak bottom-up input (e.g., current position), which becomes a likely starting point for the generative replay. Theta sequences would arise spontaneously while the agent is engaged in navigation as a manifestation of the interleaved bottom-up and top-down dynamics of the generative model, with the former acting as a filter to encode the past and the present and the latter acting as a predictor to encode the future. All these internally generated sequences are intrinsically biased, as the maps from which they are generated encode the relevance of items or their surprise in prioritized maps (or a combination of relevance and surprise in an extended model that may use both during replay, see the Discussion).

Replay statistics: similarity of spontaneous activity before and after navigation

A recent study showed that hippocampal replay statistics *after* navigation in a novel maze are better explained by sequential dynamics that already existed in the hippocampus *before* navigation, than by recent navigational experience [59]. Specifically, the study reported rank-order correlations between pre-run hippocampal activity (i.e., replays measured before navigational experience) and post-run activity (i.e., replays measured after navigational experience) to be significantly greater than rank-order correlations between run activity (i.e., hippocampal activity measured during navigational experience) and post-run activity. This finding suggests that hippocampal temporal coding depends primarily on selective changes to preconfigured network dynamics [60]. This result can be compared to a dynamic seen in our model, where preconfigured clusters can spontaneously produce repertoires of temporal sequences even before navigational experiences. The cluster that expresses the repertoire that best fits a specific maze tends to be selected (and later adapted) to represent this maze.

To test whether our model reproduces the above empirical findings, we performed the same rank-order analysis as in [59] on our generative replay and prioritized generative replay agents. We firstly calculated temporal rank-order maps (rank-maps, for short) of all *observation* units before navigational experience (pre-run), during navigational experience (run) and after navigational experience (post-run) for each of the five experimental blocks. We then measured the correlation between pairs of rank maps: *post-run* with *pre-run*, and *run* with *post-run*.

To calculate the ranks, we used either the observed location (*observation* in Fig. 1) during experienced trajectories or the predicted location during generated trajectories. For each trajectory, the observation-level units participating in the trajectory were ranked according to

their temporal order of activation in the trajectory. Figure 6A shows the rank-maps of a sample agent during a navigation episode, with the colour of each cell representing the rank of the observation in the corresponding location, from black (lowest) to white (highest) rank.

Spearman's rank order correlation was used to compute the correlations between the post-run and pre-run rank maps and between run and post-run rank maps, separately for each learning condition, replica, and block. Figure 6B shows statistical plots for each condition and episode across all 16 replicas per condition. The post-run – pre-run correlations were compared with the post-run – run correlations using two-sample t-tests and effects of block were explored using a linear regression of the correlations with block and condition as predictors, i.e. Rank-Correlation ~ condition + block.

Consistent with previous findings [59], the average pre-run – post-run rank correlation across conditions, blocks, and replicas (0.66, s.d.=0.15, n=128) was significantly greater than the average post-run – run correlation (0.36, s.d.=0.17, n=127; t=17.05, p<0.001). Note that while the original rodent study [59] focused on a single maze, here we calculated correlations across multiple, sequentially learned mazes (in pseudorandom order). We also found that the post-run – run rank correlations steadily decreased across blocks in a linear regression fit (β=-0.05, t=-3.9, p<0.001), which was not the case for the post-run – pre-run correlations (β=0.02, t= 1.35, p=0.17). This constitutes a novel prediction of our framework, which remains to be tested empirically.

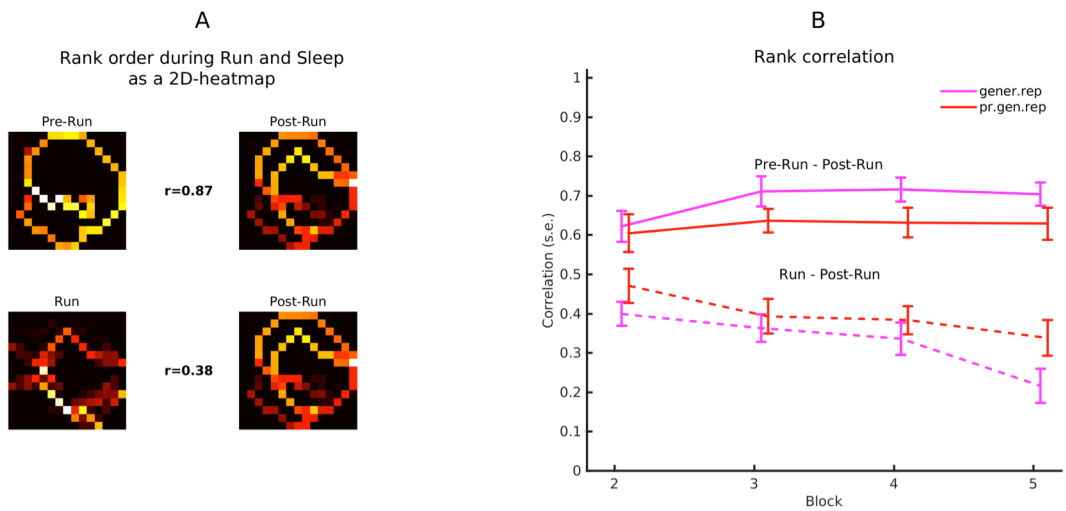


Figure 6. Network dynamics before novel navigation experiences explain replay activity better than novel experiences themselves. (A) Rank order maps of a sample learner during sample pre-run, run and post-run phases, along with rank correlations between pre-run and post-run, and run and post-run, respectively. Colours represent rank, from white (high) to black (zero). The ranks were normalized to 1 and the normalized ranks of all trajectories from a given episode were accumulated in a 2D rank-map, which was therefore normalized to sum 1. (B) Rank-correlation statistics across all n=16 replicas for each of the n=2 conditions (generative replay and prioritized generative replay) and n=5 learning blocks.

Replay statistics: spatial diffusion of replays

A recent study [11] reported that replays followed Brownian diffusion-like random trajectories even if the animal's trajectories during navigation were not Brownian. Brownian diffusion characterizes the average Euclidean distance Δx between any two locations along the path of a randomly moving particle as a power law function of travel time Δt, i.e., $\Delta x \approx \Delta t^{\alpha}$ where α is

about 0.5. This can be statistically quantified by calculating the regression slope α between the average log-Euclidian distance in space covered by neural trajectories and the corresponding log-travel time. In the animal data it was found that α ranges between 0.45-0.53 during replay (Figure 3A of [11]) and about 0.67-0.72 during navigation (Figure 4A of [11]) demonstrating that that the hippocampus can generate trajectories with statistics different from those experienced during navigation. This is similar to what happens in our model, where generative replay produces novel trajectories that are compatible with the learned maps.

We performed the same log-log regression analysis on the trajectories of our generative replay and prioritized generative replay models (considered together); see Figure 7A. Consistent with empirical findings, our model generates trajectories which approximate Brownian motion (average log-log regression coefficient across all learners and conditions α=0.46, s.d.=0.031, n=32), despite trajectories experienced during navigation reaching greater distances for the same time period (average regression coefficient α=0.75, s.d.=0.014). Figure 7B provides an intuitive illustration of these results, by showing the relative spatial diffusion of navigation and replay trajectories starting from either a central zone or the periphery (we considered trajectories starting at the 3rd time step to ensure that maps were correctly inferred). To compare the diffusion maps during navigation and replay, we calculated the average entropy of the relative diffusion maps across 5 time steps and 2 conditions (centre and periphery). Entropy was higher during replay (lower 2 rows: H = 6.63, s.d. 0.69, n=10) than during navigation (upper two rows: H = 4.89, s.d. 1.21, n=10). Overall, the quantitative and visual analysis showed that diffusion, as in animal studies, was much more directed during navigation than during replay.

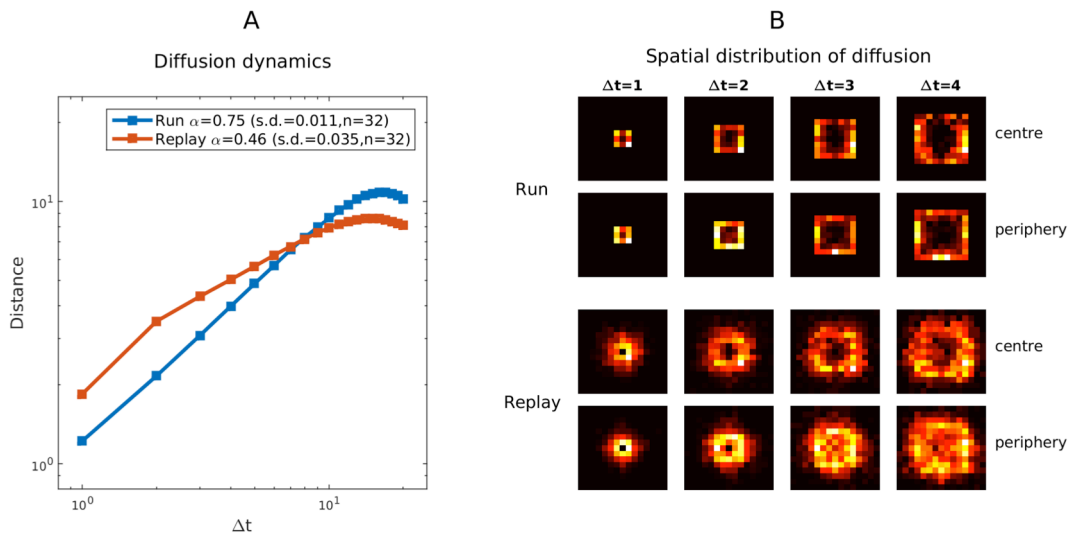


Figure 7. Trajectories during spatial experience and during replay across all learners and conditions. (A) Average point-to-point distance as a function of time-step intervals, plotted on a log-log scale. (B) Relative diffusion of locations during navigation and replay at five time steps for trajectories originating from any point of the central zone, i.e., the 5x5 central square (first and third row of panels) or the periphery, i.e., the remaining locations (second and fourth row of panels).

Putative biological implementation of the hierarchical generative model

We argued that the hippocampal formation functions as a hierarchical generative model that organizes sequential experiences into coherent but mutually exclusive spatiotemporal

contexts – or cognitive maps. In previous sections, we illustrated how the hierarchical architecture permits learning and consolidating cognitive maps using generative replay. In this section, we briefly discuss the putative neurobiological implementation of the proposed architecture; while also noting that most of its aspects are speculative and require further investigation.

Figure 8 shows a putative mapping between the components of the hierarchical generative model and the hippocampal formation, here limited to the hippocampus and the entorhinal cortex, for simplicity. The functions of the first layer may be supported by the dentate gyrus, which integrates sensory and spatial information (from MEC and LEC, respectively) to form pattern-separated, conjunctive codes for different items of experience (e.g., spatial codes such as place cells). The DG-CA3 and CA3-CA1-EC pathways have the necessary circuitry to support the encoding (recognition) and decoding (generative) processes required by our hierarchical architecture.

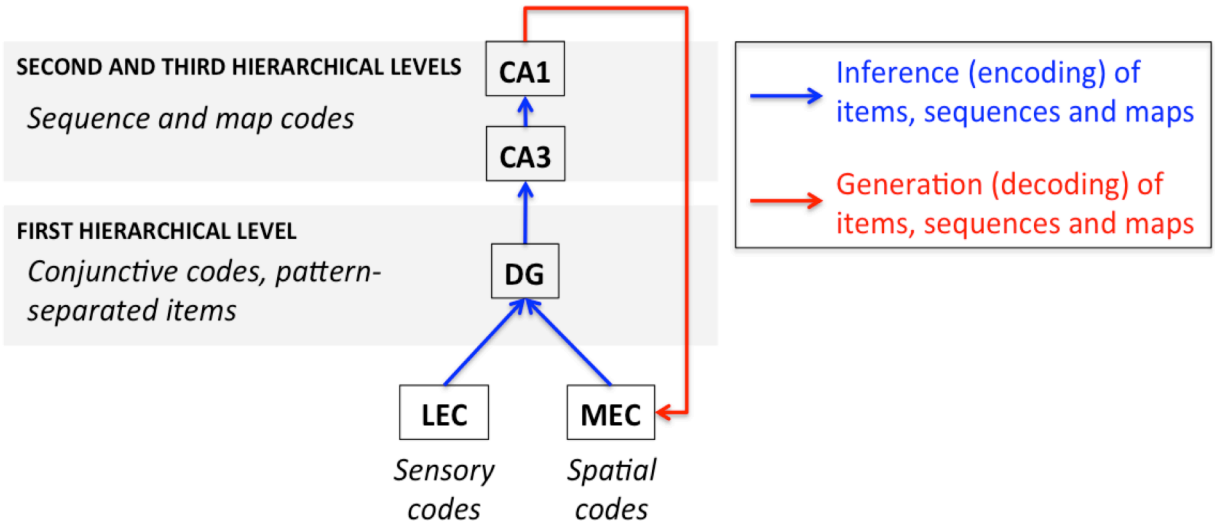


Figure 8. Putative neurobiological implementation of the hierarchical generative model (note that for simplicity, the figure only shows the EC-DG-CA3-CA1 system). LEC: lateral entorhinal cortex; MEC: medial entorhinal cortex; DG: dentate gyrus; CA: Cornu Ammonis

The model we advance here is certainly simplified. For example, our model uses simple inputs, whereas the hippocampus receives different kinds of inputs (e.g., sensory and spatial codes) from the entorhinal cortex that may already incorporate useful structure.

Furthermore, the maps that we hypothesize may be represented in a more distributed way in the hippocampal formation, with hippocampal (place) and entorhinal (grid) codes, possibly functioning as "basis functions" to encode such maps [44]. An alternative possibility is that the two different organizing principles encoded by levels two (temporal) and three (relational-spatial) in our model map to hippocampal and entorhinal structures, respectively. This would assign separate roles to the hippocampus and the entorhinal cortex, the former acting as a "sequence generator" and the latter providing a basic spatial scaffold for experiences (corresponding in our model to the "2D squares" shown in Figure 1A) or encoding (average) transition models [61]. Despite its theoretical appeal, this strict separation remains to be empirically validated, given the recurrent dynamics in the hippocampal-entorhinal circuit and the evidence of reward and goal (not just spatial or topological) information robustly coded in entorhinal cortex [62,63].

Hippocampal coding of cognitive maps

Our model assumes that the hippocampus infers cognitive maps: structured representations that define a coherent context to organize stimuli, forming spatiotemporal codes for spatial navigation [64] but also possibly non-spatial codes for other structured tasks [65–68]. While our analyses focused mostly on a parametric model that uses a fixed number of clusters, the nonparametric extension explored in the section "Nonparametric model" is better suited to capture the ability of the hippocampus to encode a very large number of experiences and to restructure itself during development and learning.

In the model, relational and contextual information is present at the level of sequences and maps, but not at the level of individual item codes. This is in keeping with evidence that in a structured, multi-compartment maze place cells code for spatial locations but not connectivity [69]. It is important to note that while item codes of different maps are visualized in the same positions of the 15x15 grids of Figure 1, they may actually correspond to the activation of different hippocampal neurons. In other words, different maps can be neurally encoded using different sets of hippocampal neurons, or by partially overlapping ensembles of neurons that "remap" across multiple maps – e.g., two neurons coding for adjacent locations in one map may code for non-adjacent locations in another map. It is this remapping mechanism that enables the disambiguation between two identical trajectories executed in two distinct spatial environments; and more broadly, avoids interference between similar experiences [70].

A second key attribute is our use of a mixture model for map probabilities, which implies that only one map is selected at a time during learning and inference. While this may be a simplification, two key phenomena – theta-paced flickering [71] and the constant cycling between representations of different trajectories during navigation [72] – suggest that place cell assemblies from different mazes are activated within different theta cycles, and map representations are not combined. Furthermore, from a computational perspective, our mixture model and alternative normative architectures, such as cloned HMM, are resistant to catastrophic forgetting [29], while our model has the advantage of high sample efficiency.

Our simulation results show that with enough capacity, the model tends to group multiple coherent experiences in a single cluster or map – which is useful for generalization and to generate novel, never experienced trajectories; see Figure 2B. However, the model does not necessarily acquire a one-to-one correspondence between clusters and mazes. Depending on the training regime and the similarity between maps, the model can in some cases develop multiple clusters for the same maze, or reuse the same clusters across multiple mazes. The first possibility – using multiple clusters for the same maze – illustrates a potential mechanism for hippocampal splitter cells [73], which show route-dependent rather than map-dependent firing. The second possibility – reusing the same cluster across multiple mazes – may explain repeating place fields often observed in repetitive (e.g., multicompartment) environments [74,75]. Furthermore, this dynamic provides a potential mechanism for generating never-observed sequences that cross two mazes, or two portions of the same maze that were never experienced together, e.g., shortcut sequences [14]; see also [76]. These shortcut sequences were empirically found while animals independently learned two portions of the same maze – where distal cues (not included in our model) may have facilitated the integration of these memories and avoided segregation into distinct maps [14]. More broadly, reusing the same clusters across multiple mazes provides a powerful mechanism for the rapid learning of novel spatial maps from a very limited set of experiences – or even prior to navigation experience [16].

Systems level perspective

Other brain areas outside those shown in Figure 5 may be related to our proposed generative model. While internally generated hippocampal dynamics are largely dependent on intrinsic (CA3) dynamics, they are usually coordinated with other brain structures, such as medial entorhinal cortex [77] and mPFC [78,79]. Prefrontal structures may play a facilitating role for hippocampal sequences to occur when needed, for example during decision-making [54], and hence mediate prospective (planning) and retrospective (learning) functions of hippocampal replay [5,18–20]. Furthermore, coordinated sequential replay across the hippocampus and PFC [79] and hippocampus and ventral striatum [80] may support goal-directed spatial navigation [81,82]. The information exchange between hippocampal and cortical structures may also support bidirectional learning. It has long been proposed that the hippocampal model can train cortical semantic systems or behavioural controllers offline. In turn, cortical systems can transfer structured, rule-related knowledge to hippocampal systems [19], potentially biasing their content and reorganizing experiences according to learned rules [17]. A theoretical possibility that remains to be tested in future research is that hippocampal and cortical structures both implement generative models that replay simultaneously and train each other during periods of coupled oscillations.

Discussion

We have proposed a novel theory of the hippocampal formation as a hierarchical generative model of spatiotemporal experiences. The theory proposes that the hippocampal generative model is hierarchically organized and includes three layers of hidden states: *items* (that disentangle representations), *sequences* (that organize items in time), and *maps* (that organize sequences in space). A key role of the hippocampal generative model is therefore organizing individual items of experience coherent spatiotemporal contexts, allowing the model to link multiple observations experienced across different navigational episodes [14]. This idea is compatible with the well-established roles of the hippocampus in cognitive map formation [83], relational tasks [84] and the spontaneous generation of sequential dynamics [21] making it a quintessential "sequence generator" [18,21].

Furthermore, the theory suggests that the hippocampus maintains several independent maps that correspond to different relational contexts in which experiences were made. From a computational perspective, the capability to learn and discriminate different contexts is key to learn multiple experiences while avoiding catastrophic forgetting [29]; whereas from a biological perspective, it might be key to support episodic and autobiographical memory [24]. Importantly, in our model, map learning starts with the selection between multiple preconfigured maps (which might sometimes correspond to different sets of neurons [71]). This favours the selection of maps that are already preconfigured for novel mazes, hence explaining why network dynamics before novel navigation experiences explain replay activity better than novel experiences themselves [59], see Figure 6. This approach has some resemblances with "selectionist" theories of learning [85], with machine learning approaches such as *reservoir computing* that select between a preconfigured set of dynamics [86–88] and the "lottery ticket" hypothesis that learning in large neural networks is largely driven by subnetworks having fortuitous initializations that render them appropriate for the task at hand [89].

Finally, the theory proposes that the hippocampal generative model supports *generative replay*: a powerful method for continual learning that uses fictive sequences generated from a learned model to train the model itself, or other (e.g., cortical) models or controllers. Accordingly, spontaneous hippocampal dynamics are manifestations of generative replay.

This theory is appealing from both machine learning and biological perspectives. From a machine learning perspective, generative replay prevents catastrophic forgetting [90] without requiring the verbatim storage of (an unbounded number of) previous experiences, as in "experience replay". Our results show that generative replay and prioritized replay agents outperform baseline agents (that do not use replay) in all our tests and also agents using experience replay (i.e., replaying from a memory buffer) in the most challenging, nonparametric setting.

Furthermore, our simulations show that generative replay mechanisms are well suited to explain the flexibility of internally generated hippocampal sequences, which are difficult to reconcile with previous theories that consider them rigid "replays" of previous experiences. This is supported by three observations. First, the hippocampus can generate novel sequences that show prospective aspects. These are explained in our model by considering that sequences follow high probability paths, and the maps from which sequences are generated implicitly encode behaviourally relevant or salient locations [10,15,43,50,58]. From a conceptual perspective, while sequences encode short-term predictions, maps encode long-term predictions of future locations that average across several episodes; this renders our idea compatible with the view that the hippocampus learns a predictive representation [91] but casts it within a hierarchical inference scheme.

Second, in our model maps can express a rich repertoire of sequential dynamics even before a specific navigational episode. The map that best fits the novel navigational episode is then selected and adapted to incorporate the experience. This mechanism implies that sequences generated before a navigational episode (preplays [16]) and after it (replays) are highly correlated, as demonstrated empirically [59] – which points to the importance of preconfigured sequential dynamics to explain fast learning in the hippocampus [12,13].

Third, in keeping with empirical findings [11], the sequences produced during generative replay follow Brownian, diffusion-like dynamics, while navigational experiences do not. This finding further exemplifies the generative nature of hippocampal sequences, which do not simply recapitulate recent experiences, either in terms of content or dynamics.

It remains to be fully assessed which of our generative replay models (or perhaps a combination of both) best explains the phenomenology of hippocampal sequences. Interestingly, a previous formal theory considered replays to be guided by two terms: a *need* term that prioritizes behaviourally relevant experience and a *gain* term that prioritizes novel experiences; but it did not provide a mechanistic implementation of this idea [51]. By noticing that need and gain terms have close correspondences with the functioning of generative replay and prioritized generative replay, respectively, our proposal could shed light on the potential mechanisms supporting these two factors that contribute to explain replay. Another intriguing possibility that remains to be tested is that the two generative replay mechanisms may be selected in context-dependent ways, depending on the activity of different neuromodulators, which are known to influence replay activity [92].

Previous theoretical and computational studies have assigned a role to the hippocampus in supporting model-based, goal-directed navigation [18,19,34,48,55,93–96]. Our proposal is consistent with these ideas but goes further by introducing the idea of hierarchical, structured representations of sequences and maps, and by explicitly modelling how these representations are learned over time. Generative replay shares resemblances with another method used to train agent controllers offline: the self-generation of experiences from a learned *transition model* of the task, as in the DYNA architecture of reinforcement learning [97]. This idea has inspired previous proposals of the role of spontaneous hippocampal activity in optimizing reinforcement learning systems offline [51]. In contrast to these ideas, our theory

views the hippocampal formation as a generative model of spatiotemporal experiences, not a transition model. Furthermore, it views spontaneous activity as a way to improve the generative model [18,94], rather than reinforcement learning controllers. From a computational perspective, learning a generative model (as done in generative replay) is substantially simpler and less error prone compared to learning a transition model (as done in DYNA). Indeed, generative replay avoids the main shortcoming of transition models used to train a controller: the fact that model errors accumulate during sequential resampling, making learning unstable. That said, the above theories are not mutually exclusive. Generative replay can be effectively used to simultaneously train the generative model and reactive controllers, as exemplified by DYNA or the world model approach [98]. Furthermore, generative replay can directly support prospective functions of goal-directed controllers, such as the prediction of future paths. Therefore, our proposal is compatible with previous proposals that the hippocampus may operate synergistically with other brain structures, such as the ventral striatum, to form a goal-directed controller for spatial navigation [58,81,82].

Though we illustrated our proposed model in the domain of spatial navigation, an interesting possibility that remains to be explored is that it may support the autonomous learning of coherent spatiotemporal contexts beyond navigation as well. Recent studies have shown that the hippocampus supports relational inference [84], temporal and rate coding of discrete sequences of non-spatial events such as (in rodents) odours [99] and button locations [100] and (in humans) of arbitrary items [17]; see also [101,102]. More broadly, several findings imply a role for the hippocampal formation in cognitive mapping in several domains of cognition outside of spatial navigation [22,103,104]. It remains to be established to what extent the hippocampus is a *general-purpose* system to form maps and sequences, and whether the generalization to non-spatial domains reuses internal codes initially developed for spatial navigation [22]. Theoretical arguments suggest that while the brain might use multiple generative models [105], these may be tuned to different (natural) statistics of stimuli. At one extreme, generative models supporting visual processing may be tuned to scenes that change gradually (in the sense that successive frames tend to be similar to one another). The hippocampal generative model might lie close to the other extreme: it may be specialized to learn sequences of arbitrary (pattern-separated) items [18]. Establishing the plausibility of these theoretical arguments is an objective for future research.

Casting the hippocampal formation as a hierarchical generative model may help reconcile its roles in multiple, apparently disconnected cognitive functions, including episodic and autobiographical memory [24,84], prospective simulations and imagination [106,107] and cognitive map-based spatial navigation [83,108] which have often been studied in isolation from one another. Standard theories of system consolidation assume that episodic memories are transferred from the hippocampus to the cortex [2,3]. However, a large body of evidence indicates that episodic memories may remain dependent on the hippocampus [23]. The theory proposed here suggests a specific way episodic memories may remain depend on the hippocampus: by being instantiated in the hippocampal generative model. Despite the generative model keeps changing with novel experiences, it is still capable to re-enact (with a varying level of fidelity) previous experiences. This idea is in keeping with the hypothesis of "constructive episodic simulation", which suggests that episodic memory is an inherently constructive process and furthermore, that it provides details to construct simulations of the future - therefore explaining why episodic memories and prospective simulations share largely the same substrate and the same constructive processes [109–111]. Our model provides a mechanistic grounding to this hypothesis, by showing that generative replay supports both the re-enactment of episodic memories and prospective simulations / imagination, that is, the spontaneous generation of novel experiences that are structurally coherent with past

experiences (as they stem from the same maps). Intriguingly, the theory of constructive episodic simulation suggests also that an imperfect memory that recombines previous experiences can be more useful to simulate future events compared to a memory that stores perfect records [112]. Our results support this idea by showing that generative replay confers systematic advantages in learning and generalization compared to the most widely used mechanism of experience replay, which epitomizes perfect memory storage.

Another important aspect of the model is that it organizes experiences into coherent spatiotemporal contexts. This makes it possible to discriminate between experiences that share similar features but occurred in different contexts, which is a key feature of episodic memory. Furthermore, we argue, the very same mechanism is crucial for cognitive map formation. As discussed above, the possibility to organize experiences within independent relational maps (at the highest level of our model hierarchy) confers significant advantages, especially during the continual learning of multiple tasks. In our model, the maps may automatically include structural priors, such as the fact that there are small gaps between adjacent locations and no preferred directionality, which implicitly embody characteristics of the physical space where an animal navigate - but more speculatively, may provide a relational scaffold to organize experiences in many more domains of cognition beyond spatial navigation [22,66,103,104]. Finally, and speculatively, the hippocampal generative model may contribute to autobiographical aspects of memory, as by continuously encoding and replaying episodes it may afford a continual sense of self that persists across time. This may also link to prospection, as a model provides not just a memory of "the past me" but also inference of "the future me" [113].

In sum, the computational framework of hierarchical generative modelling and generative replay reconciles several apparently disconnected theories of hippocampal function and provides a rigorous framework for their empirical testing.

Methods

Hierarchical spatiotemporal representation

Our theory proposes that the hippocampus implements a hierarchical generative model that organizes the *observations* it receives over time into three sets of latent codes, for *items*, *sequences* and *maps*, respectively. At the highest hierarchical level, the model learns a set of maps representing characteristic environmental (e.g., maze) properties. The model implicitly assumes that, at each time (or short time interval), its observations derive from one single map that must be inferred. A key characteristic of the generative model is that it inherently implements *generative replay*: it can stochastically generate sequences of items from the inferred map; and use the stochastically generated sequences to train the model itself (or support other cognitive task, such as trajectory prediction and planning).

Before formally describing the model, we need to introduce some notation. Let us assume that the environment is represented as a vector space $\mathbb{D} = \{(\xi_1, \dots, \xi_N) \mid \xi_1, \dots, \xi_N \in [0,1]\}$ of dimension N, which in our simulations of spatial navigation correspond to a domain of N spatial locations. An *observation* is defined as an event of "magnitude" ξ_i occurring at a location $i = 1, \dots, N$. Using the standard basis $\mathbf{e}_1 = (1,0, \dots ,0)$, $\mathbf{e}_2 = (0,1, \dots ,0)$, $\cdots$, $\mathbf{e}_i = (0, \dots ,1, \dots 0)$, $\mathbf{e}_N = (0,0, \dots ,1)$, observations are represented at the lowest level of the hierarchical model as $\boldsymbol{\xi}_i = \xi_i \mathbf{e}_i$. The default magnitude ξ_i of a noiseless observation at position *i* is set to 1.

Moving upward, the next hierarchical level infers the hidden code $\mathbf{x}$ of the observations in $\mathbb{D}$, called *items*, such that

$$\mathbf{x} = \sum_{i=1}^{N} \xi_i \mathbf{e}_i = \sum_{i=1}^{N} \boldsymbol{\xi}_i \tag{1}$$

Moving upward again, the next hierarchical level uses the same encodings to define a hidden code $\mathbf{y}$ for a *sequence* of items $S \equiv \boldsymbol{\xi}(1), \dots, \boldsymbol{\xi}(T)$ representing a trajectory through $\mathbb{D}$:

$$\mathbf{y}(1, \dots, T) = \sum_{t=1}^{T} \delta^{T-t} \mathbf{x}(t) \tag{2}$$

where $0 < \delta < 1$ ($\delta = 0.7$ in our simulations) and δ^{T-t} is an exponential decay over time. Equation (2) consists of a graded representation (or trace), where the most recent observation is encoded with the highest activation rate. Note that this coding scheme closely resembles a "bag-of-words" encoding of multiple words (observations) in a document (sequence). However, while the bag-of-word representation encodes the frequency (distribution) of elements without any order ($\delta = 1$), our method assumes few repetitions such that the graded values essentially encode item order.

Finally, at the top level, the model assumes that sequences of items $\mathbf{y}$ can be grouped together, as if they were drawn from a distribution that partitions the sequences into clusters. Each cluster k corresponds to (and is characterized by) a different feature map described by a vector of parameters $\boldsymbol{\theta}_k = \left(\theta_{k_1}, \cdots, \theta_{k_i}, \cdots, \theta_{k_N}\right) \in \mathbb{D}$ encoding the frequency of event occupancy of the $\mathbb{D}$ locations under map k. Thus, the model assumes that the representation $\mathbf{y}$ of a sequence of observations is drawn from cluster k according to a probability $p(\mathbf{y}|\boldsymbol{\theta}_k)$ and that the prior probability of this cluster is $p(k)$. These assumptions are embodied in the following generative *mixture model*:

$$p_{\boldsymbol{\theta}}(\mathbf{y}) = \sum_{k=1}^{K} p_{\boldsymbol{\theta}}(\mathbf{y}|k) p(k) \tag{3}$$

where $\equiv \{\boldsymbol{\theta}_k\}_{k=1,..,K}$, $p_{\boldsymbol{\theta}}(\mathbf{y}|k) \equiv p(\mathbf{y}|k, \boldsymbol{\theta}_k)$, and K is the number of clusters. Note that while here we set K as known and fixed, in the section "Nonparametric Method" we discuss a non-parametric extension of the model, which automatically infers the number of clusters given its observations.

The definition of the model in Equation (3) is well-posed once we specify the set of mixing probabilities $p(k)$ and the conditional distributions $p_{\boldsymbol{\theta}}(\mathbf{y}|k)$. In the proposed model, we assume that $\boldsymbol{z} \equiv \big(p(k=1), \dots, p(k=K)\big) \sim Cat(\boldsymbol{\pi})$ and the parameters $\boldsymbol{\theta}_k \sim Cat(\boldsymbol{\rho}_k)$ with $\boldsymbol{\rho}_k \sim Dir(1, \dots, N)$. Therefore, both the mixing probabilities and the structural parameters of the maps follow a categorical distribution – the latter having hyperparameters drawn from a Dirichlet distribution over $\mathbb{D}$.

In turn, we define the likelihood $p_{\boldsymbol{\theta}}(\mathbf{y}|k)$ that the kth map can generate the representation $\mathbf{y}$ of sequence of observations S as a deterministic function of $\mathbf{y}$ and the probabilistic parameters $\boldsymbol{\theta}_k$ of that map:

$$p(\mathbf{y}|\boldsymbol{\theta}_k) \propto \prod_{i=1}^{N} \theta_{k_i}{}^{\mathrm{y}_i} \tag{4}$$

where $\mathrm{y}_i = \mathbf{y} \circ \mathbf{e}_i$, with the symbol "$\circ$" denoting the Hadamard (element-wise) product. By taking the logarithm, Equation (4) becomes the scalar product between $\mathbf{y}$ and $\ln \boldsymbol{\theta}_k$:

$$ln\, p(\mathbf{y}|\boldsymbol{\theta}_k) \propto \sum_{i=1}^{N} \mathrm{y}_i \ln \theta_{k_i} = \mathbf{y} \cdot \ln \boldsymbol{\theta}_k \tag{5}$$

Note that equations (4) and (5) effectively measure the compatibility of the sequence of events under the structure of a given map, on a linear and a logarithmic scale, respectively. Indeed, equation (5) assigns higher probabilities to those maps whose spatiotemporal structure $\boldsymbol{\theta}_k$ more closely matches the allocation and temporal order of observations encoded by $\mathbf{y}$.

Note that Equation (5) affords a plausible neuronal implementation in the mechanism of weighted transmission of action potentials between synapses encoding log-probabilities of the spatiotemporal structure and presynaptic neuronal activity encoding event representations.

Hierarchical inference

To infer the posterior of the latent states upon the observation of a new event, we adopt an approximate inferential approach that considers the hierarchical and temporal dependencies of the model (Figure 1a).

First, we estimate the posterior of the mixing probability distribution $\boldsymbol{z}(t) \equiv p(k|\mathbf{y}(t))$. This is done differently at the beginning of an episode (when $t = 0$) and afterwards (when $t > 0$). At the beginning of an episode, the mixing prior $\boldsymbol{z}(0)$ is drawn from a categorical distribution $\boldsymbol{z}(0) \sim Cat(\boldsymbol{\pi})$ with hyperparameters $\boldsymbol{\pi}$ that follow a Dirichlet distribution, $\boldsymbol{\pi} \sim Dir(\boldsymbol{\alpha})$. After the calculation of the posteriors, we recursively adjust the hyperparameters $\boldsymbol{\pi}$ based on the accumulated evidence, thus allowing agents to improve the a-priori information available for the next trials; see the section on "Parameter learning" for details.

In contrast, when $t > 0$ the posterior of $\boldsymbol{z}(t)$ is calculated as follows:

$$\boldsymbol{z}(t) \propto \boldsymbol{z}(t-1)^{\tau} p(\mathbf{y}(t-1)|\boldsymbol{\Theta})^{1-\tau} \tag{6}$$

Here, τ is an exponential time constant, whose value is fixed to (the empirically determined value) $\tau = 0.9$ and $\boldsymbol{\Theta}$ is the full set of maps $\boldsymbol{\theta}_k$. For the sake of simplicity, we recall Equation (5) and use a log-scale to change Equation (6) to the linear expression:

$$\ln \boldsymbol{z}(t) \propto \tau\ \ln \boldsymbol{z}(t-1) + (1-\tau)\, \mathbf{y}(t-1) \ln \boldsymbol{\Theta} \tag{7}$$

where the matrix multiplication $\mathbf{y}(t-1)\ln\boldsymbol{\Theta}$ sums over the spatial domain.

Equation (6) determines the posterior mixing probability $\mathbf{z}_k(t)$ of each cluster k at the current time step by dynamically integrating its prior state $\mathbf{z}_k(t-1)$ and the likelihood of the current sequence, which considers the compatibility between $\mathbf{y}(t-1)$ and the cluster. We adopt a maximum a posteriori (MAP) approach to select (and adapt) which of the K maps maximizes the posterior; namely $\bar{k} = argmax(\mathbf{z}(t))$. By doing so, latent state inference guides learning to gradually form distinct maps for different environments.

At the end of the inferential process at each time step, the parameters $\boldsymbol{\rho}_{\bar{k}}$ that define the selected map $\boldsymbol{\theta}_{\bar{k}}$ are updated; see the section on "Parameter learning" for details.

Note that Equation (6) can be interpreted as Bayesian filtering, by considering that it operates in two sequential steps: generative prediction and Bayesian update. We could interpret Equations 6 and 7 as the probabilistic prediction of a set of K hidden states $k = 1 \dots K$ conditioned on the sequence code $\mathbf{y}(t-1)$. However, unlike standard Bayesian filtering, the latter is another latent state, not an observation. This process dynamically reduces uncertainty in map distributions (and the entropy of $\boldsymbol{z}(t)$), hence the inference provides increasingly more precise estimates of the mixing probabilities and map inference. This first inferential step has a plausible biological implementation, in terms of a local competition circuit that recursively accumulates evidence for each map, achieved by adding synapse-weighted bottom-up signals encoding current observations at each time step.

Second, we estimate the posterior item codes $\mathbf{x}(t)$. (Note that the hierarchal structure shown in Figure 1A might suggest that the posterior sequence $\mathbf{y}$ should be estimated before the posterior item codes $\mathbf{x}(t)$. However, we do the opposite to simplify the generation of fictive information during replay, see below). The dynamics equation for $\mathbf{x}$ considers three elements: (i) the map $\boldsymbol{\theta}_{\bar{k}}$ inferred at the highest hierarchical level, (ii) a simple physical transition model specifying that the transitions are sampled from a normal distribution centred on past observation $\boldsymbol{\xi}(t-1)$ or in other words that only adjacent locations (in 2D space) can be visited next and (iii) the prior sequence code calculated at the previous time step $\mathbf{y}(t-1)$, which is subtracted, thus implementing a form of "inhibition of return". Thus, we define the posterior of $\mathbf{x}(t)$ as follows:

$$\mathbf{x}(t) = \hat{\boldsymbol{\xi}}(t) \circ \boldsymbol{\theta}_{\bar{k}} - \mathbf{y}(t-1) \tag{8}$$

where $\hat{\boldsymbol{\xi}}(t) \sim \boldsymbol{\mathcal{N}}(\mu, \Sigma)$ is a noisy prediction about the observation with a normal distribution centered on $\mu = \boldsymbol{\xi}(t-1) + c * \boldsymbol{v}(t-1)$ and with symmetric covariance $|\Sigma| = \boldsymbol{v}(\mathrm{t}-1)$ modulated by the current velocity (i.e., the velocity at the tip) of the sequence $\mathbf{y}$, $\boldsymbol{v}(\mathrm{t}) = \boldsymbol{\xi}(t) - \boldsymbol{\xi}(t-1)$ and by an arbitrary chosen coefficient $c = 0.4$ promoting sequence continuity.

After updating the item code $\mathbf{x}(t)$, the model generates a prediction for a new observation through the maximization of the inferred item, i.e.:

$$\boldsymbol{\xi}(t) = \max_i \ (\mathbf{x}(t) \circ \mathbf{e}_i) \tag{9}$$

This mechanism is key for generative replay and the generation of fictive observations over time.

Third, we estimate the posterior sequence code $\mathbf{y}(t)$, using the following dynamics equation, which considers the hierarchical structure and causal temporal dynamics of the model:

$$\mathbf{y}(t) = \delta\mathbf{y}(t-1) + (1-\delta)\boldsymbol{\theta}_{\bar{k}} + \boldsymbol{\xi}(t) \tag{10}$$

where δ is the decay coefficient already introduced in Equation (2). In practice, Equation (10) gradually creates a spatiotemporal representation of observed sequences, by adding the item code $\boldsymbol{\xi}(t)$ corresponding to the current observation to the previous sequence code $\mathbf{y}(t-1)$, while also considering the top-down prediction provided by the inferred map $\boldsymbol{\theta}_{\bar{k}}$. Note that the item code $\boldsymbol{\xi}(t)$ is inferred during navigation, but fictive during replays.

Parameter estimation (learning)

In parallel to the aforementioned state-inference process, the model continuously infers its (hyper)parameters. In particular, every time the latent-state inference selects the kth hidden map via Bayesian filtering, it updates the model hyperparameters $\pi_{\bar{k}}$ and $\boldsymbol{\rho}_{\bar{k}}$. This enables long-term environmental information (e.g., maze properties) to be maintained across time.

To improve hyperparameter learning, we adopt the following methods: (i) we use a (small) decay constant γ to account for the volatility of information coded in the mixing priors (in our simulation, this is set to 1%), (ii) we scale the increments of the Dirichlet counters with the probability $z_{\bar{k}}(t) \equiv p(\bar{k})$ to account for the confidence in choosing $\boldsymbol{\theta}_{\bar{k}}$ as the best map. This update is done after the hierarchical inference step.

To calculate the posterior of the hyperparameters of the $\bar{k}$th mixing probability ($\pi_{\bar{k}}$), we use the following update:

$$\pi_{\bar{k}} = (1-\gamma)\pi_{\bar{k}} + z_{\bar{k}}(t). \tag{11}$$

Note that even if the value of $\pi_{\bar{k}}$ is updated at every time step, it is only used at the beginning of each trial ($t = 1$), as a parameter of the categorical distribution from which the prior $\boldsymbol{z}(0)$ is sampled. Therefore, equation (11) embodies a-priori information (from previous trials) about the probability of each cluster in the parameters $\boldsymbol{\pi}$.

To infer the Dirichlet hyperparameters of the selected map ($\boldsymbol{\rho}_{\bar{k}}$), we use the following:

$$\boldsymbol{\rho}_{\bar{k}} = \boldsymbol{\rho}_{\bar{k}} + z_{\bar{k}}(t)\mathbf{y}(t) \tag{12}$$

which accounts for both the temporal information in the sequence code $\mathbf{y}$ and belief $z_{\bar{k}}(t)$ of the choice of the current map. This update is done only when $\mathrm{t} > 1$, to estimate the most likely map associated with the sequence of observations observed so far.

Nonparametric method

The simulations described so far used a fixed number of mixture clusters K. Here we describe an extension of the previous model that automatically infers K. For this, we use an adaptation

of the Chinese Restaurant Process (CRP) prior over clustering of observations that adapts model complexity to data by creating new clusters as the model observes novel data [28,114]. Note that the CRP prior is applied only during navigation, on true data, and not during replay. This is done to prevent the generation of an excessive number of clusters.

The CRP metaphor describes how a restaurant with an infinite number of tables (clusters) is gradually filled in with a potentially infinite number of customers (observations). The first customer sits at the first table, i.e., it is assigned to the first cluster. All the other new customers that enter the restaurant can either sit at an occupied table (i.e., be assigned to existing clusters) with a probability corresponding to table occupancy $n_i/(n+\alpha)$ or sit at a new table (i.e., expand the model with a new cluster) with probability $\alpha/(n+\alpha)$, where α is a concentration parameter.

Notably, our customers are multi-dimensional real-value representations $\mathbf{y}$ and each representation is different from any of the previously observed ones. To determine whether a representation $\mathbf{y}$ should be considered as novel, we extended the classical CRP prior method to also exploit the knowledge stored in the learned cluster parameters: the maps $\boldsymbol{\theta}_k$. A representation $\mathbf{y}$ is considered as novel when the greatest likelihood $p(\mathbf{y}|\boldsymbol{\theta}_k)$ (see Formula 4) among all clusters currently in use is smaller than a threshold, $p_{thr} = 0.01$. Table occupancy corresponds to the parameters π_k of the mixing probabilities and the concentration parameter is $\alpha = 1$.

Rank-maps

Rank-order maps are used in [59] to characterize the similarity between pre-run (preplay) and post-run (replay), and between post-run and run (navigation experience). We calculate the rank map of each navigation or replay episode by accumulating the spatial trajectories $\{S^j \equiv \boldsymbol{\xi}(1), \dots, \boldsymbol{\xi}(T)\}_{j=1\dots M}$ represented in terms of normalized localized ranks during the episode. Thus, for a given episode, the rank map $\mathbf{R}$ is defined as follows:

$$\mathbf{R} = \sum_{j=1}^{M} \sum_{t=1}^{T_j} r^j(t)\mathbf{x}^{\mathbf{j}}(t) \tag{13}$$

where $r^j(t) = \frac{t}{T_j}$ is the normalized rank at time t of sequence j, T_j is the length of sequence j, and $\mathbf{x}^{\mathbf{j}}(t)$ is the location at time t of sequence j (defined in Equation 1). For replay episodes the predicted observation is used instead of the true observation (which is not available). Finally, Spearman's rank order correlation is used to calculate the similarity between the (vectorized) ranks maps. Examples of rank maps and rank correlations are shown in Fig. 6A and statistics of rank-correlations across conditions is shown in Fig. 6B.

Dynamics of Diffusion

To characterize the trajectory dynamics during navigation and replay, for each learner from each training condition during either navigation or replay, we collect the temporal (Δt) and Euclidean (L2) spatial distance Δx between each pair of points ξ_{t_1} and ξ_{t_2} (chunks) from each trajectory $S \equiv \boldsymbol{\xi}(1) \dots \boldsymbol{\xi}(T)$. We then fit a log-log regression model of the spatial distance as a function of temporal distance (i.e., duration): $\ln(\Delta x) \approx \alpha \ln(\Delta t)$. Note that the estimated regression coefficient α directly corresponds to the exponential coefficient of the power law: $\Delta x = \Delta t^{\alpha}$.

The estimation is applied on all trajectory chunks, up to a limit of 20 time steps long (this limit it selected given that the map size is 15x15 units; longer sequences would necessarily involve some systematic curvature) The results of the analyses are shown in Figure 4A.

Finally, we collect the spatial statistics of the distribution of diffusion relative to the initial position of each chunk. Some examples are shown in Figure 4B.

Acknowledgements

This research received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement Nos. 785907 and 945539 (Human Brain Project SGA2 and SGA3) to GP, the European Research Council under the Grant Agreement No. 820213 (ThinkAhead) to GP, and from the Italian Ministry of Research (PRIN2017 grant 2017KZNZLN) to IS. We gratefully acknowledge Jeremy Gordon for useful comments on this manuscript.

Supplementary Materials

Control simulations

We ran two sets of control simulations aiming to replicate the main simulation and investigate the robustness of the model and the effect of two parameters (number of replays, and item sampling error). Each simulation included n=32 replicas per condition.

First, we investigated the role of the number of replays by using twice, four times, half, and a quarter the number of navigation trials in a single episode (for which there were n=20). As in the main simulation, model evaluation was based on the proportion of incorrectly recognized mazes during the two phases of learning: navigation and retention. The results (see Suppl. Table 1) confirmed the overall advantage of generative replay methods, which (as expected) gradually reduces with fewer replays. However, using more replays did not consistently improve the performance, and in some cases slightly worsened it. This result may be due to the fact that an excessive strengthening of memory traces might decrease choice flexibility.

Second, we investigated the role of the displacement coefficient c that controls sequence continuity in formula (8) by using either half or double its value (c = 0.4 in the main simulation). Greater displacement was expected to result in less consistent sequences and therefore decreased map precision and reduced performance. The results shown in Table 1 confirmed that intuition. Doubling displacement (i.e., c = 0.8) decreased the advantage of replay and halving that error (i.e., s=0.2) did not show consistent performance improvement.

Suppl. Figure 1 shows the maps of sample learners in several conditions. These maps confirm the model robustness and suggest that lowering replays (to n=5) and increasing the displacement coefficient (c=0.8) results in less consistent maps and greater categorization error, which reduces accuracy in item- and map-inference. For example, the sample learners in the condition of large displacement coefficient (s=0.8, bottom row) have the same maps for maze 3 and 4, which is an example of categorization error: the same cluster has become preferred for two very similar mazes.

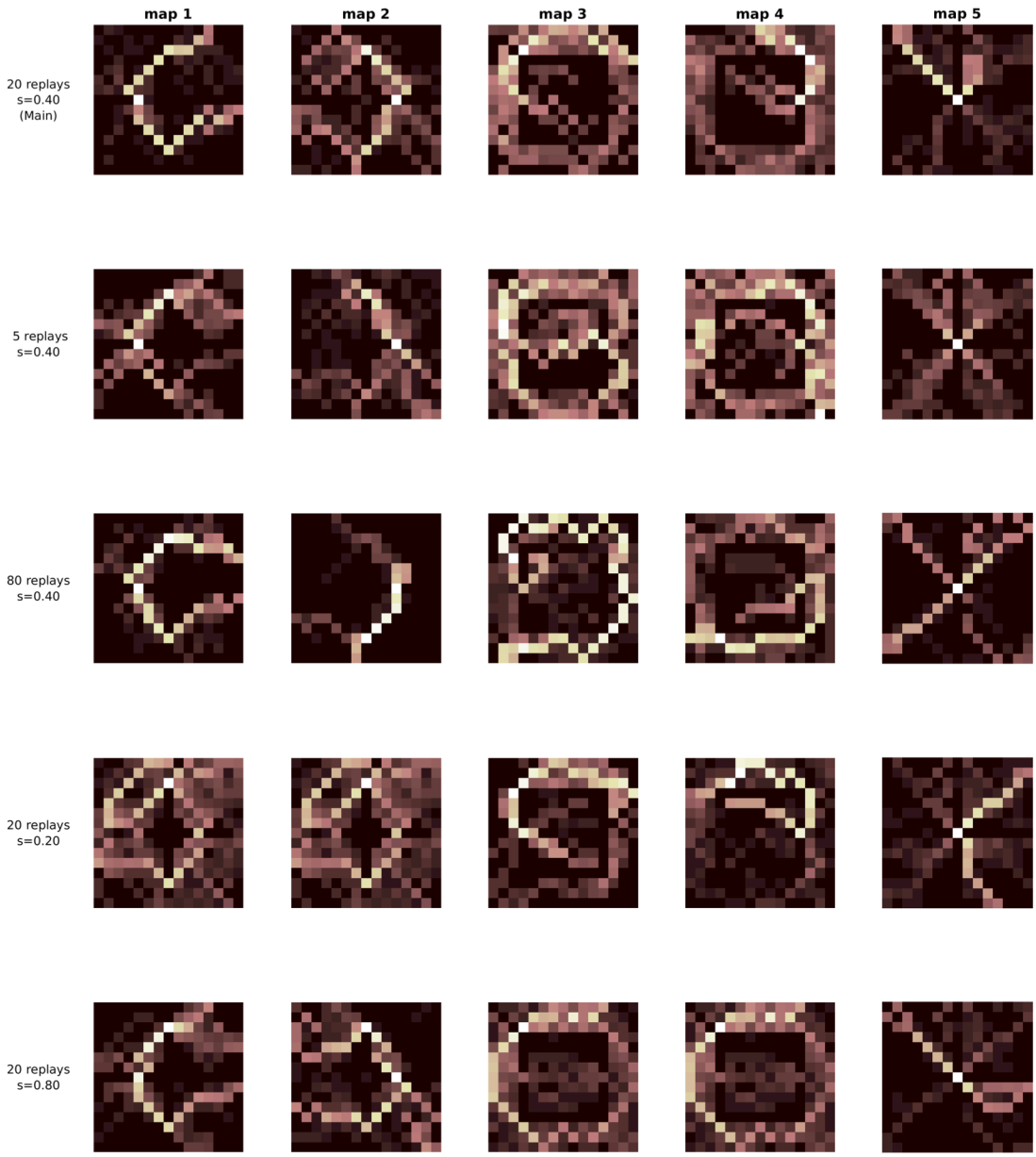
Supplementary Video

We have provided a supplementary video exemplifying the functioning of the hierarchical generative model discussed in this article. The video shows the model inputs (bottom panel) and the activity of the latent units at each hierarchical level, during the first three and the last three trials of each of the five learning blocks. For illustrative purposes, the video shows the dynamics of generative replays after each trial (labeled as "awake replays"), but these are not used for learning. Furthermore, the video shows the dynamics of generative replays after each block (labeled as "replays during sleep"), which are used for learning. The video illustrates that the hierarchical generative model learns accurate maps very rapidly, which allows accurate inference of which maze the agent is currently in. Furthermore, the generative replay mechanism generates both sequences that were actually experienced and novel sequences (sometimes including implausible sequences, as expected).

Suppl. Table 1. Control simulations. Recognition error measured during navigation (learning) and retention test. The gray rows are two replicas of the main learning condition (20 replays, s=0.4). There are n=32 replicas per condition.

	LEARNING											
		Baseline			**Exp.Replay**			**Gen.Replay**			**Prior.Gen.Replay**		
replays	**5**	0.16	(se	0.02)	0.14	(se	0.02)	0.11	(se	0.02)	0.12	(se	0.01)
	10	0.14	(se	0.02)	0.12	(se	0.02)	0.11	(se	0.02)	0.11	(se	0.02)
	20	0.16	(se	0.02)	0.10	(se	0.02)	0.07	(se	0.01)	0.05	(se	0.02)
	40	0.17	(se	0.02)	0.12	(se	0.02)	0.08	(se	0.01)	0.03	(se	0.01)
	80	0.16	(se	0.02)	0.12	(se	0.02)	0.08	(se	0.02)	0.05	(se	0.02)
c	**0,2**	0.14	(se	0.02)	0.13	(se	0.02)	0.05	(se	0.02)	0.07	(se	0.02)
	0,4	0.16	(se	0.02)	0.10	(se	0.02)	0.07	(se	0.01)	0.05	(se	0.02)
	0,8	0.16	(se	0.03)	0.12	(se	0.02)	0.08	(se	0.02)	0.15	(se	0.03)

	RETENTION											
		Baseline			**Exp.Replay**			**Gen.Replay**			**Prior.Gen.Replay**		
replays	**5**	0.21	(se	0.02)	0.20	(se	0.02)	0.17	(se	0.02)	0.18	(se	0.01)
	10	0.19	(se	0.02)	0.19	(se	0.02)	0.17	(se	0.02)	0.17	(se	0.02)
	20	0.22	(se	0.02)	0.18	(se	0.02)	0.14	(se	0.01)	0.12	(se	0.01)
	40	0.22	(se	0.01)	0.21	(se	0.01)	0.15	(se	0.01)	0.11	(se	0.01)
	80	0.21	(se	0.02)	0.22	(se	0.02)	0.17	(se	0.02)	0.14	(se	0.02)
c	**0,2**	0.20	(se	0.02)	0.21	(se	0.02)	0.13	(se	0.02)	0.13	(se	0.02)
	0,4	0.22	(se	0.02)	0.18	(se	0.02)	0.14	(se	0.01)	0.12	(se	0.01)
	0,8	0.22	(se	0.02)	0.20	(se	0.02)	0.14	(se	0.02)	0.20	(se	0.02)



Suppl. Figure 1. Control simulations, maps of sample learners. See the main text for explanation.

References

1. McCloskey, M., and Cohen, N.J. (1989). Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. In Psychology of Learning and Motivation, G. H. Bower, ed. (Academic Press), pp. 109–165.
2. McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. *102*, 419–457.
3. O'Reilly, R.C., Bhattacharyya, R., Howard, M.D., and Ketz, N. (2014). Complementary learning systems. Cogn. Sci. *38*, 1229–1248.
4. Buzsáki, G. (2015). Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. Hippocampus *25*, 1073–1188.
5. Foster, D.J. (2017). Replay comes of age. Annu. Rev. Neurosci. *40*, 581–602.
6. Foster, D.J., and Wilson, M.A. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. Nature *440*, 680–683.
7. Jadhav, S.P., Kemere, C., German, P.W., and Frank, L.M. (2012). Awake Hippocampal Sharp-Wave Ripples Support Spatial Memory. Science *336*, 1454–1458.
8. Kumaran, D., Hassabis, D., and McClelland, J.L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. Trends Cogn. Sci. *20*, 512–534.
9. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. Nature *518*, 529–533.
10. Pfeiffer, B.E., and Foster, D.J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. Nature *497*, 74–79.
11. Stella, F., Baracskay, P., O'Neill, J., and Csicsvari, J. (2019). Hippocampal Reactivation of Random Trajectories Resembling Brownian Diffusion. Neuron *102*, 450-461.e7.
12. Liu, K., Sibille, J., and Dragoi, G. (2018). Generative Predictive Codes by Multiplexed Hippocampal Neuronal Tuplets. Neuron *99*, 1329-1341.e6.
13. Dragoi, G., and Tonegawa, S. (2013). Distinct preplay of multiple novel spatial experiences in the rat. Proc. Natl. Acad. Sci. *110*, 9100–9105.
14. Gupta, A.S., van der Meer, M.A.A., Touretzky, D.S., and Redish, A.D. (2010). Hippocampal replay is not a simple function of experience. Neuron *65*, 695–705.
15. Ólafsdóttir, H.F., Barry, C., Saleem, A.B., Hassabis, D., and Spiers, H.J. (2015). Hippocampal place cells construct reward related sequences through unexplored space. eLife *4*, e06063.
16. Dragoi, G., and Tonegawa, S. (2011). Preplay of future place cell sequences by hippocampal cellular assemblies. Nature *469*, 397–401.
17. Liu, Y., Dolan, R.J., Kurth-Nelson, Z., and Behrens, T.E.J. (2019). Human Replay Spontaneously Reorganizes Experience. Cell.
18. Pezzulo, G., Kemere, C., and Meer, M. van der (2017). Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition. Ann. N. Y. Acad. Sci. *1396*, 144–165.
19. Penagos, H., Varela, C., and Wilson, M.A. (2017). Oscillations, neural computations and learning during wake and sleep. Curr. Opin. Neurobiol. *44*, 193–201.
20. Pezzulo, G., Donnarumma, F., Maisto, D., and Stoianov, I. (2019). Planning at decision time and in the background during spatial navigation. Curr. Opin. Behav. Sci. *29*, 69–76.
21. Buzsáki, G., and Tingley, D. (2018). Space and Time: The Hippocampus as a Sequence

Generator. Trends Cogn. Sci. *22*, 853–869.
22. Buzsáki, G., and Moser, E.I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. Nat. Neurosci. *16*, 130–138.
23. Yonelinas, A.P., Ranganath, C., Ekstrom, A.D., and Wiltgen, B.J. (2019). A contextual binding theory of episodic memory: systems consolidation reconsidered. Nat. Rev. Neurosci. *20*, 364–375.
24. Tulving, E. (2002). Episodic Memory: From Mind to Brain. Annu. Rev. Psychol. *53*, 1–25.
25. Sanders, H., Wilson, M.A., and Gershman, S.J. (2020). Hippocampal remapping as hidden state inference. eLife *9*, e51140.
26. Gershman, S.J., Blei, D.M., and Niv, Y. (2010). Context, learning, and extinction. Psychol. Rev. *117*, 197–209.
27. Collins, A.G., and Frank, M.J. (2013). Cognitive control over learning: creating, clustering, and generalizing task-set structure. Psychol. Rev. *120*, 190.
28. Stoianov, I., Genovesio, A., and Pezzulo, G. (2015). Prefrontal Goal Codes Emerge as Latent States in Probabilistic Value Learning. J. Cogn. Neurosci.
29. Rikhye, R.V., Gothoskar, N., Guntupalli, J.S., Dedieu, A., Lázaro-Gredilla, M., and George, D. (2020). Learning cognitive maps as structured graphs for vicarious evaluation. bioRxiv, 864421.
30. Heald, J.B., Lengyel, M., and Wolpert, D.M. (2020). Contextual inference underlies the learning of sensorimotor repertoires.
31. Mocanu, D.C., Vega, M.T., Eaton, E., Stone, P., and Liotta, A. (2016). Online Contrastive Divergence with Generative Replay: Experience Replay without Storing Data. ArXiv161005555 Cs.
32. Shin, H., Lee, J.K., Kim, J., and Kim, J. (2017). Continual Learning with Deep Generative Replay. ArXiv170508690 Cs.
33. van de Ven, G.M., Siegelmann, H.T., and Tolias, A.S. (2020). Brain-inspired replay for continual learning with artificial neural networks. Nat. Commun. *11*, 4069.
34. Stoianov, I.P., Pennartz, C.M.A., Lansink, C.S., and Pezzulo, G. (2018). Model-based spatial navigation in the hippocampus-ventral striatum circuit: A computational analysis. PLOS Comput. Biol. *14*, e1006316.
35. Rao, R.P., and Ballard, D.H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci *2*, 79–87.
36. Friston, K.J. (2005). A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci *360*, 815–836.
37. Dabaghian, Y., Brandt, V.L., and Frank, L.M. (2014). Reconceiving the hippocampal map as a topological template. eLife *3*, e03476.
38. Donnarumma, F., Maisto, D., and Pezzulo, G. (2016). Problem Solving as Probabilistic Inference with Subgoaling: Explaining Human Successes and Pitfalls in the Tower of Hanoi. PLOS Comput. Biol. *12*, e1004864.
39. Maisto, D., Donnarumma, F., and Pezzulo, G. (2016). Nonparametric Problem-Space Clustering: Learning Efficient Codes for Cognitive Control Tasks. Entropy *18*, 61.
40. Maisto, D., Donnarumma, F., and Pezzulo, G. (2015). Divide et impera: subgoaling reduces the complexity of probabilistic inference and problem solving. J. R. Soc. Interface *12*, 20141335.
41. Gershman, S.J. (2018). The Successor Representation: Its Computational Logic and Neural Substrates. J. Neurosci. *38*, 7193–7200.
42. Aoki, Y., Igata, H., Ikegaya, Y., and Sasaki, T. (2019). The Integration of Goal-Directed Signals onto Spatial Maps of Hippocampal Place Cells. Cell Rep. *27*, 1516-1527.e5.
43. Bhattarai, B., Lee, J.W., and Jung, M.W. (2020). Distinct effects of reward and navigation history on hippocampal forward and reverse replays. Proc. Natl. Acad. Sci. *117*, 689–697.

44. Whittington, J.C.R., Muller, T.H., Mark, S., Chen, G., Barry, C., Burgess, N., and Behrens, T.E.J. (2020). The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. Cell *183*, 1249-1263.e23.
45. Stachenfeld, K.L., Botvinick, M.M., and Gershman, S.J. (2017). The hippocampus as a predictive map. Nat. Neurosci. *20*, 1643–1653.
46. Moore, A.W., and Atkeson, C.G. (1993). Prioritized sweeping: Reinforcement learning with less data and less time. Mach. Learn. *13*, 103–130.
47. Fuhs, M.C., and Touretzky, D.S. (2007). Context learning in the rodent hippocampus. Neural Comput. *19*, 3173–3215.
48. Penny, W.D., Zeidman, P., and Burgess, N. (2013). Forward and Backward Inference in Spatial Cognition. PLoS Comput Biol *9*, e1003383.
49. Rikhye, R.V., Guntupalli, J.S., Gothoskar, N., Lázaro-Gredilla, M., and George, D. (2019). Memorize-Generalize: An online algorithm for learning higher-order sequential structure with cloned Hidden Markov Models. bioRxiv, 764456.
50. Ambrose, R.E., Pfeiffer, B.E., and Foster, D.J. (2016). Reverse Replay of Hippocampal Place Cells Is Uniquely Modulated by Changing Reward. Neuron *91*, 1124–1136.
51. Mattar, M.G., and Daw, N.D. (2017). Prioritized memory access explains planning and hippocampal replay. bioRxiv, 225664.
52. Foster, D.J., and Wilson, M.A. (2007). Hippocampal theta sequences. Hippocampus *17*, 1093–1099.
53. Johnson, A., and Redish, A.D. (2007). Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. J Neurosci *27*, 12176–12189.
54. Redish, A.D. (2016). Vicarious trial and error. Nat. Rev. Neurosci. *17*, 147–159.
55. Lisman, J., and Redish, A.D. (2009). Prediction, Sequences and the Hippocampus. Philos. Trans. R. Soc. B Biol. Sci. *364*, 1193–1201.
56. Sanders, H., Rennó-Costa, C., Idiart, M., and Lisman, J. (2015). Grid Cells and Place Cells: An Integrated View of their Navigational and Memory Function. Trends Neurosci. *38*, 763–775.
57. Shimazaki, H. (2018). Neurons as an Information-theoretic Engine. ArXiv151207855 Phys. Q-Bio.
58. Wikenheiser, A.M., and Redish, A.D. (2015). Hippocampal theta sequences reflect current goals. Nat. Neurosci. *18*, 289–294.
59. Liu, K., Sibille, J., and Dragoi, G. (2019). Preconfigured patterns are the primary driver of offline multi-neuronal sequence replay. Hippocampus *29*, 275–283.
60. Dragoi, G. (2020). Cell assemblies, sequences and temporal coding in the hippocampus. Curr. Opin. Neurobiol. *64*, 111–118.
61. Whittington, J., Muller, T., Mark, S., Barry, C., and Behrens, T. (2018). Generalisation of structural knowledge in the hippocampal-entorhinal system. In Advances in Neural Information Processing Systems 31, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds. (Curran Associates, Inc.), pp. 8484–8495.
62. Boccara, C.N., Nardin, M., Stella, F., O'Neill, J., and Csicsvari, J. (2019). The entorhinal cognitive map is attracted to goals. Science *363*, 1443–1447.
63. Butler, W.N., Hardcastle, K., and Giocomo, L.M. (2019). Remembered reward locations restructure entorhinal spatial maps. Science *363*, 1447–1452.
64. Plitt, M.H., and Giocomo, L.M. (2019). Experience dependent contextual codes in the hippocampus. bioRxiv, 864090.
65. Aronov, D., Nevers, R., and Tank, D.W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. Nature *543*, 719–722.
66. Constantinescu, A.O., O'Reilly, J.X., and Behrens, T.E.J. (2016). Organizing conceptual knowledge in humans with a gridlike code. Science *352*, 1464–1468.
67. Mok, R.M., and Love, B.C. (2019). A non-spatial account of place and grid cells based on

clustering models of concept learning. Nat. Commun. *10*, 1–9.
68. Gulli, R.A., Duong, L.R., Corrigan, B.W., Doucet, G., Williams, S., Fusi, S., and Martinez-Trujillo, J.C. (2020). Context-dependent representations of objects and space in the primate hippocampus during virtual navigation. Nat. Neurosci. *23*, 103–112.
69. Duvelle, É., Grieves, R.M., Liu, A., Jedidi-Ayoub, S., Holeniewska, J., Harris, A., Nyberg, N., Donnarumma, F., Lefort, J.M., Jeffery, K.J., et al. (2021). Hippocampal place cells encode global location but not connectivity in a complex space. Curr. Biol. *31*, 1221-1233.e9.
70. Colgin, L.L., Moser, E.I., and Moser, M.-B. (2008). Understanding memory through hippocampal remapping. Trends Neurosci. *31*, 469–477.
71. Jezek, K., Henriksen, E.J., Treves, A., Moser, E.I., and Moser, M.-B. (2011). Theta-paced flickering between place-cell maps in the hippocampus. Nature *478*, 246–249.
72. Kay, K., Chung, J.E., Sosa, M., Schor, J.S., Karlsson, M.P., Larkin, M.C., Liu, D.F., and Frank, L.M. (2020). Constant Sub-second Cycling between Representations of Possible Futures in the Hippocampus. Cell *180*, 552-567.e25.
73. Grieves, R.M., Wood, E.R., and Dudchenko, P.A. (2016). Place cells on a maze encode routes rather than destinations. eLife *5*.
74. Spiers, H.J., Hayman, R.M.A., Jovalekic, A., Marozzi, E., and Jeffery, K.J. (2015). Place field repetition and purely local remapping in a multicompartment environment. Cereb. Cortex N. Y. N 1991 *25*, 10–25.
75. Grieves, R.M., Jenkins, B.W., Harland, B.C., Wood, E.R., and Dudchenko, P.A. (2016). Place field repetition and spatial learning in a multicompartment environment. Hippocampus *26*, 118–134.
76. Liu, Y., Kurth-Nelson, Z., and Behrens, T. (2018). Non-spatial neural replay in building and updating world models in humans. In Cosyne 2018.
77. Chenani, A., Sabariego, M., Schlesiger, M.I., Leutgeb, J.K., Leutgeb, S., and Leibold, C. (2019). Hippocampal CA1 replay becomes less prominent but more rigid without inputs from medial entorhinal cortex. Nat. Commun. *10*, 1341.
78. Shin, J.D., Tang, W., and Jadhav, S.P. (2019). Dynamics of Awake Hippocampal-Prefrontal Replay for Spatial Learning and Memory-Guided Decision Making. Neuron *104*, 1110-1125.e7.
79. Tang, W., Shin, J.D., and Jadhav, S.P. (2020). Multiple time-scales of decision making in the hippocampus and prefrontal cortex. bioRxiv, 2020.10.17.343699.
80. Lansink, C.S., Goltstein, P.M., Lankelma, J.V., McNaughton, B.L., and Pennartz, C.M. (2009). Hippocampus leads ventral striatum in replay of place-reward information. PLoS Biol. *7*, e1000173.
81. Pezzulo, G., van der Meer, M.A.A., Lansink, C.S., and Pennartz, C.M.A. (2014). Internally generated sequences in learning and executing goal-directed behavior. Trends Cogn. Sci. *18*, 647–657.
82. van der Meer, M., Kurth-Nelson, Z., and Redish, A.D. (2012). Information processing in decision-making systems. The Neuroscientist *18*, 342–359.
83. Tolman, E.C. (1948). Cognitive maps in rats and men. Psychol. Rev. *55*, 189–208.
84. Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. Nat. Rev. Neurosci. *1*, 41–50.
85. Edelman, G.M. (1987). Neural Darwinism: The Theory of Neuronal Group Selection. (Basic Books).
86. Lukoševičius, M., and Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. Comput. Sci. Rev. *3*, 127–149.
87. Jaeger, H., Maass, W., and Principe, J. (2007). Special issue on echo state networks and liquid state machines.
88. Dominey, P., Arbib, M., and Joseph, J.P. (1995). A model of corticostriatal plasticity for

learning oculomotor associations and sequences. J. Cogn. Neurosci. *7*, 311–336.
89. Frankle, J., and Carbin, M. (2019). The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks. ArXiv180303635 Cs.
90. Shin, H., Lee, J.K., Kim, J., and Kim, J. (2017). Continual Learning with Deep Generative Replay. ArXiv170508690 Cs.
91. Recanatesi, S., Farrell, M., Lajoie, G., Deneve, S., Rigotti, M., and Shea-Brown, E. (2018). Signatures and mechanisms of low-dimensional neural predictive manifolds. bioRxiv, 471987.
92. Atherton, L.A., Dupret, D., and Mellor, J.R. (2015). Memory trace replay: the shaping of memory consolidation by neuromodulation. Trends Neurosci. *38*, 560–570.
93. Pezzulo, G., Rigoli, F., and Chersi, F. (2012). A mixed instrumental controller can combine habitual and goal-directed choice. In Proceedings of CogSci 2012.
94. Pezzulo, G., Zorzi, M., and Corbetta, M. (2021). The secret life of predictive brains: what's spontaneous activity for? Trends Cogn. Sci. *25*, 730–743.
95. Wittkuhn, L., Chien, S., Hall-McMaster, S., and Schuck, N.W. (2021). Replay in minds and machines. Neurosci. Biobehav. Rev. *129*, 367–388.
96. Cazin, N., Alonso, M.L., Chiodi, P.S., Pelc, T., Harland, B., Weitzenfeld, A., Fellous, J.-M., and Dominey, P.F. (2019). Reservoir computing model of prefrontal cortex creates novel combinations of previous navigation sequences from hippocampal place-cell replay with spatial reward propagation. PLOS Comput. Biol. *15*, e1006624.
97. Sutton, R.S. (1991). Integrated modeling and control based on reinforcement learning and dynamic programming. In Advances in Neural Information Processing Systems, D. S. Touretzky, ed., pp. 471–478.
98. Ha, D., and Schmidhuber, J. (2018). World Models. ArXiv180310122 Cs Stat.
99. Terada, S., Sakurai, Y., Nakahara, H., and Fujisawa, S. (2017). Temporal and Rate Coding for Discrete Event Sequences in the Hippocampus. Neuron *94*, 1248-1262.e4.
100. Miller, K.J., Botvinick, M.M., and Brody, C.D. (2017). Dorsal hippocampus contributes to model-based planning. Nat. Neurosci. *20*, 1269–1276.
101. Schuck, N.W., and Niv, Y. (2019). Sequential replay of nonspatial task states in the human hippocampus. Science *364*.
102. Shahbaba, B., Li, L., Agostinelli, F., Saraf, M., Elias, G.A., Baldi, P., and Fortin, N.J. (2019). Hippocampal ensembles represent sequential relationships among discrete nonspatial events. bioRxiv, 840199.
103. Behrens, T.E.J., Muller, T.H., Whittington, J.C.R., Mark, S., Baram, A.B., Stachenfeld, K.L., and Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. Neuron *100*, 490–509.
104. Bellmund, J.L.S., Gärdenfors, P., Moser, E.I., and Doeller, C.F. (2018). Navigating cognition: Spatial codes for human thinking. Science *362*, eaat6766.
105. Pezzulo, G., Donnarumma, F., Iodice, P., Maisto, D., and Stoianov, I. (2017). Model-Based Approaches to Active Perception and Control. Entropy *19*, 266.
106. Buckner, R.L. (2010). The role of the hippocampus in prediction and imagination. Annu Rev Psychol *61*, 27–48, C1-8.
107. Schacter, D.L., Addis, D.R., Hassabis, D., Martin, V.C., Spreng, R.N., and Szpunar, K.K. (2012). The Future of Memory: Remembering, Imagining, and the Brain. Neuron *76*, 677–694.
108. O'keefe, J., and Nadel, L. (1978). The hippocampus as a cognitive map (Clarendon Press Oxford).
109. Schacter, D.L., and Addis, D.R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. Philos Trans R Soc Lond B Biol Sci *362*, 773–786.
110. Schacter, D.L., Addis, D.R., and Buckner, R.L. (2007). Remembering the past to imagine the future: the prospective brain. Nat Rev Neurosci *8*, 657–661.

111. Szpunar, K.K. (2010). Episodic Future Thought: An Emerging Concept. Perspect. Psychol. Sci. *5*, 142–162.
112. Schacter, D.L., and Addis, D.R. (2007). The ghosts of past and future. Nature *445*, 27–27.
113. Addis, D.R., Cheng, T., Roberts, R.P., and Schacter, D.L. (2011). Hippocampal contributions to the episodic simulation of specific and general future events. Hippocampus *21*, 1045–1052.
114. Gershman, S.J., and Blei, D.M. (2012). A tutorial on Bayesian nonparametric models. J. Math. Psychol. *56*, 1–12.