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 to make this possible, the hippocampal generative model is endowed with strong inductive biases to pattern-separate individual items of experience (at the first hierarchical layer), organize them into sequences (at the second layer) and then cluster them into maps (at the 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. Our experiments show that the hierarchical model using generative replay is able to learn and retain efficiently multiple spatial navigation trajectories, organizing them into separate spatial maps. Furthermore, it reproduces flexible and prospective aspects of hippocampal dynamics that have been 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 the novel acquired knowledge from interfering with past memories, causing their *catastrophic forgetting* [1]. The hippocampus has long been considered important to solve 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, hence preventing catastrophic forgetting [2,3].

The idea that the hippocampus stores and reactivates episodic memories offline received empirical support from rodent studies during spatial navigation. In phases of wakeful rest during spatial navigation and during 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 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, too. 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 such as 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 all its learning episodes in a memory buffer; and then it replays all them from the buffer when it learns a novel task C, preventing their catastrophic forgetting – similar to how hippocampal “replay” putatively works [8,9].

While the view that the hippocampus stores episodes in a memory buffer and "replays" them is consolidated 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 “replay” the order in which they were actually experienced [17].

In sum, internally generated hippocampal sequences during sleep or wakeful rest go beyond previous experiences and have prospective aspects that are difficult to reconcile with the idea of a “replay” of experiences from a memory buffer [5,18–21]; but an alternative theoretical framework has still 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., rodent trajectories during spatial navigation) 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 offline resampling of fictive experiences from the generative model, rather than from the verbatim replay of memories from a buffer.

The new theory is based on two main assumptions. The first assumption is that the hippocampal formation encodes experiences by *learning a hierarchical generative model* of 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) 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), hence supporting their sequential processing. 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). In other words, the hierarchical structure provides the model with *inductive biases* to disentangle *items* of experience and organize them into *sequences* (temporally ordered items) and *maps* (spatially clustered sequences).

The second assumption is that internally generated hippocampal sequences are manifestations of the inferential dynamics of the generative model, not "replays" from a memory buffer. 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 or to train separate models or controllers. 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 [22–24]. Furthermore, it is appealing from a biological perspective, as it explains flexible and prospective aspects of internally generated hippocampal sequences, which are difficult to reconcile with previous theories.

In the following section, we present a series of simulations that illustrate the computational efficacy ("Continual learning of multiple spatial maps with generative replay") and biological relevance ("Biological interpretation of generative replays and supporting evidence") of the novel hypothesis.

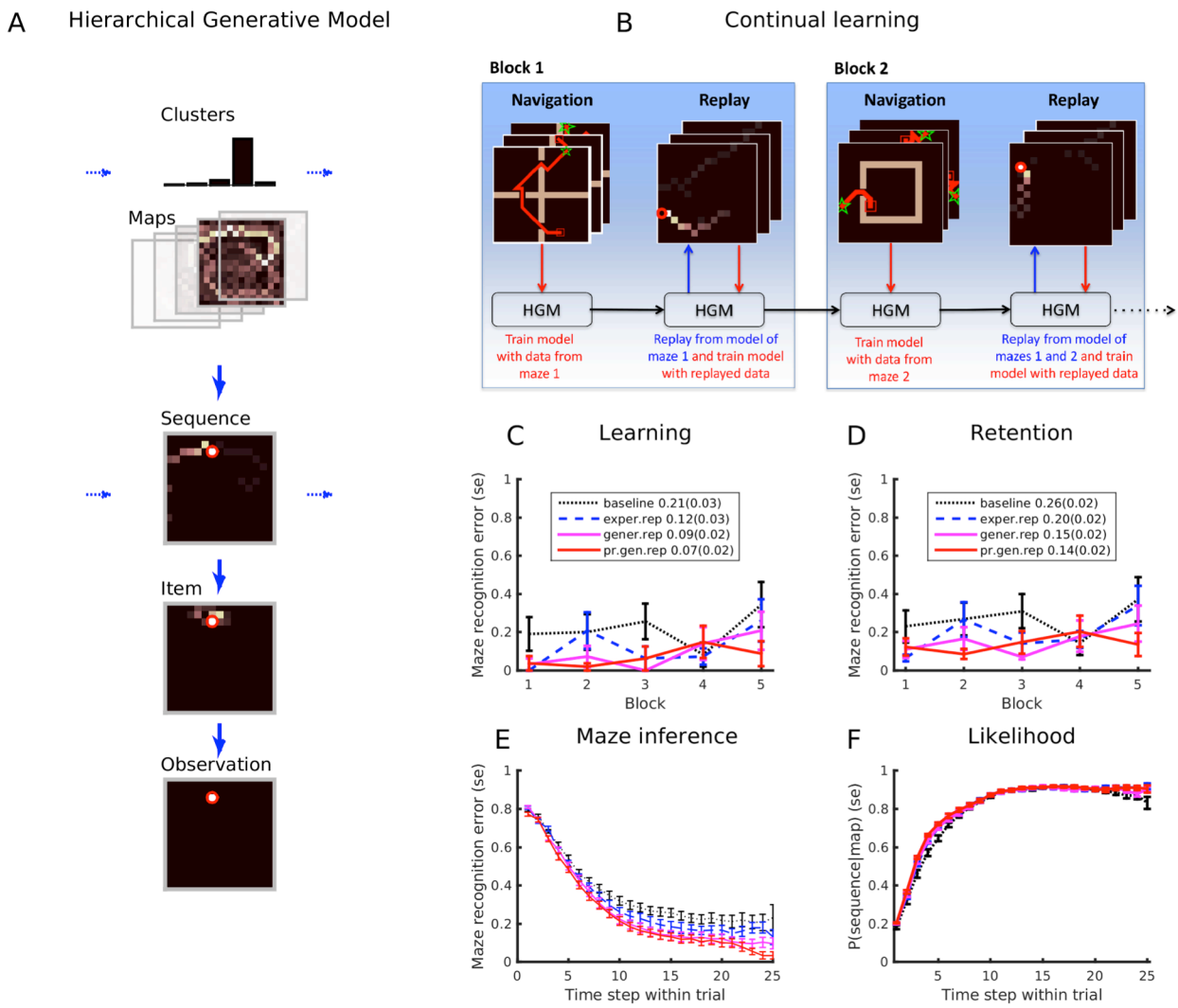

Figure 1. Hippocampal generative model. A: Components of the hierarchical generative model, from top to bottom: clusters / maps, sequences, items and observations. The red circle denotes the agent's current position. Colour codes correspond to probabilities, from high (white) to low (black). See the 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 applies 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 respective generative models. 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 learning agents (i.e., inference of the correct maze), measured during each block, after training on 5 trajectories of each block. D: Same as C, but measured at the end of the five learning blocks, on 20 novel trials per block. (E,F): Dynamics of inference. E: Same as D, but as a function of time step within trials (25 time steps are shown). 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 in 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".

The continual learning task that the four learning agents face consists in learning five 15x15 mazes, using a single generative model, see Figure 1B. The agents learn the five mazes in different blocks (mazes were selected randomly without repetitions for each agent and block). During each block, the four learning agents receive as data the same 20 spatial trajectories from the maze they are currently in and use these data to update their hierarchical generative models. The trajectory data are generated by a separate model-based Bayesian controller that, for each trial and maze, started from a random location and reached a random goal location, by following a near-optimal trajectory [25].

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 the learning agents are endowed with the same hierarchical generative model, shown in Figure 1A (see the Methods section for details). The generative model comprises three layers of hidden states, coding for *items*, *sequences* and *maps*, respectively; and a layer coding the current agent *observation*. We start describing the model from the bottom-up, i.e., from 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 trajectory data, but just one data point at the time.

The first hidden layer of the hierarchical model encodes *item* codes, putatively corresponding to place cells. Item codes are inferred probabilistically, using a scheme analogous to predictive coding [26,27], 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.

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 previous sequence of items, bottom-up information from the last inferred item, and the structure of the currently active (inferred) map.

The third hidden layer of the hierarchical model encodes *maps* 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 model maintains a (categorical) probability distribution over the clusters of the mixture model. The cluster having the highest probability corresponds to the currently inferred map, i.e., the map the agent believes it is in, based on its observed trajectory data. This probability distribution is continuously updated, in a Bayesian way, as the agent

observes new data. Only the currently inferred map is updated during learning and used to generate fictive experiences during generative replays (see later).

The "content" of each cluster (i.e., the map) is a probability distribution over an arbitrary structure. In our simulations of spatial navigation, the structure is a 2D grid of locations, coherent with the view that the hippocampus encodes a space's topological information, and relations between locations, such as adjacency [28]. Hence, the maps used in this simulation automatically provide *structural priors* to organize experience, such as the fact that there are "no gaps" between adjacent locations and trajectories can proceed in both ways. Furthermore, maps do not simply represent items in spatial locations but also their *behavioural relevance* – in their associated probability distributions. These probability distributions are updated using visitation counts during both real and fictive experiences (see later) and come to reflect the expectation of visiting each location in the future - or empirical priors in Bayesian parlance (note that using other methods, such as algorithmic complexity [29–31] and successor representations [32], 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 [33]. The fact that maps are *biased* by behavioural relevance reflects the idea that some aspects of value are already incorporated in the hippocampal map and influence replay statistics [34] - which is an aspect that we will discuss below, in the section on "Biological interpretation of generative replays and supporting evidence".

In our simulations, the generative model uses five clusters. This is done to facilitate the analyses, as there is potentially a one-to-one mapping between the five clusters and the five mazes (and indeed our results will show that most simulations converge to this solution, or to a close approximation). However, we later 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 as data 20 spatial trajectories from the maze they are currently in (experiencing each trajectory just once in pseudorandom order) and use these data to train their generative models. 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 replays for training their generative models offline, between blocks: they replay 20 trajectories and use these fictive data to update their generative models, in exactly the same way they do with real trajectories. However, they use different replay mechanisms. The second agent uses *experience replay*: it replays trajectories randomly selected amongst all those experienced before. The third agent uses *generative replay*: it resamples trajectories from their inferred map, i.e., the one corresponding to the cluster having currently the highest probability (note that as the probabilities are continuously updated, successive replays can be generated by different maps). The fourth agent uses *prioritized generative replay*: it resamples trajectories from its generative model, but using *prioritized maps* 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 measures the difference from the (prior) probabilities encoded in the maps, for each location. Using prioritized maps permits resampling trajectories corresponding to the portions of the maze that needs more updating, similar to prioritized sweeping in

reinforcement learning [35] and to empirical observations that sequential activations that were not present before a navigation episode were replayed more afterwards [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., in which maze they are) – which in Bayesian terms, is a form of *latent state (map)* inference [36–38]. For this, we first establish, for each agent, what is the preferred (i.e., most likely) maze for each cluster at the end of learning (note that in almost all cases, one cluster is much more likely than the others, see below). We consider this preferred cluster to be "ground truth", or the cluster that the agent would select if it had enough data. Then we consider, for each step, how many times the agent assigns the highest probability to the preferred cluster, after the first 10 observations of each trial (but we exclude the first 5 trials of each block from this analysis). Our results show that the two agents using generative replay and prioritized generative replay outperform the baseline and the experience replay agents: they infer more often the correct map, given their input trajectory data.

Figure 1D shows the same learning error, but computed during a separate (retention) test session executed after the agents experienced all the 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 particularly challenging, as the agents cannot accumulate evidence (about the maze they are in) across trials. Furthermore, learning novel mazes could lead to the catastrophic forgetting of previous mazes. Again, our results show that the two agents using generative replay and prioritized generative replay outperform the baseline and the experience replay agents, suggesting that the former are less prone to catastrophic forgetting.

Figures 1E and 1F characterize the dynamic effect of evidence accumulation *within* trials. They plot the performance of the agents during the test session against the first 25 time steps of each trial, during which they gradually accumulate information about both the maze and the trajectories they observe (recall that the agents receive one observation at a time and never observe entire trajectories). Our results show that as they accumulate more evidence within trials, all agents increase their capacity to correctly infer the maze they are in (Figure 1E) and their capacity to generate sequences from the correct map (Figure 1F). The latter is equivalent to reconstruction accuracy: a performance measure that is 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 the experience replay agents, in terms of both speed and accuracy of inference.

The reasons why generative replay and prioritized generative replay outperform the other two agents become apparent 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 the five mazes they experienced.

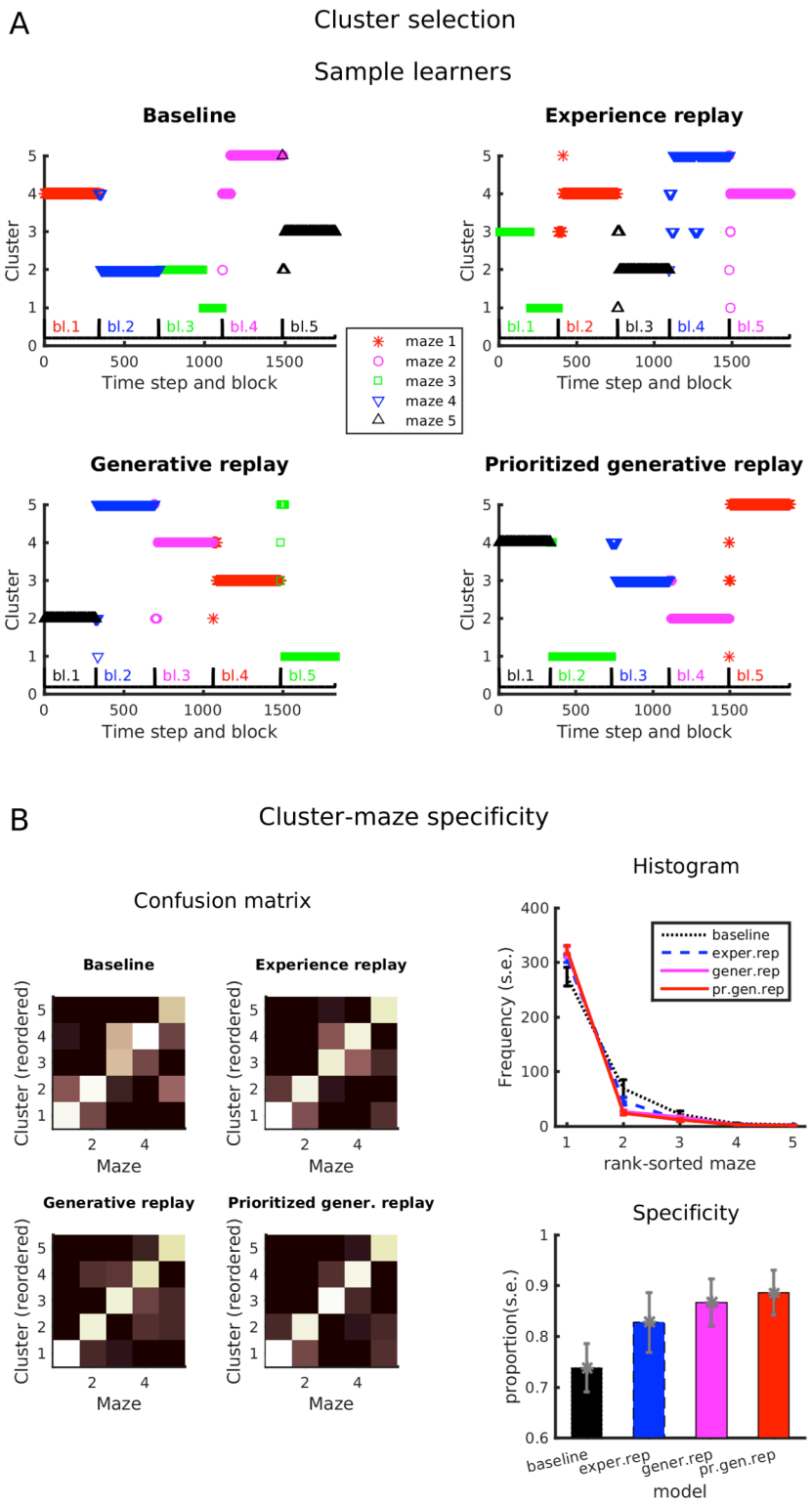

Figure 2. Cluster specificity during the continual learning experiment. (A) Cluster selection during navigation in sample replicas of each agent. The abscissa is double-labelled with time step (below) and colour coded blocks (above). The ordinate indexes the clusters (1 to 5) selected during the blocks. 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).

The cluster specificity of generative replay and prioritized generative replay agents can be further appreciated by looking at their (sample) learned maps, see Figure 3A. These maps encode spatial-topological information of 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 will discuss biological aspects of generative replay. It is also worth noting that generative replay and prioritized generative replay agents (but not baseline or experience replay agents) can generate novel trajectories during replay. Figure 3B shows an example novel trajectory generated by generative replay, superimposed with an actually experienced trajectory. The percentage of novel locations virtually explored during replays is quantified in Figure 3C. The possibility to generate novel trajectories that respect spatial-topological constraints of maps are key to explain the performance of generative replay and prioritized generative replay agents, compared to experience replay agents.

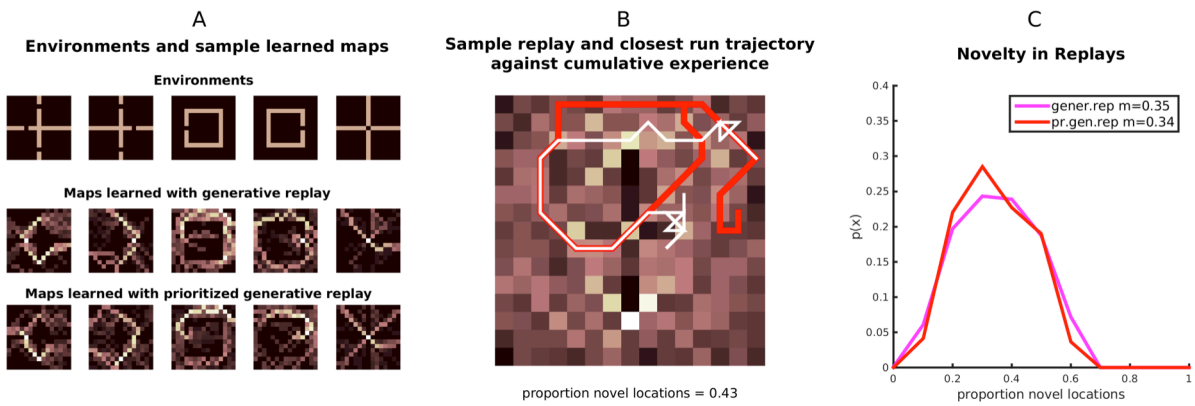

Figure 3. Examples of maps and replays of generative replay agents. (A) The five mazes used in the simulation (top) and some example maps learned by generative replay and prioritized generative replay agents. Colours represent probabilities, from white (high) to black (low). (B) Example novel trajectory 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.

Nonparametric model

So far, we discussed a generative model having a fixed number of clusters (five). However, we conducted the same continual learning experiment using 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. At difference with the previous simulations, here each maze was presented 4 times in different blocks, resulting in 20 navigation blocks. This is because nonparametric methods generally require more training data than parametric methods (as they need to infer the appropriate number of clusters). The performance of the nonparametric method approximates the parametric method: recognition error is 0.261 (se 0.029) for baseline; 0.148 (se 0.017) for experience replay; 0.097 (se 0.016) for generative replay, and 0.095 (se 0.019) for prioritized generative replay, respectively (n=16 replicas of each agent). As expected from a nonparametric method, all the agents show some redundancy in their learned clusters (Figure 4A) but they are still able to reliably map each maze to at least one cluster – as evident from the confusion matrices of Figure 4B. As for the parametric method, the two agents using generative replay and prioritized generative replay develop more specific clusters, which may explain their better performance. These

results illustrate that our approach to continual learning does not necessarily require a-priori information about the number of clusters, but can infer it from data.

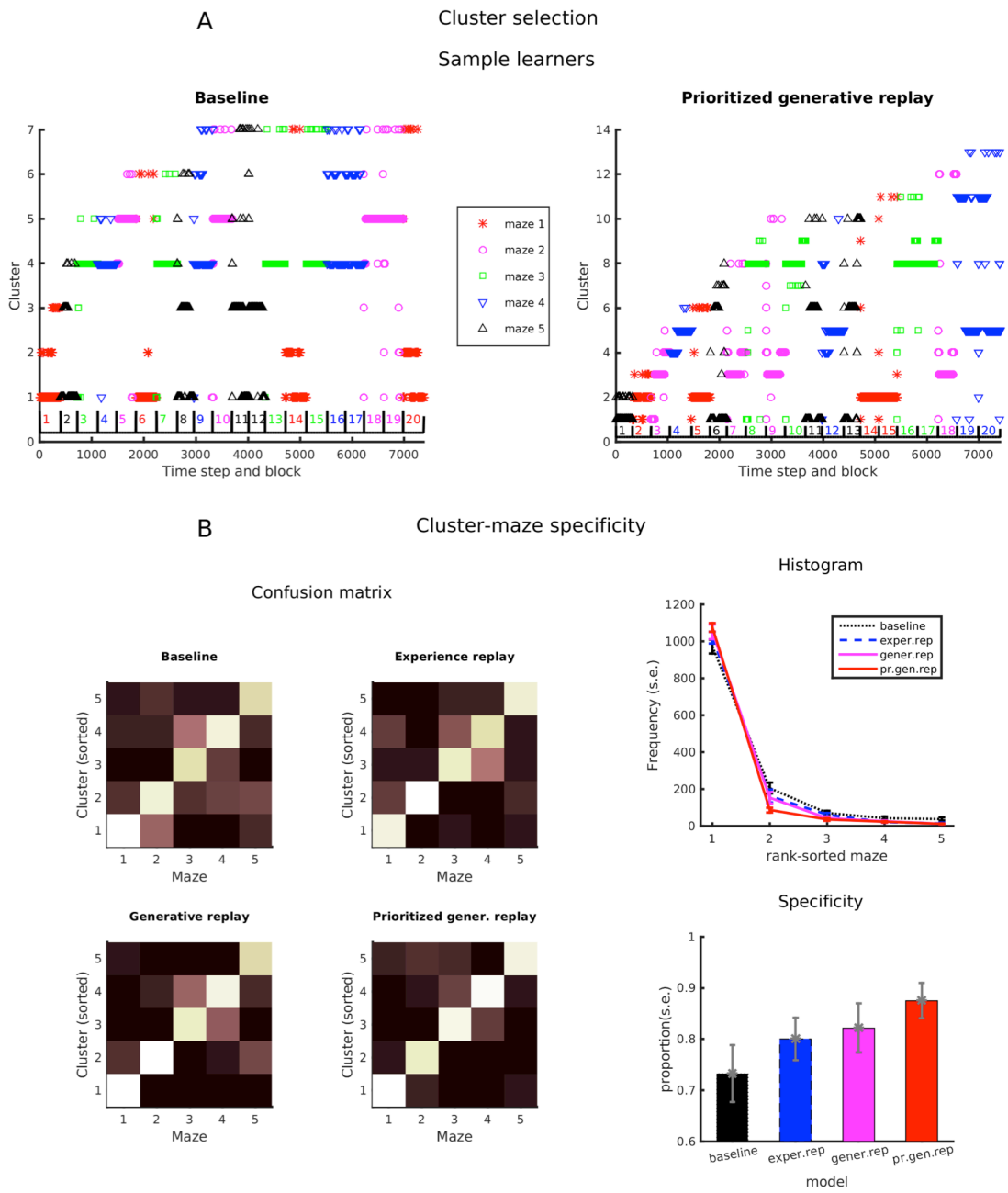

Figure 4. Clustering emerged during continuous nonparametric learning. (A) Cluster selection during navigation in sample baseline and prioritized generative replay agents. Note the greater maze-cluster consistency in the latter agent. (B) Analysis of the consistency of cluster-maze selection across all 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, 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.

Biological interpretation and supporting evidence for generative replay

We argued that generative replay provides a novel perspective on internally generated hippocampal sequences, suggesting that they are *sequences resampled from the latent map (prior) distribution*, rather than replays 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 replays during sleep and their sensitivity to goal and reward information

Generative replay functions both when stimuli are absent and when they are present, but differently. When externals stimuli are absent of weak (e.g., when the animal sleeps), generative replays produces sequences entirely composed of fictive observations; this mechanism is sometimes called *inputless decoder* in machine learning. The starting point of each generative replay is sampled from the currently inferred map, i.e., the map that currently has the highest probability. It is worth reminding that the maps used by the generative replay agent encode the *probability* of items at given locations (i.e., empirical priors in Bayesian terms), which reflect their behavioural relevance. The maps used by the prioritized generative replay agent encode instead *surprise* information; however, after goal-directed navigation, when new rewards are encountered (as in [39]), also surprise maps tend to peak at behaviourally relevant (e.g., goal) locations, where they diverge the most from prior probabilities. Hence, in the absence of external stimuli, generative replay events will tend to start from behaviourally relevant locations, such as goal locations, see Figure 5. This explains the finding that *hippocampal replays during sleep* tend to start form rewarded locations and proceed "backwards" to start positions during a trial [39]. Note that by selectively resampling sequences that encode goals, generative replay further biases the agent's maps to focus on goal and other behaviourally relevant locations (or items within them).

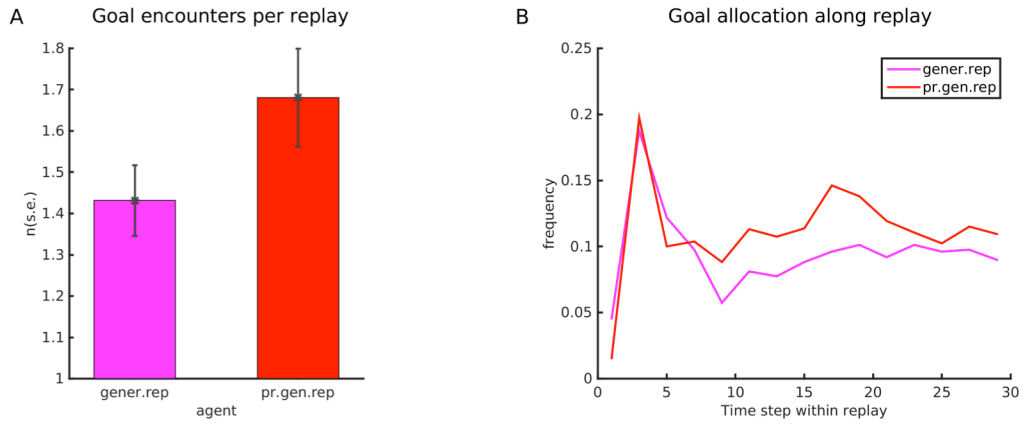

Figure 5. Analysis of goal-sensitivity during generative replays in the absence of external stimuli (akin to hippocampal replays during sleep). (A) Frequency of fictive visits to goal locations during generative replay. (B) Frequency of fictive visits to goal locations during the generative replays (lasting 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 explain replays during awake state and theta sequences

When an animal is engaged in a behavioural task, it receives sequences of (e.g., location) observations, which can provide a context and a starting point for generative replays. In this case, generative replays may support forward prediction: they may produce fictive observations (conditioned on the currently inferred map) that start from the currently inferred location and proceed along a path of high probability. Generative replay in the context of a behavioural task can potentially explain two important hippocampal phenomena.

First, it can explain *replays during the awake state*. At difference from replays during sleep, replays during the awake state often start around the animal's position and are biased towards goal locations [10]. This can be explained by considering that during awake state, the generative model may receive some weak observation (e.g., the animal's 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, such as goal locations. Hence, in this theory, the observed differences between replays during sleep and the awake state would only depend on the ways they are initialized: by sampling from a map or with 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 cells, observed during navigation within each hippocampal theta cycle [40]. Theta sequences have been implied in both memory encoding and the prediction of future spatial positions along the animal's current spatial trajectory [41–43]. 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 [44] for the related suggestion that these two processes may be realized in early and late phases of each theta cycle. 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 [45], 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 [46], which in our model would depend on the fact that the maps generating top-down predictions are biased.

To recap, in our model, replays during sleep, replays during awake state and theta sequences would all be manifestations of the dynamics of the same hierarchical generative model, but in different conditions. Replays during sleep are the "purest" form of (inputless) generative replay, whose starting point is sampled from the probabilistic maps. Note that during learning, the agent replays from multiple maps and not just the last experienced one (it selects a map for each replay, based on current cluster probabilities). This mechanism nicely explains why hippocampal replays can include distal experiences, as shown empirically [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 if the two maps were combined during replay).

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 [47]. 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 [48]. This is similar to what happens in our model, where preconfigured clusters can spontaneously produce repertoires of temporal sequences even before navigational experiences. The cluster that expresses the repertoire that fits better a specific maze tends to be selected (and later adapted) to represent such maze.

To test whether our model reproduces the above empirical findings, we performed the same rank-order analysis as in [47] on our generative replay and prioritized generative replay models. 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; and then correlated the rank maps calculated during post-run with pre-run, and during run with post-run.

To calculate the ranks, we used either the observed location (*observation* in Fig. 1) during run or the predicted locations during fictive (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. The ranks were normalized to 1 and the normalized ranks of all trajectories from a given episode were accumulated on a 2D rank-map,

which was therefore normalized to sum 1 (see Methods). Figure 6A shows the rank-maps of a sample learner 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 epoch. Figure 6A shows the rank-correlations between the sample rank-maps. Figure 6B shows statistical plots for each condition and episode across all the 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 sought using a linear regression of the correlations with block and condition as predictors, i.e. Rank-Correlation ~ condition + block.

Consistent with previous findings [47], 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 [47] 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.

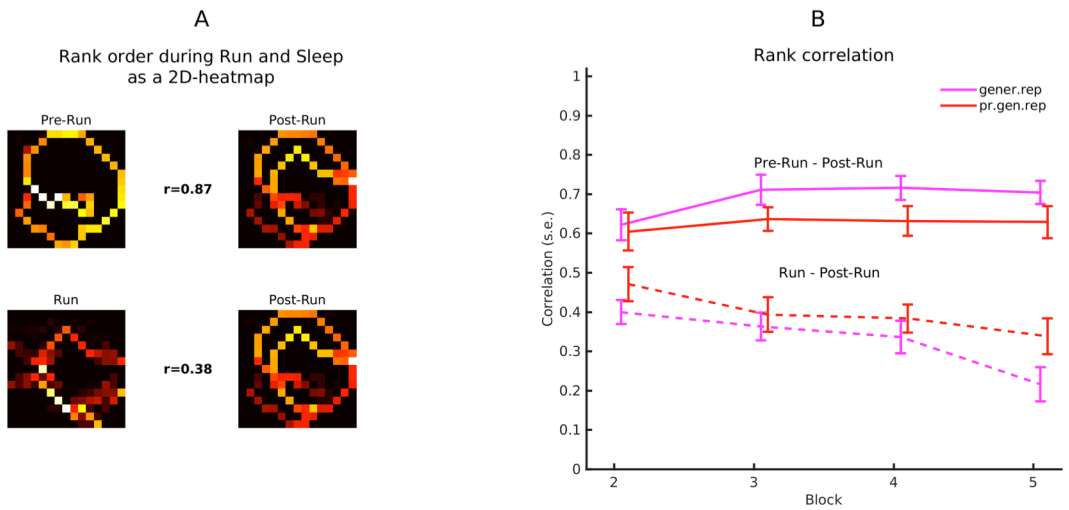

Figure 6. Network dynamics before novel navigation experiences explain replay activity better than novel experiences themselves. (A) Rank order maps of a sample learner during sample pre-run, run and post-run phases, along with rank correlations between pre-run and post-run, and run and post-run, respectively. Colours represent rank, from white (high) to black (zero). (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 epochs.

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 could 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-travelling time. In the animal data it was found that α was about 0.45-0.53 during replay (Figure 3A of [11]) and about 0.67-0.72 during navigation (Figure 4A of [11]) - which demonstrates that that the hippocampus can generate trajectories having different statistics from those experienced during navigation. This is similar to what happens in our model, where generative replay permits generating 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. As reported empirically, trajectories produced by our model during generative replay 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 reach 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 qualify 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 more uniform 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 graphical analysis showed that diffusion, as in the animal data, was much more directed during navigation than during replay.

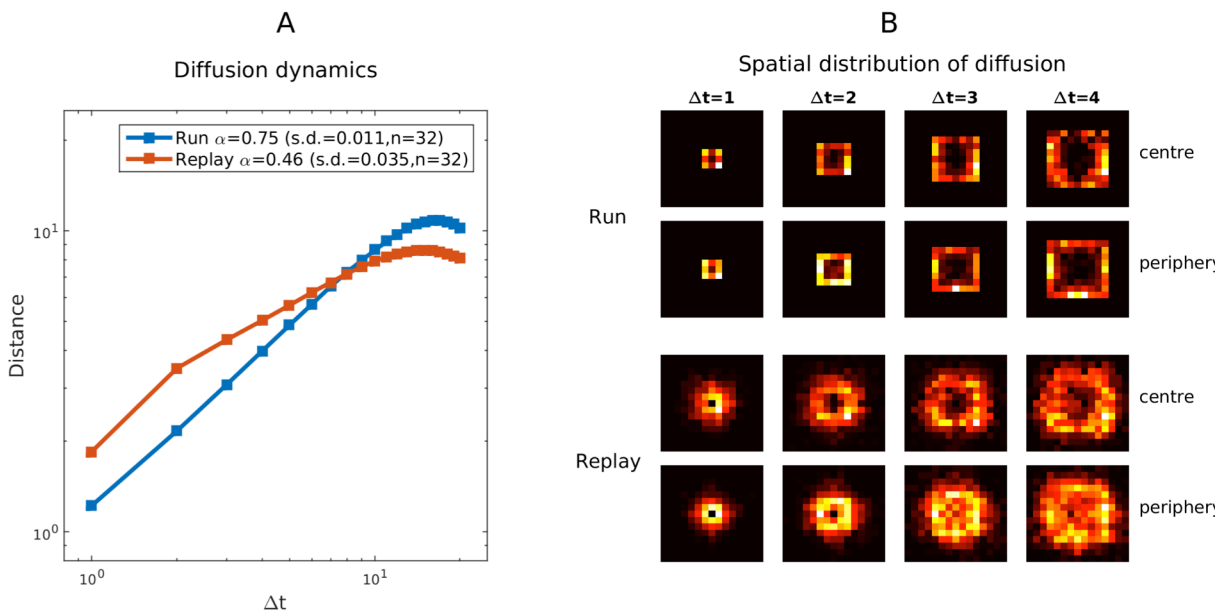

Figure 7. Trajectories during spatial experience and during replay across all learners and conditions. (A) Average point-to-point distance as a function of time-step intervals, plotted on a log-log scale. (B) Relative diffusion of locations during navigation and replay at five time steps for trajectories 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. 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 demand future investigations.

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 the hierarchical architecture.

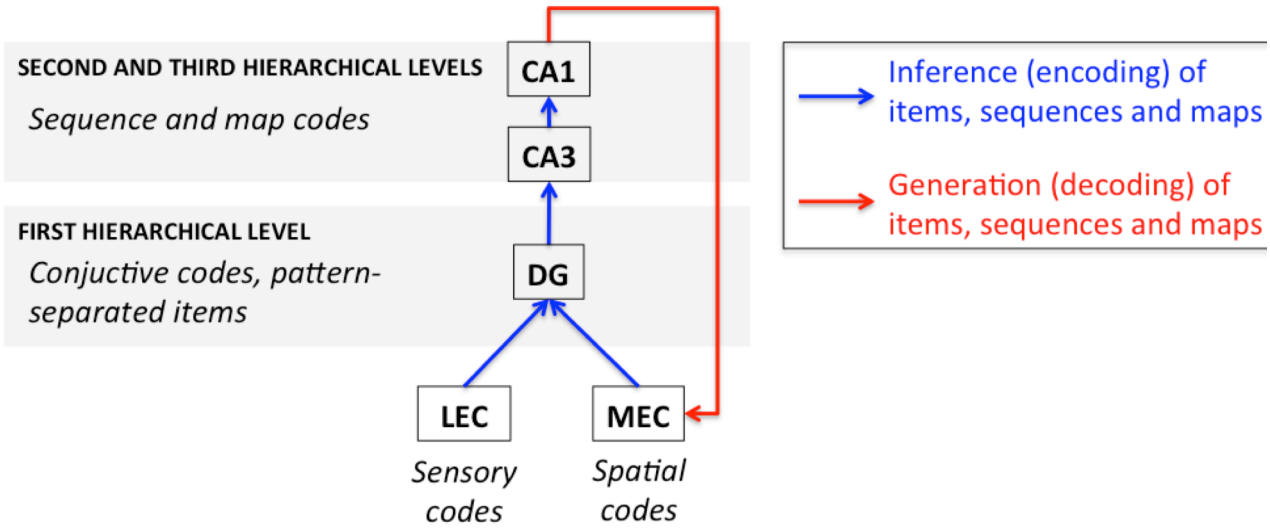

Figure 8. Putative neurobiological implementation of the hierarchical generative model (note that for simplicity, the figure only shows the EC-DG-CA3-CA1 system). LEC: lateral entorhinal cortex; MEC: medial entorhinal cortex; DG: dentate gyrus; CA: Cornu Ammonis

The maps 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 [49]. An alternative possibility is that the two different organizing principles encoded by levels two and three in our model – temporal and relational-spatial, respectively – 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 [50]. 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 [51,52].

Hippocampal coding of cognitive maps

Our model assumes that maps are structured representations that define a coherent context to organize stimuli, forming spatiotemporal codes for spatial navigation [53] but also possibly non-spatial codes for other tasks [54–57]. 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 plausibly 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.

Importantly, different maps may be neurally encoded using different sets of hippocampal neurons, or by partially overlapping sets of neurons that however "remap" across multiple maps - which implies that two neurons coding for adjacent locations in one map may code for non-adjacent locations in another map. It is this remapping mechanism that permits distinguishing two identical trajectories executed in two distinct spatial environments; and more broadly, avoiding interference between similar experiences [58]. Therefore, despite when we illustrate the model (e.g., in Figure 1) item codes of different maps are placed in the same location of the 15x15 maps, they may correspond to the activation of different hippocampal neurons.

Our model uses a mixture model for map probabilities, which implies that only one map is active at each time during learning and inference. While this may be a simplification, two key phenomena - theta-paced flickering [59] and the constant cycling between representations of different trajectories during navigation [60] - 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 closely related architectures, such as cloned HMM, are resistant to catastrophic forgetting [61].

Grouping multiple experiences in a single cluster or map is useful for generalization and to generate novel, never experienced trajectories; see Figure 2B. However, our model does not necessarily acquire a one-to-one correspondence between clusters and mazes. Depending on the training regime, and the similarity between novel and existing maps, the model can develop multiple clusters for the same maze or rather 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 [62], 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 [63,64]. Furthermore, it 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 [65]). These shortcut sequences were empirically found while animals learned independently two portions of the same maze – where distal cues (not included in our model) may have facilitated the integration of these memories rather than their segregation into fully 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 before navigation experience [16].

Systems level

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 [66] and mPFC [67,68]. Prefrontal structures may play a permissive role for hippocampal sequences to occur when needed, such as for example, during decision-making [42], hence mediating prospective (planning) and retrospective (learning) functions of hippocampal replays [5,18–21]. Furthermore, coordinated sequential replays across the hippocampus and PFC [68] and hippocampus and ventral striatum [69] may support goal-directed spatial navigation [70,71]. The cross-talk 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 [20], potentially contributing to bias their content and reorganize 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 proposed a novel theory of the hippocampal formation as a hierarchical generative model of spatiotemporal experiences. The first tenet of the theory is 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). In other words, it uses both maps and sequences as structures (or *inductive biases*) to organize items of experience. Organizing experiences into sequences and maps helps forming a coherent spatiotemporal context, permitting the model to link to the same latent map sets of trajectories experienced across different navigational episodes [14]. This idea is compatible with the well-established roles of the hippocampus in cognitive map formation [72] (and relatedly, in relational tasks) and the spontaneous generation of sequential dynamics [73] - which makes it a quintessential "sequence generator" [18,73].

The second tenet of the theory is that the hippocampal generative model supports generative replay: a powerful method for continual learning that uses fictive replays 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 perspective. From a machine learning perspective, generative replay prevents catastrophic forgetting [74] without requiring the verbatim storage of (an unbounded number of) previous experiences, as in "experience replay". Our results show that agents using generative replay, and especially prioritized generative replay, have superior performance compared to baseline agents (with no replay) and agents using experience replay (i.e., replaying from a memory buffer).

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.

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 maps from which sequences are generated implicitly encode behaviourally relevant (e.g., goal) or salience locations - which can also be seen as distal predictions [10,15,39,46] [34]. In other words, while sequences encode short-term predictions, maps encode long-term, average predictions of future locations. Our view is then compatible with the idea tat the hippocampus learns a predictive representation [75] but proposes that it is hierarchical.

Second, in our model maps that 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 novel experience. This mechanism implies that sequences generated before a navigational episode (preplays [16]) and after it (replays) are highly correlated, as demonstrated empirically [47] - 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, despite navigational experiences do not. This finding further exemplifies the generative nature of hippocampal sequences, which do not simply recapitulate recent experiences, neither in terms of content or dynamics.

The two methods we investigated – *generative replay* and its variant *prioritized generative replay* - are both appealing from both machine learning and biological perspectives; and it remains to be fully assessed which one (or which combination or both) best explains the phenomenology of hippocampal sequences. An intriguing possibility that remains to be tested in future research is that to the extent that neuromodulation influences replay expression [76], the activity of different neuromodulators may produce the selection of contextually-appropriate generative replay methods.

Previous theoretical and computational studies assigned a role to the hippocampus in supporting model-based, goal-directed navigation [18–20,25,37,43,77,78]. Our theory is in agreement with these ideas but goes beyond them, by introducing the idea of hierarchical, structured representations of sequences and maps, and by modelling explicitly how these representations are learned over time. Generative replay shares resemblances with another method 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 [79]. This idea has inspired previous proposals of the hippocampal spontaneous activity as a way to optimize reinforcement learning systems offline [80]. At difference with these ideas, our theory views the hippocampal formation as a generative model of spatiotemporal experiences, not a transition model; and spontaneous activity as a way to learn experiences, not reinforcement learning (RL) controllers – in keeping with the well know role of the hippocampus in episodic memory and latent learning, beyond RL. 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 using a transition model to train a controller: the fact that model errors accumulate when one resamples sequentially from them, making learning instable. Having said this, the above theories are not mutually exclusive. Generative replay can be used effectively to train reactive controllers, as hypothesized in DYNA, or support prospective functions of goal-directed controllers, such as the prediction of future paths. Therefore, our proposal is compatible with previous theories that the hippocampus may operate synergistically with other brain structures, such as the ventral striatum, to form a goal-directed controller for spatial navigation [71].

While we exemplified the proposed model in the domain of spatial navigation, it would support the autonomous learning of coherent spatiotemporal contexts for non-navigation tasks, too. Recent studies have shown that the hippocampus supports relational inference [81], temporal and rate coding of discrete sequences of non-spatial events such as (in rodents) odours [82] and button locations [83] and (in humans) of arbitrary items [17]; see also [84,85]. More broadly, novel findings imply the hippocampal formation in cognitive mapping in several domains of cognition [86–88]. 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 [88]. Theoretical

arguments suggest that while the brain might use multiple generative models [89], 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 immediately reconciles its roles in multiple, apparently disconnected cognitive functions, including episodic and autobiographic memory [81,90], imagination [91,92] and cognitive map-based spatial navigation [72,93]. These functions have often been studied in isolation and remain challenging to reconcile within an integrated theoretical proposal. The link between episodic memory and imagination becomes clear if one considers that generative replay supports both the rapid acquisition of novel episodic memories and the spontaneous generation of novel, never experienced sequences that are however structurally coherent with past experiences (as they stem from the same maps). The role of the hippocampus in cognitive map formation becomes apparent when one considers that maps may automatically include structural priors, such as the fact that there are "no gaps", which provide a relational scaffold to organize experiences in multiple domains of cognition. Finally, and speculatively, the hippocampal generative model may contribute to autobiographic aspects od memory, as by continuously encoding and replaying episodes it may afford a sense of continual self across time. This may also link to prospective function, as a model provides not just a memory of “the past me” but also inference of “the future me”.

In sum, the computational framework of hierarchical generative modelling and generative replay reconciles several apparently disconnected ideas on 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 needs to 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 (but also for 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 of the observations into $\mathbb{D}$, called *item*, 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 code for a *sequence* of items $S \equiv \boldsymbol{\xi}(1), \dots, \boldsymbol{\xi}(T)$ in $\mathbb{D}$:

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

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, 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 no or few repetitions, with the graded values that essentially encode the order of items.

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 that map. 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 $\Theta \equiv \{\boldsymbol{\theta}_k\}_{k=1,..,K}$, $p_{\boldsymbol{\Theta}}(\mathbf{y}|k) \equiv p(\mathbf{y}|k, \boldsymbol{\theta}_k)$, 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:

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

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

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

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 $\boldsymbol{\theta}_k$ matches closer 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, hence permitting to improve the a-priori information available for the next trials; see the section on "Parameter learning" for details.

Differently, 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 found value of) $\tau = 0.9$ and $\boldsymbol{\Theta}$ is the whole 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 approach to select (and adapt) which of the K maps maximizes the posterior; namely $\bar{k} = argmax(\boldsymbol{z}(t))$. By doing so, latent state inference guides learning to gradually form distinct maps for different environments.

At the end of the inferential process of 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, differently from the 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, by adding at each time step synapse-weighted bottom-up signals encoding current observations.

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 dynamic 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 encoding the fact that the locations of successive items change gradually, and hence prescribing transitions sampled from a normal distribution centered on past observation $\boldsymbol{\xi}(t-1)$, 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 $\hat{\boldsymbol{\xi}}(t) \sim \boldsymbol{\mathcal{N}}(\boldsymbol{\xi}(t-1), \Sigma)$ is an observation estimated from an elliptic distribution, with covariance Σ encoding information about the current velocity (i.e., the velocity at the tip) of the sequence $\mathbf{y}$.

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 dynamic 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 permits carrying over long-term environmental information (e.g., maze properties) 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 best map. The updating is done at the end of the hierarchical inference.

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:

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

that 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 updating is done only when $t > 1$, to estimate the most likely map associated to 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 novel clusters as the model observes novel data [94,95]. 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 on 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 [47] 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, Speakman's rank order correlation is used to calculate the similarity between the (vectorized) ranks maps. Examples of rank maps and rank correlations are shown on Fig. 6A and statistics of rank-correlations across conditions is shown on 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 every two 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 the diffusion of location 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.

References

1. McCloskey, M. & Cohen, N. J. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem. in *Psychology of Learning and Motivation* (ed. Bower, G. H.) vol. 24 109–165 (Academic Press, 1989).
2. McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. 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 (1995).
3. O'Reilly, R. C., Bhattacharyya, R., Howard, M. D. & Ketz, N. Complementary learning systems. *Cogn. Sci.* **38**, 1229–1248 (2014).
4. Buzsáki, G. Hippocampal sharp wave-ripple: A cognitive biomarker for episodic memory and planning. *Hippocampus* **25**, 1073–1188 (2015).
5. Foster, D. J. Replay comes of age. *Annu. Rev. Neurosci.* **40**, 581–602 (2017).
6. Foster, D. J. & Wilson, M. A. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* **440**, 680–683 (2006).
7. Jadhav, S. P., Kemere, C., German, P. W. & Frank, L. M. Awake Hippocampal Sharp-Wave Ripples Support Spatial Memory. *Science* **336**, 1454–1458 (2012).
8. Kumaran, D., Hassabis, D. & McClelland, J. L. What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends Cogn. Sci.* **20**, 512–534 (2016).
9. Mnih, V. *et al.* Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
10. Pfeiffer, B. E. & Foster, D. J. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79 (2013).
11. Stella, F., Baracskay, P., O'Neill, J. & Csicsvari, J. Hippocampal Reactivation of Random Trajectories Resembling Brownian Diffusion. *Neuron* **102**, 450-461.e7 (2019).
12. Liu, K., Sibille, J. & Dragoi, G. Generative Predictive Codes by Multiplexed Hippocampal

Neuronal Tuplets. *Neuron* **99**, 1329-1341.e6 (2018).

13. Dragoi, G. & Tonegawa, S. Distinct preplay of multiple novel spatial experiences in the rat. *Proc. Natl. Acad. Sci.* **110**, 9100–9105 (2013).
14. Gupta, A. S., van der Meer, M. A. A., Touretzky, D. S. & Redish, A. D. Hippocampal replay is not a simple function of experience. *Neuron* **65**, 695–705 (2010).
15. Ólafsdóttir, H. F., Barry, C., Saleem, A. B., Hassabis, D. & Spiers, H. J. Hippocampal place cells construct reward related sequences through unexplored space. *eLife* **4**, e06063 (2015).
16. Dragoi, G. & Tonegawa, S. Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature* **469**, 397–401 (2011).
17. Liu, Y., Dolan, R. J., Kurth-Nelson, Z. & Behrens, T. E. J. Human Replay Spontaneously Reorganizes Experience. *Cell* (2019) doi:10.1016/j.cell.2019.06.012.
18. Pezzulo, G., Kemere, C. & Meer, M. van der. Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition. *Ann. N. Y. Acad. Sci.* **1396**, 144–165 (2017).
19. Pezzulo, G., van der Meer, M. A. A., Lansink, C. S. & Pennartz, C. M. A. Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* **18**, 647–657 (2014).
20. Penagos, H., Varela, C. & Wilson, M. A. Oscillations, neural computations and learning during wake and sleep. *Curr. Opin. Neurobiol.* **44**, 193–201 (2017).
21. Pezzulo, G., Donnarumma, F., Maisto, D. & Stoianov, I. Planning at decision time and in the background during spatial navigation. *Curr. Opin. Behav. Sci.* **29**, 69–76 (2019).
22. Mocanu, D. C., Vega, M. T., Eaton, E., Stone, P. & Liotta, A. Online Contrastive Divergence with Generative Replay: Experience Replay without Storing Data. *ArXiv161005555 Cs* (2016).
23. Shin, H., Lee, J. K., Kim, J. & Kim, J. Continual Learning with Deep Generative Replay.

ArXiv170508690 Cs (2017).

24. van de Ven, G. M., Siegelmann, H. T. & Tolias, A. S. Brain-inspired replay for continual learning with artificial neural networks. *Nat. Commun.* **11**, 4069 (2020).
25. Stoianov, I. P., Pennartz, C. M. A., Lansink, C. S. & Pezzulo, G. Model-based spatial navigation in the hippocampus-ventral striatum circuit: A computational analysis. *PLOS Comput. Biol.* **14**, e1006316 (2018).
26. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* **2**, 79–87 (1999).
27. Friston, K. J. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* **360**, 815–836 (2005).
28. Dabaghian, Y., Brandt, V. L. & Frank, L. M. Reconceiving the hippocampal map as a topological template. *eLife* **3**, e03476 (2014).
29. Donnarumma, F., Maisto, D. & Pezzulo, G. Problem Solving as Probabilistic Inference with Subgoaling: Explaining Human Successes and Pitfalls in the Tower of Hanoi. *PLOS Comput. Biol.* **12**, e1004864 (2016).
30. Maisto, D., Donnarumma, F. & Pezzulo, G. Nonparametric Problem-Space Clustering: Learning Efficient Codes for Cognitive Control Tasks. *Entropy* **18**, 61 (2016).
31. Maisto, D., Donnarumma, F. & Pezzulo, G. Divide et impera: subgoaling reduces the complexity of probabilistic inference and problem solving. *J. R. Soc. Interface* **12**, 20141335 (2015).
32. Gershman, S. J. The Successor Representation: Its Computational Logic and Neural Substrates. *J. Neurosci.* **38**, 7193–7200 (2018).
33. Aoki, Y., Igata, H., Ikegaya, Y. & Sasaki, T. The Integration of Goal-Directed Signals onto Spatial Maps of Hippocampal Place Cells. *Cell Rep.* **27**, 1516-1527.e5 (2019).
34. Bhattarai, B., Lee, J. W. & Jung, M. W. Distinct effects of reward and navigation history on

hippocampal forward and reverse replays. *Proc. Natl. Acad. Sci.* **117**, 689–697 (2020).

35. Moore, A. W. & Atkeson, C. G. Prioritized sweeping: Reinforcement learning with less data and less time. *Mach. Learn.* **13**, 103–130 (1993).
36. Fuhs, M. C. & Touretzky, D. S. Context learning in the rodent hippocampus. *Neural Comput.* **19**, 3173–3215 (2007).
37. Penny, W. D., Zeidman, P. & Burgess, N. Forward and Backward Inference in Spatial Cognition. *PLoS Comput Biol* **9**, e1003383 (2013).
38. Sanders, H., Wilson, M. A. & Gershman, S. J. *Hippocampal Remapping as Hidden State Inference*. (2019).
39. Ambrose, R. E., Pfeiffer, B. E. & Foster, D. J. Reverse Replay of Hippocampal Place Cells Is Uniquely Modulated by Changing Reward. *Neuron* **91**, 1124–1136 (2016).
40. Foster, D. J. & Wilson, M. A. Hippocampal theta sequences. *Hippocampus* **17**, 1093–1099 (2007).
41. Johnson, A. & Redish, A. D. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J Neurosci* **27**, 12176–12189 (2007).
42. Redish, A. D. Vicarious trial and error. *Nat. Rev. Neurosci.* **17**, 147–159 (2016).
43. Lisman, J. & Redish, A. D. Prediction, Sequences and the Hippocampus. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 1193–1201 (2009).
44. Sanders, H., Rennó-Costa, C., Idiart, M. & Lisman, J. Grid Cells and Place Cells: An Integrated View of their Navigational and Memory Function. *Trends Neurosci.* **38**, 763–775 (2015).
45. Shimazaki, H. Neurons as an Information-theoretic Engine. *ArXiv151207855 Phys. Q-Bio* (2018) doi:10.1007/978-3-319-71976-4.
46. Wikenheiser, A. M. & Redish, A. D. Hippocampal theta sequences reflect current goals. *Nat. Neurosci.* **18**, 289–294 (2015).
47. Liu, K., Sibille, J. & Dragoi, G. Preconfigured patterns are the primary driver of offline

multi-neuronal sequence replay. *Hippocampus* **29**, 275–283 (2019).

48. Dragoi, G. Cell assemblies, sequences and temporal coding in the hippocampus. *Curr. Opin. Neurobiol.* **64**, 111–118 (2020).
49. Whittington, J. C. *et al.* The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalisation in the hippocampal formation. *bioRxiv* 770495 (2019) doi:10.1101/770495.
50. Whittington, J., Muller, T., Mark, S., Barry, C. & Behrens, T. Generalisation of structural knowledge in the hippocampal-entorhinal system. in *Advances in Neural Information Processing Systems 31* (eds. Bengio, S. et al.) 8484–8495 (Curran Associates, Inc., 2018).
51. Boccara, C. N., Nardin, M., Stella, F., O'Neill, J. & Csicsvari, J. The entorhinal cognitive map is attracted to goals. *Science* **363**, 1443–1447 (2019).
52. Butler, W. N., Hardcastle, K. & Giocomo, L. M. Remembered reward locations restructure entorhinal spatial maps. *Science* **363**, 1447–1452 (2019).
53. Plitt, M. H. & Giocomo, L. M. Experience dependent contextual codes in the hippocampus. *bioRxiv* 864090 (2019) doi:10.1101/864090.
54. Aronov, D., Nevers, R. & Tank, D. W. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature* **543**, 719–722 (2017).
55. Constantinescu, A. O., O'Reilly, J. X. & Behrens, T. E. J. Organizing conceptual knowledge in humans with a gridlike code. *Science* **352**, 1464–1468 (2016).
56. Mok, R. M. & Love, B. C. A non-spatial account of place and grid cells based on clustering models of concept learning. *Nat. Commun.* **10**, 1–9 (2019).
57. Gulli, R. A. *et al.* Context-dependent representations of objects and space in the primate hippocampus during virtual navigation. *Nat. Neurosci.* **23**, 103–112 (2020).
58. Colgin, L. L., Moser, E. I. & Moser, M.-B. Understanding memory through hippocampal remapping. *Trends Neurosci.* **31**, 469–477 (2008).

59. Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I. & Moser, M.-B. Theta-paced flickering between place-cell maps in the hippocampus. *Nature* **478**, 246–249 (2011).

60. Kay, K. *et al.* Constant Sub-second Cycling between Representations of Possible Futures in the Hippocampus. *Cell* **180**, 552-567.e25 (2020).

61. Rikhye, R. V. *et al.* Learning cognitive maps as structured graphs for vicarious evaluation. *bioRxiv* 864421 (2020) doi:10.1101/864421.

62. Grieves, R. M., Wood, E. R. & Dudchenko, P. A. Place cells on a maze encode routes rather than destinations. *eLife* **5**, (2016).

63. Spiers, H. J., Hayman, R. M. A., Jovalekic, A., Marozzi, E. & Jeffery, K. J. Place field repetition and purely local remapping in a multicompartment environment. *Cereb. Cortex N. Y. N 1991* **25**, 10–25 (2015).

64. Grieves, R. M., Jenkins, B. W., Harland, B. C., Wood, E. R. & Dudchenko, P. A. Place field repetition and spatial learning in a multicompartment environment. *Hippocampus* **26**, 118–134 (2016).

65. Liu, Y., Kurth-Nelson, Z. & Behrens, T. Non-spatial neural replay in building and updating world models in humans. in *Cosyne 2018* (2018).

66. Chenani, A. *et al.* Hippocampal CA1 replay becomes less prominent but more rigid without inputs from medial entorhinal cortex. *Nat. Commun.* **10**, 1341 (2019).

67. Shin, J. D., Tang, W. & Jadhav, S. P. Dynamics of Awake Hippocampal-Prefrontal Replay for Spatial Learning and Memory-Guided Decision Making. *Neuron* **104**, 1110-1125.e7 (2019).

68. Tang, W., Shin, J. D. & Jadhav, S. P. Multiple time-scales of decision making in the hippocampus and prefrontal cortex. *bioRxiv* 2020.10.17.343699 (2020) doi:10.1101/2020.10.17.343699.

69. Lansink, C. S., Goltstein, P. M., Lankelma, J. V., McNaughton, B. L. & Pennartz, C. M. Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biol.* **7**,

e1000173 (2009).

70. Pezzulo, G., van der Meer, M. A. A., Lansink, C. S. & Pennartz, C. M. A. Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* **18**, 647–657 (2014).
71. van der Meer, M., Kurth-Nelson, Z. & Redish, A. D. Information processing in decision-making systems. *The Neuroscientist* **18**, 342–359 (2012).
72. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).
73. Buzsáki, G. & Tingley, D. Space and Time: The Hippocampus as a Sequence Generator. *Trends Cogn. Sci.* **22**, 853–869 (2018).
74. Shin, H., Lee, J. K., Kim, J. & Kim, J. Continual Learning with Deep Generative Replay. *ArXiv170508690 Cs* (2017).
75. Recanatesi, S. *et al.* Signatures and mechanisms of low-dimensional neural predictive manifolds. *bioRxiv* 471987 (2018) doi:10.1101/471987.
76. Atherton, L. A., Dupret, D. & Mellor, J. R. Memory trace replay: the shaping of memory consolidation by neuromodulation. *Trends Neurosci.* **38**, 560–570 (2015).
77. Pezzulo, G., Rigoli, F. & Chersi, F. A mixed instrumental controller can combine habitual and goal-directed choice. in *Proceedings of CogSci 2012* (2012).
78. van der Meer, M., Kurth-Nelson, Z. & Redish, A. D. Information processing in decision-making systems. *The Neuroscientist* **18**, 342–359 (2012).
79. Sutton, R. S. Integrated modeling and control based on reinforcement learning and dynamic programming. in *Advances in Neural Information Processing Systems* (ed. Touretzky, D. S.) vol. 3 471–478 (1991).
80. Mattar, M. G. & Daw, N. D. Prioritized memory access explains planning and hippocampal replay. *bioRxiv* 225664 (2017) doi:10.1101/225664.
81. Eichenbaum, H. A cortical-hippocampal system for declarative memory. *Nat. Rev. Neurosci.*

1, 41–50 (2000).

82. Terada, S., Sakurai, Y., Nakahara, H. & Fujisawa, S. Temporal and Rate Coding for Discrete Event Sequences in the Hippocampus. *Neuron* **94**, 1248-1262.e4 (2017).
83. Miller, K. J., Botvinick, M. M. & Brody, C. D. Dorsal hippocampus contributes to model-based planning. *Nat. Neurosci.* **20**, 1269–1276 (2017).
84. Schuck, N. W. & Niv, Y. Sequential replay of nonspatial task states in the human hippocampus. *Science* **364**, (2019).
85. Shahbaba, B. *et al.* Hippocampal ensembles represent sequential relationships among discrete nonspatial events. *bioRxiv* 840199 (2019) doi:10.1101/840199.
86. Behrens, T. E. J. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* **100**, 490–509 (2018).
87. Bellmund, J. L. S., Gärdenfors, P., Moser, E. I. & Doeller, C. F. Navigating cognition: Spatial codes for human thinking. *Science* **362**, eaat6766 (2018).
88. Buzsáki, G. & Moser, E. I. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nat. Neurosci.* **16**, 130–138 (2013).
89. Pezzulo, G., Donnarumma, F., Iodice, P., Maisto, D. & Stoianov, I. Model-Based Approaches to Active Perception and Control. *Entropy* **19**, 266 (2017).
90. Tulving, E. Episodic Memory: From Mind to Brain. *Annu. Rev. Psychol.* **53**, 1–25 (2002).
91. Buckner, R. L. The role of the hippocampus in prediction and imagination. *Annu Rev Psychol* **61**, 27–48, C1-8 (2010).
92. Schacter, D. L. *et al.* The Future of Memory: Remembering, Imagining, and the Brain. *Neuron* **76**, 677–694 (2012).
93. O'keefe, J. & Nadel, L. *The hippocampus as a cognitive map*. vol. 3 (Clarendon Press Oxford, 1978).
94. Stoianov, I., Genovesio, A. & Pezzulo, G. Prefrontal Goal Codes Emerge as Latent States in

Probabilistic Value Learning. *J. Cogn. Neurosci.* (2015).

95. Gershman, S. J. & Blei, D. M. A tutorial on Bayesian nonparametric models. *J. Math. Psychol.* **56**, 1–12 (2012).