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, Via San Martino della Battaglia 44, Rome 00185, Italy

2 *Institute for High Performance Computing and Networking, National Research Council, Via P. Castellino, 111, Naples 80131, Italy*

** Contact: giovanni.pezzulo@istc.cnr.it*

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, for the sake of 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 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 to avoid that the novel acquired knowledge interferes with past memories, causing their *catastrophic forgetting* [1].

The hippocampus has long been considered as central to the solution of 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 second, semantic memory system located in the cortex, which gradually integrates and consolidates memories [2], [3].

The idea that the hippocampus stores and reactivates episodic memories offline received empirical support from rodent studies during spatial navigation – a domain in which the hippocampus is particularly important. These studies have shown that that in phases of wakeful rest during spatial navigation, and during subsequent sleep, the rodent hippocampus spontaneously regenerates sequences of neural activations (within so-called *sharp wave ripples*), which resemble sequences observed during actual animal behaviour [4] (e.g., sequences of place cells that are active when the animal follows a specific spatial trajectory), in a time-compressed manner [5]. Such *internally generated hippocampal sequences* are often interpreted as "replays" of previous experiences from an episodic memory buffer located in the hippocampus; and their disruption was shown to impair memory-based navigation [6].

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, where the same agent architecture has to learn multiple tasks, one after the other [7]. In most neural network architectures, learning a novel task (task C) would imply forgetting previously learned tasks (tasks A and B), as the memories of C would interfere with those of A and B. An agent using experience replay can avoid this problem. Essentially, when it 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.

Yet, rodent and human studies have shown that the hippocampus does not literally "replay" previous experiences from a memory buffer. Internally generated hippocampal sequences during sleep or wakeful rest can depict paths to future goal locations rather than only past trajectories, possibly supporting planning and imagination beyond purely memory function [8]; random trajectories, resembling Brownian diffusion [9]; trajectories that have never been directly explored but yet reflect prior spatial knowledge [10], [11], such as shortcuts [12] and future event locations [13] (and in human studies of non-spatial domains, learned rules [14]); or even *preplay* sequences of cells that later on (i.e., when the environment is experienced) map onto real trajectories [15].

These findings indicate that internally generated hippocampal sequences during sleep or wakeful rest have flexible and prospective aspects [5], [16]–[19], which are difficult to reconcile with the idea of "replays" of experiences from a memory buffer. Here we advance a novel theoretical proposal to explain the flexible aspects of internally generated hippocampal sequences and their roles in episodic memory and beyond.

We propose that the hippocampal formation is a *hierarchical generative model* that organizes sequential experiences (e.g., rodent trajectories during spatial navigation), into a set of coherent but 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.

The theory is based on two main assumptions. The first assumption is that the hippocampal formation encodes experiences by *learning a hierarchical generative model*, as opposed to storing them in a memory system. The proposed hierarchical structure 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. At this first layer, the item codes are pattern-separated and hence deprived of context. The role of the second and third layers is to organize the items into coherent spatiotemporal contexts. For this, 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 coherent and mutually exclusive spatiotemporal 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 firstly disentangle *items* of experience and then organize them into *sequences* (temporally ordered items) and *maps* (spatially clustered sequences).

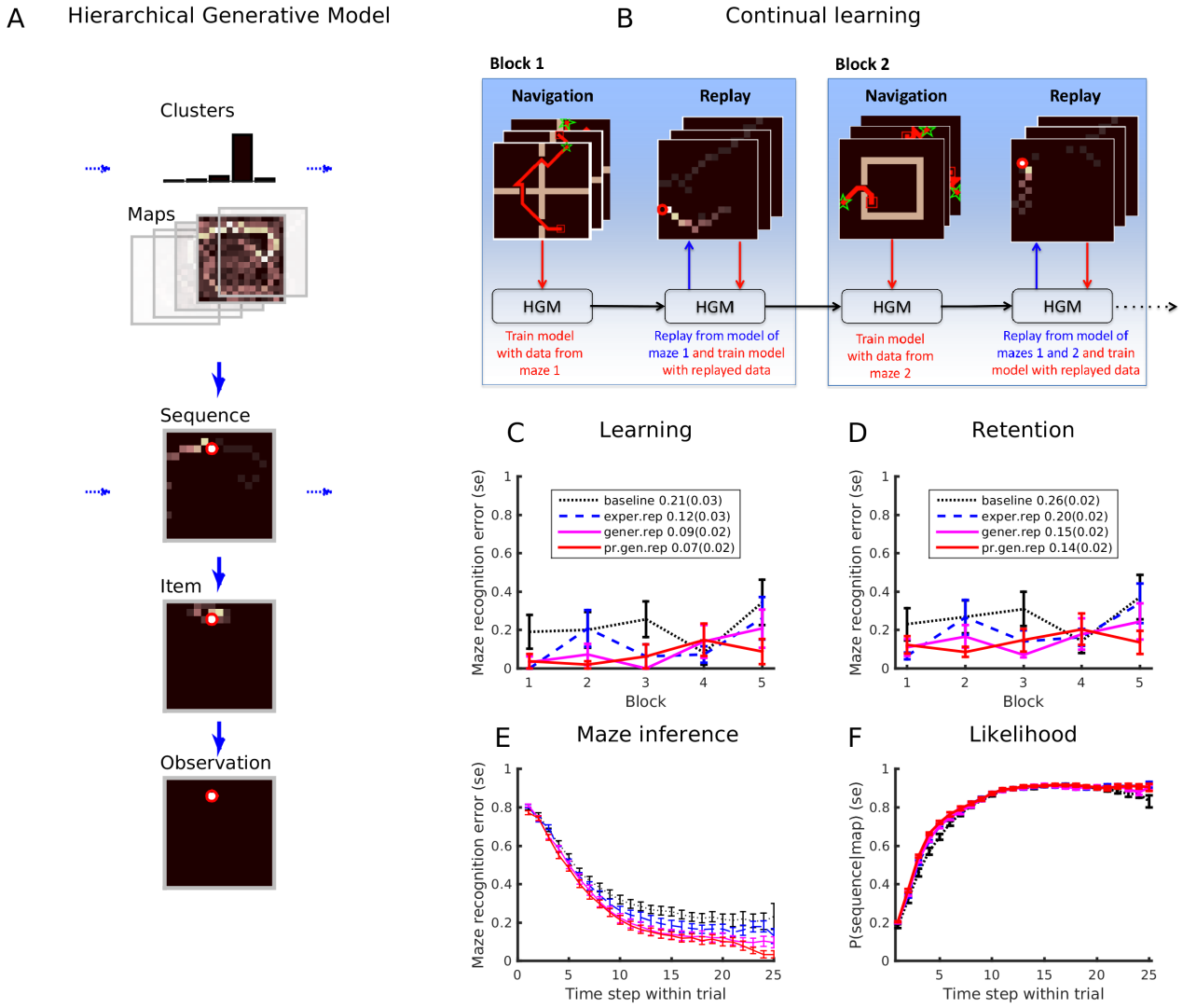

Figure 1. Hippocampal generative model. A: Components of the hierarchical generative model. From top to bottom, clusters / maps, sequences, items and observations. The red circle denotes the agent's current position. Colour codes correspond to probabilities, from high (white) to low (black). See the main text for explanation. Note that the current implementation focuses on spatial navigation and uses codes for observation, item, sequence, and maps that can be directly mapped to spatial locations within 2D maps. However, the method is more general and all the

above codes could be mapped to other, arbitrary spaces, or factorized (e.g., into separate maps where items are objects or locations), providing their structural (e.g., sequential) relations are preserved. B: Structure of the continual learning experiment. The experiment includes five blocks and for each block, the learning agents trains the generative model based on 20 trajectory data from one single maze. Between blocks, they can replay experiences. 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.

The second assumption is that internally generated hippocampal sequences are manifestations of the inferential dynamics of the generative model, not "replays" from a memory. Specifically, the hippocampal generative model supports *generative replay*: a method for continual learning that uses fictive experiences resampled from a generative model rather than memorized experiences, as in experience replay. Generative replay uses fictive experiences to self-train the model that generates them (as well as to train separate models or controllers) and is effective in preventing catastrophic forgetting [20]–[22]. Furthermore, by using a model, generative replay avoids the main shortcoming of experience replay: the necessity of unbounded memory resources to learn multiple tasks. Furthermore, generative replay explains flexible and prospective aspects of internally generated hippocampal sequences, which are difficult to reconcile with experience replay. Generative replay is also similar to 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 [23]. However, learning a generative model is substantially simpler than learning a transition model and less error prone. Generative replay hence 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.

Results

In this section, we present the simulation of a continual learning task, consisting in learning multiple spatial mazes from (simulated) navigation trajectories. We compare four learning agents that use the same hierarchical generative model (Figure 1A) but different replay methods: no replay, experience replay, generative replay, and a novel variant of generative replay ("prioritized generative replay"). Besides illustrating the efficacy of generative replay for continual learning, this simulation also shows how the generative model supports spatial map formation, inference, and the generation of fictive experiences that capture essential aspects of internally generated hippocampal sequences.

The continual learning task

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 different mazes in different blocks (mazes are selected randomly for each agent and block). During each block, the four learning agents receive as data 20 spatial trajectories from the maze they are current in and use these trajectory data to update their hierarchical generative models. These trajectory data are generated by a separate model-based Bayesian controller that, for each

trial and maze, starts from a random location and reaches a random goal location, by following a near-optimal trajectory [24]. 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 a formal specification). 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 only thing the agent observes is its current spatial location in the maze. In other words, it does not observe the whole trajectory data simultaneously, but one location after the other.

The first hidden layer of the hierarchical model encodes observations into *item* codes. The item codes are inferred probabilistically, using a scheme similar to predictive coding [25], [26], which integrates four sources of information: top-down predictions from the higher hierarchical layers, bottom-up observations, lateral information from the previous sequence, and a (fixed) model of movement dynamics.

The second hidden layer of the hierarchical model encodes *sequences* of items, which are inferred by considering two sources of information: lateral information from the previous sequence, and bottom-up information from the last inferred item.

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 observations can belong to different clusters (i.e., different trajectories can belong to different spatial maps). The model maintains a (categorical) probability distribution over the clusters of the mixture model, with the cluster having the highest probability corresponding to the "currently inferred" map, i.e., the map the agent believes it is in, based on the trajectory data it is observing. This probability distribution is continuously updated, in a Bayesian way, as the agent observes new data. The currently inferred map is the (only) one that is updated during learning and used to generate fictive experiences.

The "content" of each cluster (i.e., the map) is a probability distribution over an arbitrary structure, which in our simulations is a 2D spatial structure. Importantly, given their probabilistic representation, and the way they are trained, maps are biased: they encode not just items in their spatial locations but also simultaneously the (prior) probability of visiting these locations in the future, under a goal-directed policy. This is intuitively zero at locations that cannot be crossed (e.g., walls) and high in goal and other behaviourally relevant locations, such as junction points or bottlenecks [27]. In the current implementation, the probabilities are updated using visitation counts during both real and fictive experiences; but other methods such as algorithmic complexity [28]–[30] and successor representations [31] produce similar results. The reason why goal (and other behaviourally relevant) locations are assigned higher probabilities is due to the fact that the training trajectory data come from a goal-directed controller; the distribution would have been flatter if we used a random controller instead. This bias towards behaviourally relevant locations will become important later, when we will discuss training using generative replay.

In our simulations, the generative model uses five clusters. This is done to facilitate the analyses, as there is (potential) a one-to-one mapping between the five clusters and the five mazes; and our results will show that most simulations converge to this solution or a close approximation. However, the methods are more general and can extend to an arbitrary number of clusters, using a nonparametric method that automatically selects the best number of clusters given the agent's observations (see section on "Nonparametric Method").

In parallel, for each cluster, the model maintains an auxiliary, *prioritized map* that encodes the average surprise of the model at each location (rather than visitation counts). Intuitively, the prioritized maps describe how unpredicted each location was, possibly marking portions of the map that need more updating, similar to prioritized sweeping in reinforcement learning [32]. This bias towards surprising locations will become important later, when we will discuss training using prioritized generative replay.

Training procedure

As discussed above, the four 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 to select the first item.

Simulation results

The results of the continual learning experiment are shown in Figure 1C-F. Figure 1C shows the agents' learning error, measured in terms of 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 [33]–[35]. For this, we first establish (for each agent) what is the most probable (or preferred) maze for each cluster at the end of the learning (Figure 2B-C reveal that in almost all cases one cluster is much more probable than the others). This preferred cluster is the "ground truth", or the cluster that the agent would select if it had enough data. Then we consider at 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 outperform the baseline and the experience replay agents, which means that they infer more often in which map they are given the trajectory data they are experiencing.

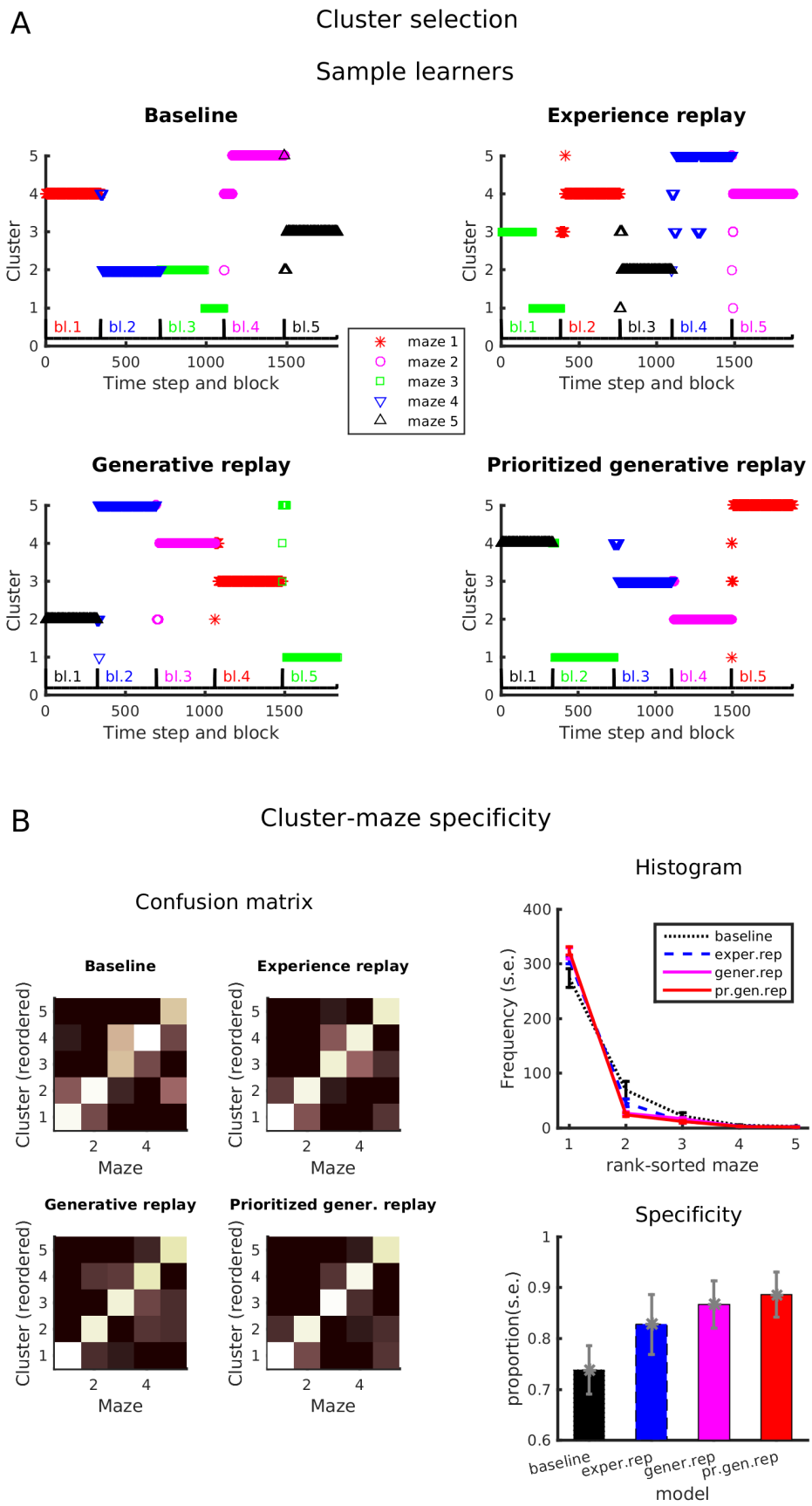

Figure 2. Clustering emerged during the continual learning experiment. (A) Cluster selection during navigation in sample replicas of each agent. The abscissa is double-labelled with time step (below) and colour coded blocks (above). The ordinate indexes the clusters (1 to 5) selected

during the blocks. The legend colour-and-symbol labels the maze used in each block. In many cases, agents select one single cluster for each block (especially after offline replay). (B) Analysis of cluster-maze specificity across all n=16 replicas of the four learning agents. Left: average cluster-maze confusion matrices. Right: Average frequency of cluster selection across clusters as a function of cluster-specific maze-preference rank (top) and proportion of the top-rank cluster selected during navigation (bottom).

Figure 1D shows the same learning error, but computed during a separate (retention) test session executed after the agents experienced all the five blocks. Note that during this retention test session, we reset the cluster probabilities at each trial, hence the agent treats each new trajectory as independent; this makes the test particularly challenging, as the agents cannot consider the evidence they accumulated when observing previous sequences. Furthermore, learning novel mazes could lead to the catastrophic forgetting of older mazes. Again, our results show that the two agents using generative replay outperform the baseline and the experience replay agents, suggesting that they suffer much less from 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 are observing (recall that the agents receive one observation after the other and never observe entire trajectories, nor they know the maze or location in which they are). Our results show that as they accumulate more evidence within trials, the agents' performance increases, both in terms of their capacity to infer the correct 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. Coherent with the above overall performance measures, the two agents using generative replay outperform the other two, both in terms of speed and accuracy of inference.

To better understand the learning procedure, Figure 2A plots the clusters that sample replicas of the learning agents actually select during training. The results show that all the agents tend to select one cluster for each maze. However, the two agents using generative replay do this more systematically; and especially the agent using prioritized generative replay. This is confirmed by Figure 2B, which shows the cluster-maze "confusion matrices" of the four agents: the two agents using generative replay tend to assign all the trajectories of the same maze to the same cluster, hence disambiguating more clearly the five mazes they experienced.

Grouping multiple experiences in a single cluster or map is useful for generalization (e.g., to generate never-experienced or shortcut trajectories; see below). 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 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 [36], which show route-dependent rather than map-dependent firing. The second possibility – reusing the same clusters across multiple mazes – 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 [12] (see also [37]). These shortcut sequences were empirically found while animals learned independently two portions of the same maze – where distal cues may have facilitated the integration of these memories rather than their segregation into fully distinct

maps [12]. 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 actual navigation experience [15].

Generative replay of sequences in the presence or absence of external stimuli

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. Generative replay provides a novel explanation of internally generated hippocampal sequences, suggesting that they are *resampled sequences from the latent map (prior) distribution* in the absence of external stimuli (this is sometimes called *inputless decoder* in machine learning), rather than replays of previous experiences as commonly assumed.

Importantly, the hierarchical model spontaneously generates sequences both in the absence of external stimuli and in their presence. In the absence of external stimuli (e.g., during generative replay), the whole sequence consists in fictive observations. The starting points of generative replays are sampled from the currently inferred map, i.e., the map that currently has the highest probability (but note that this probability is continuously updated, and hence successive generative replays can be generated by different maps). Importantly, the maps are biased, as they encode the *probability* of items at given locations, which indirectly signals their importance (or surprise in prioritized maps). This implies that in the absence of external stimuli, generative replay events will tend to start from important locations, such as goal locations, as shown empirically [8], [38]; see Figure 3. Note that by selectively resampling sequences that encode goals, generative replay further biases the agent's maps to focus on goal- (or more broadly, behaviourally relevant) locations that have adaptive value.

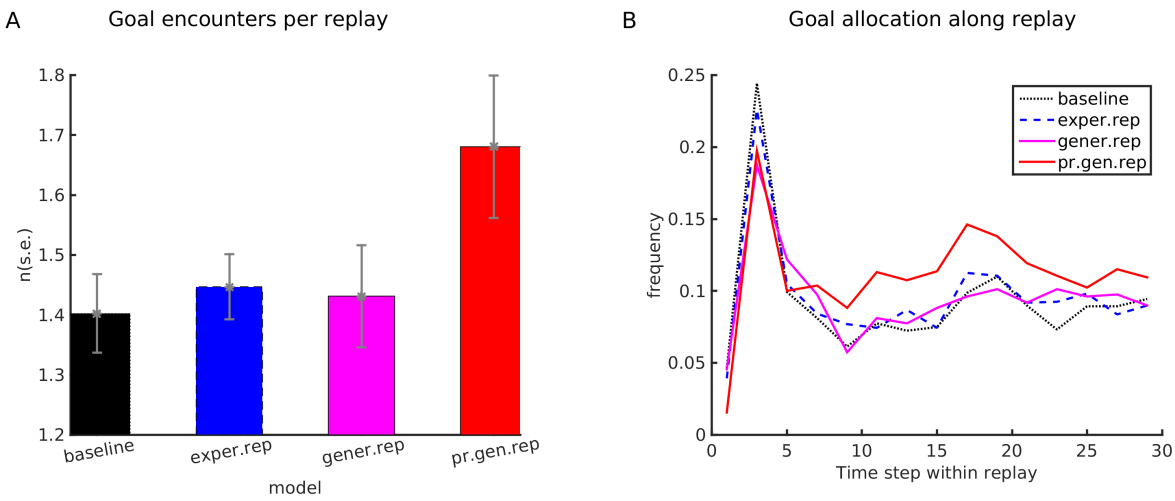

Figure 3. Analysis of goal-sensitivity during generative replay in the absence of external stimuli (akin to hippocampal replays during sleep). Note that in order to compare the four agents, in this analysis they are all allowed to do generative replay after each learning block (but still the baseline and experience replay agents do not learn from their generative replays). (A) Frequency of fictive visits to goal locations during generative replay. (B) Frequency of fictive visits to goal locations along generative replays (with all generative replays lasting 30 time steps for simplicity). Generative replays tend to start close to goal locations (but also show some goal sensitivity near the end of the fictive trajectories).

In the presence of external inputs, such as a currently experienced sequence, the hierarchical model supports *sequence forward prediction*, i.e., the prediction of the continuation of the sequence, conditioned on the current observation and the currently inferred map. This mechanism can potentially explain two important hippocampal phenomena.

First, at difference with replays during sleep, repays occurring during awake state can start around the animal's position and be biased towards goal locations [8]. 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. Then, for the very same reason explained above (that latent maps are biased) sequence predictions would be biased towards important items or locations, such as goal locations. This would imply that the differences between replays during awake state and sleep would depend on the way generative replays are initialized, given the presence (during awake state) or absence (during sleep) of bottom-up observations – with the former more biased to originate from the animal's current location to reach behaviourally relevant locations, and the latter more biased to originate from behaviourally relevant locations.

Second, during navigation, the hippocampal encodes *theta sequences* [39], or sequences of place cells, within each theta cycle. Theta sequences have been implied in both memory encoding and the prediction of future spatial positions along the animal's current spatial trajectory [40]–[42]. 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 [43] 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 [44], 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 [45], which in our model would occur given 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 arise in different conditions. Replays during sleep are the "purest" (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 [12]. 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 importance of items or their surprise in prioritized maps (or a combination of importance and surprise if one combines these maps during replay).

Nonparametric method

So far, we discussed a model in which the number of cluster is fixed a-priori. However, a straightforward nonparametric extension of the same method permits an open-ended expansion of the number of clusters as the agent receives novel observations; see the Methods section for details. The resulting nonparametric model is plausibly better suited to capture the ability of the hippocampus to potentially encode a very large number of experiences and to restructure itself during development and learning.

We tested the efficacy of the nonparametric method in the same structure learning experiment introduced above, with the only difference that each maze was presented 4 times in different blocks, resulting in 20 navigation blocks. This procedure provides more training data, which are required to show the efficiency of clustering by the nonparametric method; see Figure 4.

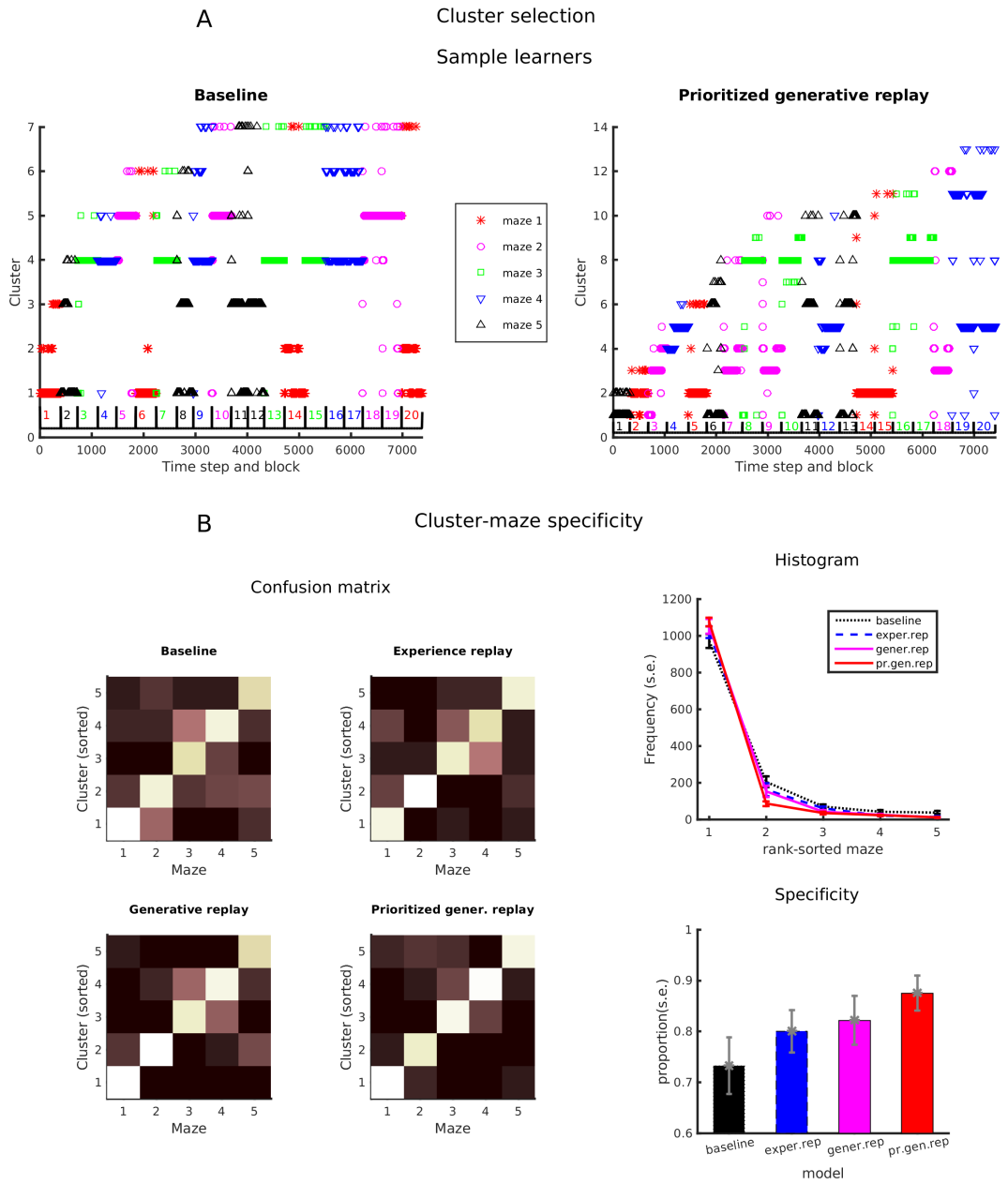

Figure 4. Clustering emerged during continuous nonparametric learning. (A) Cluster selection during navigation in sample baseline and prioritized generative replay agents. Note the greater maze-cluster consistency in the latter agent. (B) Analysis of the consistency of cluster-maze

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.

Our results show that the model performance was in line with the model with fixed number of clusters; with the four agents (baseline, experience replay, generative replay and prioritized experience replay) showing a maze recognition error of 0.261 (se 0.029), 0.148 (se 0.017), 0.097 (se 0.016), and 0.095 (se 0.019), respectively (n=16 replicas of each agent). As expected from a nonparametric method, all the agents learned the implicit trajectory clustering task with some redundancy compared to the parametric method (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 generative replay agents develop more specific clusters, which would explain their better performance. These results illustrate that our generative replay approach to continual learning does not necessarily require a-priori information about the number of clusters but can directly infer it from data. This (nonparametric) challenge is similar to what the hippocampal formation has to face when learning multiple experiences. Note however that the hippocampal formation could use a number of (distal sensory) cues to help determining, for example, whether the maze has changed, which are instead unavailable to our nonparametric agents.

Putative biological implementation of the hierarchical generative model

We described the hippocampal formation as a hierarchical generative model that organizes sequential experiences into coherent spatiotemporal contexts. In this section, we briefly discuss the putative neurobiological implementation of the proposed architecture (see Figure 5); while also noting that most aspects of this neurobiological architecture are speculative and demand future investigations.

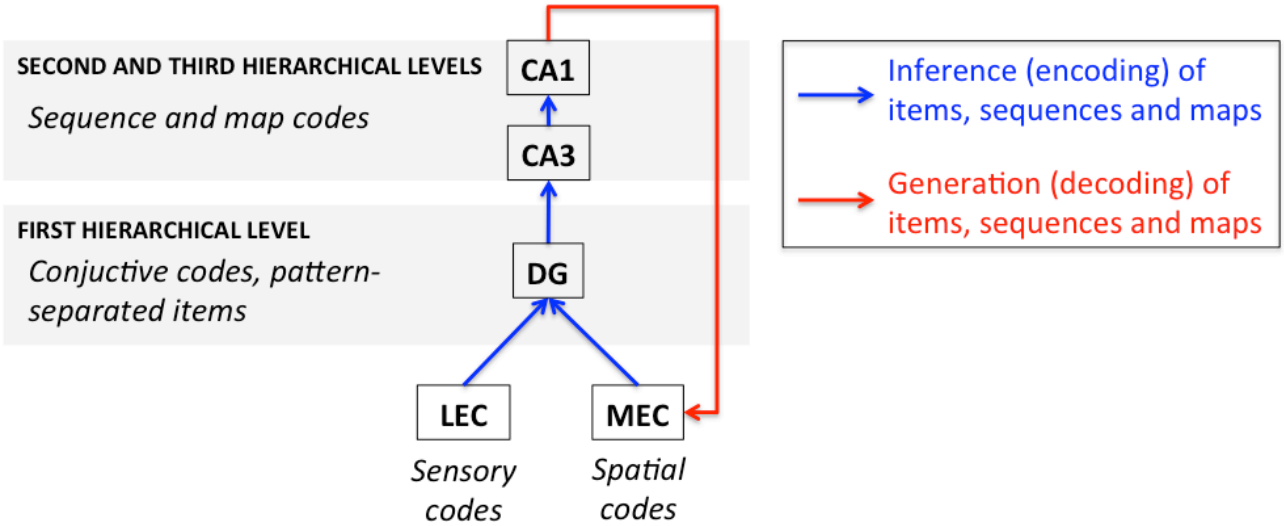

Figure 5. Putative neurobiological implementation of the hierarchical generative model (note that for space limitations, the figure only shows the EC-DG-CA3-CA1 system). LEC: lateral entorhinal cortex; MEC: medial entorhinal cortex; DG: dentate gyrus; CA: Cornu Ammonis.

Figure 5 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 illustrative purposes. 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 *inference* (encoding) and *generation* (decoding) processes implemented by the hierarchical architecture.

The maps may be represented in a more distributed way in the hippocampal formation, with hippocampal (e.g., place) and entorhinal (e.g., grid) codes, possibly functioning as "basis functions" to encode such maps [46]. 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 directly to hippocampal and entorhinal structures. This would assign separate roles to the hippocampus and the entorhinal cortex, the former acting as a "sequence predictor" 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 [47]. 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 relational and spatial) information robustly coded in entorhinal cortex [48], [49].

Our model assumes that maps are structured representations that define a coherent context to organize stimuli, forming spatiotemporal codes for spatial navigation [50] but also possibly non-spatial codes for other tasks [51]–[54]. It is worth noting that we used a mixture model for map probabilities, which implies that only one map is active at each time during learning and inference. This may be a simplification (to be possibly relaxed in future work), but evidence of theta-paced flickering [55] indicates that place cell assemblies from different mazes are activated within different theta cycles – suggesting that one map at the time might be activated. Computational arguments support our hypothesis, too, given that mixture models like the one we propose here, or related architectures such as cloned HMM, tend to be resistant to catastrophic forgetting [56].

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 also in part influenced by cortical structures – including not just medial entorhinal cortex [57] but also mPFC [58]. Prefrontal structures may play at least two important roles in the proposed model. First, they may play a permissive role for hippocampal sequences to occur when needed, such as for example, during decision-making [41], hence mediating prospective (planning) and retrospective (learning) role of hippocampal replays [5], [16]–[19]. Second, the cross-talk between hippocampal and cortical structure may permit their coordinated functioning and (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 [18], potentially contributing to bias their content and reorganize experiences according to learned rules [14]. A theoretical possibility that remains to be tested in future research is whether during periods of coupled oscillations, hippocampal and cortical structures simultaneously replay from their respective generative models and train each other. This idea can apply to subcortical structures that strongly cross-talk with the hippocampus, such as the ventral striatum. It has been proposed that the hippocampus and the ventral striatum may jointly implement a behavioural controller for goal-directed spatial navigation [59], [60]; and generative replay in these structures can help training the model.

Discussion

Our novel theoretical proposal casts the hippocampal formation as a hierarchical generative model of spatiotemporal experiences. The proposed 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 (and also external, e.g., cortical models or controllers). This method prevents catastrophic forgetting without requiring the verbatim storage of (an unbounded number of) previous experiences, as in "experience replay" [20]–[22]. 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 memory not model). Furthermore, our simulations show that generative replay mechanisms are well suited to explain the flexibility of internally generated hippocampal sequences [8], [13], beyond the mere replay of previous experience. It remains to be fully established whether standard *generative replay*, its variant *prioritized generative replay* or a combination of both methods best explain the growing empirical evidence on internally generated hippocampal sequences. The proposal that neuromodulation influences replay expression [61] suggests that under different contexts (defined by the activity of different neuromodulators), the selection of different generative replay methods may occur; but this hypothesis remains to be systematically tested in future research.

We proposed that the hippocampal generative model is hierarchically organized, to form 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. This idea is compatible with the well-established roles of the hippocampus in cognitive map formation [62] (and relatedly, in relational tasks) and the spontaneous generation of sequential dynamics [63], possibly based at least in part on preconfigured hippocampal dynamics that are sequential in nature [10], [11]. Organizing experiences into sequences and maps may help forming a coherent spatiotemporal context, permitting the model to link to the same latent map sets of trajectories experienced across different trials [12]. Furthermore, starting the inference with a preconfigured set of maps (acquired from previous experiences) that can be later adapted when maze-specific observations are gathered may help explaining preplay phenomena [15] that are more puzzling to explain in associative learning frameworks.

Another way to consider the role of sequences and maps in our architecture is considering that they afford *hierarchical predictions*, with sequences affording short-term predictions and (biased) maps affording long-term, average predictions of future locations. Our view is then compatible with the idea tat the hippocampus learns a predictive representation [64] but proposes that it is hierarchical.

While we exemplified the functioning of the proposed model in the domain of spatial navigation, its architecture is generic and could learn coherent spatiotemporal contexts also for non-navigation experiences, providing that they have the appropriate statistical structure. Recent studies have shown that the hippocampus supports temporal and rate coding of discrete sequences of non-spatial events such as (in rodents) odours [65] and button locations [66] and (in humans) of arbitrary items [14]; see also [67], [68]. More broadly, novel findings imply the hippocampal formation in cognitive mapping in several domains of cognition [69]–[71]. 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 [71]. Theoretical arguments suggest that while the brain might use multiple generative models, 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 and may be specialized to learn sequences of arbitrary (pattern-separated) items[16]. Establishing the plausibility of these theoretical arguments is an objective for future research.

Previous theoretical and computational studies assigned a role to the hippocampus in supporting model-based, goal-directed navigation [16]–[18], [24], [34], [42], [72], [73]. Our theory is generally in agreement with these ideas but goes beyond them, by introducing the idea of hierarchical, structured representations for sequences and maps, and by modelling explicitly how these representations are acquired over time. Importantly, our theory focuses on the hippocampal formation as a generative model of spatiotemporal experiences, not (or not necessarily) a transition model or a behavioural controller for goal-directed navigation, as proposed in the above models. Yet it is important to consider that the proposed generative model includes the essential components to train external controllers offline using generative replay, to provide predictions to external controllers, or even to form a controller for goal-directed navigation, when coupled with external structures such as the ventral striatum. Therefore, our proposed model could support all the previously hypothesized (model-based) functions of the hippocampus – but the plausibility of these or other possibilities remains to be established.

Casting the hippocampal formation as a hierarchical generative model immediately reconciles its roles in multiple, apparently disconnected cognitive functions, including episodic and autobiographic memory [74], [75], imagination [76], [77] and cognitive-map-based (or goal-directed) spatial navigation [62], [78]. These functions have often been studied in isolation and remain challenging to reconcile within an integrated theoretical proposal.

The idea of generative replay helps reconciling the roles of the hippocampus episodic memory and imagination. On the one hand, it supports the rapid acquisition of multiple episodic memories and can be used to train both the hippocampal generative model and other systems, (e.g., cortical models or controllers). On the other hand, it supports the spontaneous generation of novel, unexperienced sequences that are structurally coherent with past experiences. The spontaneously generated events can be biased to include behaviourally relevant (e.g., goal) or salient locations [8], [38], [45] as these are sampled from biased maps that represent behavioural relevance and saliency probabilistically. The hierarchical structure permits organizing experiences into coherent spatiotemporal contexts, explaining why the hippocampal formation is a “sequence generator” [16], [63] and plays a key role in cognitive map formation and goal-directed spatial navigation (and relational inference outside purely spatial domains). 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. However, 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 implements *generative replay*: it can stochastically generate items and sequences starting from its observations and inferred map; and use the stochastically generated sequences for training itself.

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 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 $S \equiv \boldsymbol{\xi}(1), \dots, \boldsymbol{\xi}(T)$:

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

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 the sequence of observations can be grouped together (according to any of their common characteristics), as if they were drawn from a distribution that partitions the sequences into clusters. Each cluster corresponds to a different map and the model assumes that its current sequence of observations belong to only one map at a time. These assumptions are embodied in the following generative *mixture model*:

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

where K is the number of maps. Note that while here we set K as known and fixed, in the section on "Nonparametric Method" we discuss a non-parametric extension of the model, which automatically infers the best number of clusters given its observations.

The definition of the model in Equation (3) is well-posed once we specify the quantities $p(c=k)$ and $p_{\boldsymbol{\theta}}(\mathbf{y}|k)$. The former is the mixing probability of the kth cluster to be in the partition, whose distribution is characterized by the parameters $\boldsymbol{\Theta} \equiv \{\boldsymbol{\theta}_k\}_{k=1,..,K}$. Each parameter $\boldsymbol{\theta}_k = \left(\theta_{k_1}, \cdots, \theta_{k_i}, \cdots, \theta_{k_N}\right) \in \mathbb{D}$ represents a particular map and encodes the frequency of event occupancy of the $\mathbb{D}$ locations under the map k. In the proposed model, we assume that $\boldsymbol{z} \equiv \left(p(c=1), \dots, p(c=K)\right) \sim Cat(\boldsymbol{\pi})$ and $\boldsymbol{\theta}_k \sim Cat(\boldsymbol{\rho}_k)$ 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, $p_{\boldsymbol{\theta}}(\mathbf{y}|k) \equiv p(\mathbf{y}|\boldsymbol{\theta}_k)$ represents the likelihood that the kth map can account for the sequence $\mathbf{y}$ of observations, and can be written down as a deterministic function of the probabilistic parameters $\boldsymbol{\theta}_k$:

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

where $\mathrm{y}_i = \mathbf{y} \circ \mathbf{e}_i$, with the symbol "$\circ$" denoting the Hadamard product. Here, we made the simplifying assumption that observations are independent. By making the logarithm, Equation (4) becomes the scalar product between $\mathbf{y}$ and $\ln \boldsymbol{\theta}_k$:

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

Thus, Equations (4) and (5) effectively measure the similarity between the sequence of events and the structure of a 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 the 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 calculate the values of the hidden states of the computational model, we adopt a Bayesian dynamical inferential process. The hierarchical inference starts by estimating the posteriors of the mixing probabilities $\boldsymbol{z}(t) \equiv p(c|\mathbf{y}(t))$. At the first time step, $(t=1)$, $\boldsymbol{z}(0)$ is drawn from a categorical distribution, i.e., $\boldsymbol{z}(0) \sim Cat(\boldsymbol{\pi})$, with hyperparameters $\boldsymbol{\pi}$ that follow a Dirichlet

distribution, $\boldsymbol{\pi} \sim Dir(\boldsymbol{\alpha})$. At the end of the inferential process of each time step, the hyperparameters $\boldsymbol{\pi}$ are recursively adjusted, based on the accumulated evidence, hence permitting to improving the a-priori information available for the next trials; see the section on "Parameter learning" for details.

Differently, when $t > 1$, $\boldsymbol{z}(t)$ is given by the expression:

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

Here, τ is an exponential time constant, whose value was fixed to (the empirically found value of) $\tau = 0.9$ and $\boldsymbol{\Theta}$ is the whole set of the 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:

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

Equation (6) determines the mixing probabilities at the current time by integrating their past posteriors with the information on past sequence $\mathbf{y}(t-1)$ available from the lower level. We adopt a Maximum A Posteriori approach, to select which of the K maps maximizes posteriors; namely $\bar{k} = argmax(\boldsymbol{z}(t))$. At the end of the inferential process of each time step (for $t > 1$), the parameters $\boldsymbol{\rho}_{\bar{k}}$ that define the 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, prediction and update. We could interpret Equations (6 and 7) as the prediction of a set of K hidden states $c = 1, \dots, c = K$, conditioned on the previous sequence code $\mathbf{y}(t-1)$; but contrary to standard Bayesian filtering, the latter is another hidden 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 evaluations for 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.

The hierarchal structure shown in Figure 1A suggests that, at this point, one should update the sequence $\mathbf{y}$. Conversely, to simplify the generation of fictive information during replay (see below), our model estimates item codes $\mathbf{x}(t)$ before sequence codes. 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 (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 sequence code at the previous time step $\mathbf{y}(t-1)$, which is subtracted, thus implementing a form of inhibition of return. Thus, we define $\mathbf{x}(t)$ updating as:

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

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 velocity of the sequence $\mathbf{y}$.

After updating the item code $\mathbf{x}(t)$, the model generates a prediction for the current 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

Finally, the inferential process updates the sequence code $\mathbf{y}(t)$, using the following dynamic equation:

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

where δ is the decay coefficient already introduced in Equation (2). In practice, Equation (10) gradually creates a spatiotemporal representation of observation sequences, by adding the item code $\boldsymbol{\xi}(t)$ for 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 the previous section, we mentioned that another process coexists with the hierarchical inference: parameter learning. Every time the inference selects the $\bar{k}$th hidden map via "filtering", the related hyperparameter components $\pi_{\bar{k}}$ and $\boldsymbol{\rho}_{\bar{k}}$ are updated, thus permitting to carry over information across time steps. To improve hyperparameter learning, we adopted the following methods: (i) we used a decay constant γ to account for the volatility of information coded in the maps (in our simulation, this is a small decay factor, equal to 1%), (ii) we added the probability of $z_{\bar{k}}(t) \equiv p(c = \bar{k})$, to account for the confidence in choosing $\boldsymbol{\theta}_{\bar{k}}$ as best map, and (iii) we utilize the value $\pi_{\bar{k}}$ only when $t = 1$ and the value $\boldsymbol{\rho}_{\bar{k}}$ for every $t > 1$; in both cases, the updating is done at the end of the hierarchical inference.

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

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

Even if the value of $\pi_{\bar{k}}$ is updated at every time step, it is only used at the beginning of each trial (when $t = 1$), as a parameter of the categorical distribution from which $\boldsymbol{z}(0)$ is sampled. Therefore, Equation (10) embodies in the parameters $\boldsymbol{\pi}$ a-priori information (from previous trials) about the probability of each cluster.

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

$$\boldsymbol{\rho}_{\bar{k}} = (1-\gamma)\boldsymbol{\rho}_{\bar{k}} + z_{\bar{k}}(t)\mathbf{y}(t) \tag{12}$$

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

Nonparametric method

Our simulations above used a fixed number of mixture clusters K. Here we describe an extension of the previous model that automatically infers K. For this, we used an adaptation of the Chinese Restaurant Process (CRP) prior over clustering of observations that adapts model complexity to data by allowing creating novel clusters as it observes novel data [79], [80]. Note that the CRP prior was applied only during navigation, on true data, and not during replay.

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.

Here, 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 $\mathbf{y}$ is to be considered as novel, we exploited the acquired cluster structure knowledge, or maps $\boldsymbol{\theta}_k$. A representation $\mathbf{y}$ was considered as novel when the greatest likelihood $p(\mathbf{y}|\boldsymbol{\theta}_k)$ (see Formula 4) among all clusters currently in use was smaller than a threshold, $p_{thr} = 0.01$. Table occupancy corresponded to the parameters π_k of the mixing probabilities and the concentration parameter was $\alpha = 1$.

Acknowledgements. The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 820213, ThinkAhead); and the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific Grant Agreement No. 785907 (Human Brain Project SGA2). The GEFORCE Titan GPU card used for this research was donated by the NVIDIA Corp.

References

[1] M. McCloskey and N. J. Cohen, 'Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem', in *Psychology of Learning and Motivation*, vol. 24, G. H. Bower, Ed. Academic Press, 1989, pp. 109–165.

[2] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly, '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.*, vol. 102, no. 3, pp. 419–457, Jul. 1995.

[3] R. C. O'Reilly, R. Bhattacharyya, M. D. Howard, and N. Ketz, 'Complementary learning systems', *Cogn. Sci.*, vol. 38, no. 6, pp. 1229–1248, Aug. 2014, doi: 10.1111/j.1551-6709.2011.01214.x.

[4] G. Buzsáki, 'Hippocampal sharp wave-ripple: A cognitive biomarker for episodic

memory and planning', *Hippocampus*, vol. 25, no. 10, pp. 1073–1188, 2015.
[5] D. J. Foster, 'Replay comes of age', *Annu. Rev. Neurosci.*, vol. 40, pp. 581–602, 2017.
[6] S. P. Jadhav, C. Kemere, P. W. German, and L. M. Frank, 'Awake Hippocampal Sharp-Wave Ripples Support Spatial Memory', *Science*, vol. 336, no. 6087, pp. 1454–1458, Jun. 2012, doi: 10.1126/science.1217230.
[7] M. G. Mattar and N. D. Daw, 'Prioritized memory access explains planning and hippocampal replay', *bioRxiv*, p. 225664, Nov. 2017, doi: 10.1101/225664.
[8] B. E. Pfeiffer and D. J. Foster, 'Hippocampal place-cell sequences depict future paths to remembered goals', *Nature*, vol. 497, no. 7447, pp. 74–79, May 2013, doi: 10.1038/nature12112.
[9] F. Stella, P. Baracskay, J. O'Neill, and J. Csicsvari, 'Hippocampal Reactivation of Random Trajectories Resembling Brownian Diffusion', *Neuron*, vol. 102, no. 2, pp. 450-461.e7, Apr. 2019, doi: 10.1016/j.neuron.2019.01.052.
[10] K. Liu, J. Sibille, and G. Dragoi, 'Generative Predictive Codes by Multiplexed Hippocampal Neuronal Tuplets', *Neuron*, vol. 99, no. 6, pp. 1329-1341.e6, Sep. 2018, doi: 10.1016/j.neuron.2018.07.047.
[11] G. Dragoi and S. Tonegawa, 'Distinct preplay of multiple novel spatial experiences in the rat', *Proc. Natl. Acad. Sci.*, vol. 110, no. 22, pp. 9100–9105, 2013.
[12] A. S. Gupta, M. A. A. van der Meer, D. S. Touretzky, and A. D. Redish, 'Hippocampal replay is not a simple function of experience.', *Neuron*, vol. 65, no. 5, pp. 695–705, Mar. 2010, doi: 10.1016/j.neuron.2010.01.034.
[13] H. F. Ólafsdóttir, C. Barry, A. B. Saleem, D. Hassabis, and H. J. Spiers, 'Hippocampal place cells construct reward related sequences through unexplored space', *eLife*, vol. 4, p. e06063, Jun. 2015, doi: 10.7554/eLife.06063.
[14] Y. Liu, R. J. Dolan, Z. Kurth-Nelson, and T. E. J. Behrens, 'Human Replay Spontaneously Reorganizes Experience', *Cell*, Jul. 2019, doi: 10.1016/j.cell.2019.06.012.
[15] G. Dragoi and S. Tonegawa, 'Preplay of future place cell sequences by hippocampal cellular assemblies', *Nature*, vol. 469, no. 7330, pp. 397–401, 2011.
[16] G. Pezzulo, C. Kemere, and M. van der Meer, 'Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition', *Ann. N. Y. Acad. Sci.*, vol. 1396, pp. 144–165, 2017.
[17] G. Pezzulo, M. A. A. van der Meer, C. S. Lansink, and C. M. A. Pennartz, 'Internally generated sequences in learning and executing goal-directed behavior', *Trends Cogn. Sci.*, vol. 18, no. 12, pp. 647–657, 2014, doi: 10.1016/j.tics.2014.06.011.
[18] H. Penagos, C. Varela, and M. A. Wilson, 'Oscillations, neural computations and learning during wake and sleep', *Curr. Opin. Neurobiol.*, vol. 44, pp. 193–201, Jun. 2017, doi: 10.1016/j.conb.2017.05.009.
[19] G. Pezzulo, F. Donnarumma, D. Maisto, and I. Stoianov, 'Planning at decision time and in the background during spatial navigation', *Curr. Opin. Behav. Sci.*, vol. 29, pp. 69–76, Oct. 2019, doi: 10.1016/j.cobeha.2019.04.009.
[20] H. Shin, J. K. Lee, J. Kim, and J. Kim, 'Continual Learning with Deep Generative Replay', *ArXiv170508690 Cs*, May 2017.
[21] D. C. Mocanu, M. T. Vega, E. Eaton, P. Stone, and A. Liotta, 'Online Contrastive Divergence with Generative Replay: Experience Replay without Storing Data', *ArXiv161005555 Cs*, Oct. 2016.
[22] G. M. van de Ven and A. S. Tolias, 'Generative replay with feedback connections as a general strategy for continual learning', *ArXiv Prepr. ArXiv180910635*, 2018.
[23] R. S. Sutton, 'Integrated modeling and control based on reinforcement learning and dynamic programming', in *Advances in Neural Information Processing Systems*, Morgan-Kaufmann., vol. 3, D. S. Touretzky, Ed. 1991, pp. 471–478.

[24] I. P. Stoianov, C. M. A. Pennartz, C. S. Lansink, and G. Pezzulo, 'Model-based spatial navigation in the hippocampus-ventral striatum circuit: A computational analysis', *PLOS Comput. Biol.*, vol. 14, no. 9, p. e1006316, Sep. 2018, doi: 10.1371/journal.pcbi.1006316.
[25] R. P. Rao and D. H. Ballard, 'Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects', *Nat Neurosci*, vol. 2, no. 1, pp. 79–87, Jan. 1999, doi: 10.1038/4580.
[26] K. J. Friston, 'A theory of cortical responses.', *Philos Trans R Soc Lond B Biol Sci*, vol. 360, no. 1456, pp. 815–836, Apr. 2005, doi: 10.1098/rstb.2005.1622.
[27] Y. Aoki, H. Igata, Y. Ikegaya, and T. Sasaki, 'The Integration of Goal-Directed Signals onto Spatial Maps of Hippocampal Place Cells', *Cell Rep.*, vol. 27, no. 5, pp. 1516-1527.e5, Apr. 2019, doi: 10.1016/j.celrep.2019.04.002.
[28] F. Donnarumma, D. Maisto, and G. Pezzulo, 'Problem Solving as Probabilistic Inference with Subgoaling: Explaining Human Successes and Pitfalls in the Tower of Hanoi', *PLOS Comput. Biol.*, vol. 12, no. 4, p. e1004864, Apr. 2016, doi: 10.1371/journal.pcbi.1004864.
[29] D. Maisto, F. Donnarumma, and G. Pezzulo, 'Nonparametric Problem-Space Clustering: Learning Efficient Codes for Cognitive Control Tasks', *Entropy*, vol. 18, no. 2, p. 61, Feb. 2016, doi: 10.3390/e18020061.
[30] D. Maisto, F. Donnarumma, and G. Pezzulo, 'Divide et impera: subgoaling reduces the complexity of probabilistic inference and problem solving', *J. R. Soc. Interface*, vol. 12, no. 104, p. 20141335, Mar. 2015, doi: 10.1098/rsif.2014.1335.
[31] S. J. Gershman, 'The Successor Representation: Its Computational Logic and Neural Substrates', *J. Neurosci.*, vol. 38, no. 33, pp. 7193–7200, Aug. 2018, doi: 10.1523/JNEUROSCI.0151-18.2018.
[32] A. W. Moore and C. G. Atkeson, 'Prioritized sweeping: Reinforcement learning with less data and less time', *Mach. Learn.*, vol. 13, no. 1, pp. 103–130, Oct. 1993, doi: 10.1007/BF00993104.
[33] M. C. Fuhs and D. S. Touretzky, 'Context learning in the rodent hippocampus', *Neural Comput.*, vol. 19, no. 12, pp. 3173–3215, 2007.
[34] W. D. Penny, P. Zeidman, and N. Burgess, 'Forward and Backward Inference in Spatial Cognition', *PLoS Comput Biol*, vol. 9, no. 12, p. e1003383, Dec. 2013, doi: 10.1371/journal.pcbi.1003383.
[35] H. Sanders, M. A. Wilson, and S. J. Gershman, 'Hippocampal Remapping as Hidden State Inference', Center for Brains, Minds and Machines (CBMM), bioRxiv, 2019.
[36] R. M. Grieves, E. R. Wood, and P. A. Dudchenko, 'Place cells on a maze encode routes rather than destinations', *eLife*, vol. 5, 10 2016, doi: 10.7554/eLife.15986.
[37] Y. Liu, Z. Kurth-Nelson, and T. Behrens, 'Non-spatial neural replay in building and updating world models in humans', in *Cosyne 2018*, 2018.
[38] R. E. Ambrose, B. E. Pfeiffer, and D. J. Foster, 'Reverse Replay of Hippocampal Place Cells Is Uniquely Modulated by Changing Reward', *Neuron*, vol. 91, no. 5, pp. 1124–1136, Sep. 2016, doi: 10.1016/j.neuron.2016.07.047.
[39] D. J. Foster and M. A. Wilson, 'Hippocampal theta sequences', *Hippocampus*, vol. 17, no. 11, pp. 1093–1099, 2007, doi: 10.1002/hipo.20345.
[40] A. Johnson and A. D. Redish, 'Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point.', *J Neurosci*, vol. 27, no. 45, pp. 12176–12189, Nov. 2007, doi: 10.1523/JNEUROSCI.3761-07.2007.
[41] A. D. Redish, 'Vicarious trial and error', *Nat. Rev. Neurosci.*, vol. 17, no. 3, pp. 147–159, Mar. 2016, doi: 10.1038/nrn.2015.30.
[42] J. Lisman and A. D. Redish, 'Prediction, Sequences and the Hippocampus', *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 364, no. 1521, pp. 1193–1201, Dec. 2009, doi: 10.1098/rstb.2008.0316.
[43] H. Sanders, C. Rennó-Costa, M. Idiart, and J. Lisman, 'Grid Cells and Place Cells: An

Integrated View of their Navigational and Memory Function', *Trends Neurosci.*, vol. 38, no. 12, pp. 763–775, Dec. 2015, doi: 10.1016/j.tins.2015.10.004.
[44] H. Shimazaki, 'Neurons as an Information-theoretic Engine', *ArXiv151207855 Phys. Q-Bio*, 2018, doi: 10.1007/978-3-319-71976-4.
[45] A. M. Wikenheiser and A. D. Redish, 'Hippocampal theta sequences reflect current goals', *Nat. Neurosci.*, vol. 18, no. 2, pp. 289–294, Feb. 2015, doi: 10.1038/nn.3909.
[46] J. C. Whittington, T. H. Muller, S. Mark, C. Barry, N. Burgess, and T. E. Behrens, 'The Tolman-Eichenbaum Machine: Unifying space and relational memory through generalisation in the hippocampal formation', *bioRxiv*, p. 770495, Sep. 2019, doi: 10.1101/770495.
[47] J. Whittington, T. Muller, S. Mark, C. Barry, and T. Behrens, 'Generalisation of structural knowledge in the hippocampal-entorhinal system', in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Curran Associates, Inc., 2018, pp. 8484–8495.
[48] C. N. Boccara, M. Nardin, F. Stella, J. O'Neill, and J. Csicsvari, 'The entorhinal cognitive map is attracted to goals', *Science*, vol. 363, no. 6434, pp. 1443–1447, 29 2019, doi: 10.1126/science.aav4837.
[49] W. N. Butler, K. Hardcastle, and L. M. Giocomo, 'Remembered reward locations restructure entorhinal spatial maps', *Science*, vol. 363, no. 6434, pp. 1447–1452, 29 2019, doi: 10.1126/science.aav5297.
[50] M. H. Plitt and L. M. Giocomo, 'Experience dependent contextual codes in the hippocampus', *bioRxiv*, p. 864090, Dec. 2019, doi: 10.1101/864090.
[51] D. Aronov, R. Nevers, and D. W. Tank, 'Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit', *Nature*, vol. 543, no. 7647, pp. 719–722, Mar. 2017, doi: 10.1038/nature21692.
[52] A. O. Constantinescu, J. X. O'Reilly, and T. E. J. Behrens, 'Organizing conceptual knowledge in humans with a gridlike code', *Science*, vol. 352, no. 6292, pp. 1464–1468, Jun. 2016, doi: 10.1126/science.aaf0941.
[53] R. M. Mok and B. C. Love, 'A non-spatial account of place and grid cells based on clustering models of concept learning', *Nat. Commun.*, vol. 10, no. 1, pp. 1–9, Dec. 2019, doi: 10.1038/s41467-019-13760-8.
[54] R. A. Gulli *et al.*, 'Context-dependent representations of objects and space in the primate hippocampus during virtual navigation', *Nat. Neurosci.*, vol. 23, no. 1, pp. 103–112, Jan. 2020, doi: 10.1038/s41593-019-0548-3.
[55] K. Jezek, E. J. Henriksen, A. Treves, E. I. Moser, and M.-B. Moser, 'Theta-paced flickering between place-cell maps in the hippocampus', *Nature*, vol. 478, no. 7368, pp. 246–249, Oct. 2011, doi: 10.1038/nature10439.
[56] R. V. Rikhye, N. Gothoskar, J. S. Guntupalli, A. Dedieu, M. Lázaro-Gredilla, and D. George, 'Learning cognitive maps for vicarious evaluation', *bioRxiv*, p. 864421, Dec. 2019, doi: 10.1101/864421.
[57] A. Chenani, M. Sabariego, M. I. Schlesiger, J. K. Leutgeb, S. Leutgeb, and C. Leibold, 'Hippocampal CA1 replay becomes less prominent but more rigid without inputs from medial entorhinal cortex', *Nat. Commun.*, vol. 10, no. 1, p. 1341, 2019.
[58] J. D. Shin, W. Tang, and S. P. Jadhav, 'Dynamics of Awake Hippocampal-Prefrontal Replay for Spatial Learning and Memory-Guided Decision Making', *Neuron*, vol. 104, no. 6, pp. 1110-1125.e7, Dec. 2019, doi: 10.1016/j.neuron.2019.09.012.
[59] G. Pezzulo, M. A. A. van der Meer, C. S. Lansink, and C. M. A. Pennartz, 'Internally generated sequences in learning and executing goal-directed behavior', *Trends Cogn. Sci.*, vol. 18, no. 12, pp. 647–657, 2014, doi: 10.1016/j.tics.2014.06.011.
[60] M. van der Meer, Z. Kurth-Nelson, and A. D. Redish, 'Information processing in decision-making systems', *The Neuroscientist*, vol. 18, no. 4, pp. 342–359, 2012.

[61] L. A. Atherton, D. Dupret, and J. R. Mellor, 'Memory trace replay: the shaping of memory consolidation by neuromodulation', *Trends Neurosci.*, vol. 38, no. 9, pp. 560–570, Sep. 2015, doi: 10.1016/j.tins.2015.07.004.
[62] E. C. Tolman, 'Cognitive maps in rats and men', *Psychol. Rev.*, vol. 55, pp. 189–208, 1948.
[63] G. Buzsáki and D. Tingley, 'Space and Time: The Hippocampus as a Sequence Generator', *Trends Cogn. Sci.*, vol. 22, no. 10, pp. 853–869, Oct. 2018, doi: 10.1016/j.tics.2018.07.006.
[64] S. Recanatesi, M. Farrell, G. Lajoie, S. Deneve, M. Rigotti, and E. Shea-Brown, 'Signatures and mechanisms of low-dimensional neural predictive manifolds', *bioRxiv*, p. 471987, Dec. 2018, doi: 10.1101/471987.
[65] S. Terada, Y. Sakurai, H. Nakahara, and S. Fujisawa, 'Temporal and Rate Coding for Discrete Event Sequences in the Hippocampus', *Neuron*, vol. 94, no. 6, pp. 1248-1262.e4, Jun. 2017, doi: 10.1016/j.neuron.2017.05.024.
[66] K. J. Miller, M. M. Botvinick, and C. D. Brody, 'Dorsal hippocampus contributes to model-based planning', *Nat. Neurosci.*, vol. 20, no. 9, pp. 1269–1276, Sep. 2017, doi: 10.1038/nn.4613.
[67] N. W. Schuck and Y. Niv, 'Sequential replay of nonspatial task states in the human hippocampus', *Science*, vol. 364, no. 6447, Jun. 2019, doi: 10.1126/science.aaw5181.
[68] B. Shahbaba *et al.*, 'Hippocampal ensembles represent sequential relationships among discrete nonspatial events', *bioRxiv*, p. 840199, Nov. 2019, doi: 10.1101/840199.
[69] T. E. J. Behrens *et al.*, 'What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior', *Neuron*, vol. 100, no. 2, pp. 490–509, Oct. 2018, doi: 10.1016/j.neuron.2018.10.002.
[70] J. L. S. Bellmund, P. Gärdenfors, E. I. Moser, and C. F. Doeller, 'Navigating cognition: Spatial codes for human thinking', *Science*, vol. 362, no. 6415, p. eaat6766, Nov. 2018, doi: 10.1126/science.aat6766.
[71] G. Buzsáki and E. I. Moser, 'Memory, navigation and theta rhythm in the hippocampal-entorhinal system', *Nat. Neurosci.*, vol. 16, no. 2, pp. 130–138, Feb. 2013, doi: 10.1038/nn.3304.
[72] G. Pezzulo, F. Rigoli, and F. Chersi, 'A mixed instrumental controller can combine habitual and goal-directed choice', in *Proceedings of CogSci 2012*, 2012.
[73] M. van der Meer, Z. Kurth-Nelson, and A. D. Redish, 'Information processing in decision-making systems', *The Neuroscientist*, vol. 18, no. 4, pp. 342–359, 2012.
[74] E. Tulving, 'Episodic Memory: From Mind to Brain', *Annu. Rev. Psychol.*, vol. 53, no. 1, pp. 1–25, 2002, doi: 10.1146/annurev.psych.53.100901.135114.
[75] H. Eichenbaum, 'A cortical-hippocampal system for declarative memory', *Nat. Rev. Neurosci.*, vol. 1, no. 1, pp. 41–50, Oct. 2000, doi: 10.1038/35036213.
[76] R. L. Buckner, 'The role of the hippocampus in prediction and imagination.', *Annu Rev Psychol*, vol. 61, pp. 27–48, C1-8, 2010, doi: 10.1146/annurev.psych.60.110707.163508.
[77] D. L. Schacter, D. R. Addis, D. Hassabis, V. C. Martin, R. N. Spreng, and K. K. Szpunar, 'The Future of Memory: Remembering, Imagining, and the Brain', *Neuron*, vol. 76, no. 4, pp. 677–694, Nov. 2012, doi: 10.1016/j.neuron.2012.11.001.
[78] J. O'keefe and L. Nadel, *The hippocampus as a cognitive map*, vol. 3. Clarendon Press Oxford, 1978.
[79] I. Stoianov, A. Genovesio, and G. Pezzulo, 'Prefrontal Goal Codes Emerge as Latent States in Probabilistic Value Learning', *J. Cogn. Neurosci.*, 2015.
[80] S. J. Gershman and D. M. Blei, 'A tutorial on Bayesian nonparametric models', *J. Math. Psychol.*, vol. 56, no. 1, pp. 1–12, 2012.