The hippocampus as a predictive map

Kimberly L. Stachenfeld[1,2,*], **Matthew M. Botvinick**[1,3], **and Samuel J. Gershman**[4]

[1]DeepMind, London, UK
[2]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA
[3]Gatsby Computational Neuroscience Unit, University College London, London, UK
[4]Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA, USA
[*]stachenfeld@google.com

ABSTRACT

A cognitive map has long been the dominant metaphor for hippocampal function, embracing the idea that place cells encode a geometric representation of space. However, evidence for predictive coding, reward sensitivity, and policy dependence in place cells suggests that the representation is not purely spatial. We approach this puzzle from a reinforcement learning perspective: what kind of spatial representation is most useful for maximizing future reward? We show that the answer takes the form of a predictive representation. This representation captures many aspects of place cell responses that fall outside the traditional view of a cognitive map. Furthermore, we argue that entorhinal grid cells encode a low-dimensional basis set for the predictive representation, useful for suppressing noise in predictions and extracting multiscale structure for hierarchical planning.

Introduction

Learning to predict long-term reward is fundamental to the survival of many animals. Some species may go days, weeks or even months before attaining primary reward, during which time aversive states must be endured. Evidence suggests that the brain has evolved multiple solutions to this reinforcement learning (RL) problem[1]. One solution is to learn a model or "cognitive map" of the environment[2], which can then be used to generate long-term reward predictions through simulation of future states[1]. However, this solution is computationally intensive, especially in real-world environments where the space of future possibilities is virtually infinite. An alternative "model-free" solution is to learn, from trial-and-error, a value function mapping states to long-term reward predictions[3]. However, dynamic environments can be problematic for this approach, because changes in the distribution of rewards necessitates complete relearning of the value function.

Here, we argue that the hippocampus supports a third solution: learning of a "predictive map" that represents each state in terms of its successor states[4,5]. This representation is sufficient for long-term reward prediction, is learnable using a simple, biologically plausible algorithm, and explains a wealth of data from studies of the hippocampus.

Our primary focus is on understanding the computational function of hippocampal place cells, which respond selectively when an animal occupies a particular location in space[6]. A classic and still influential view of place cells is that they collectively furnish an explicit map of space[7,8]. This map can then be employed as the input to a model-based[9–11] or model-free[12,13] RL system for computing the value of the animal's current state. In contrast, the predictive map theory views place cells as encoding predictions of future states, which can then be combined with reward predictions to compute values. This theory can account for why the firing of place cells is modulated by variables like obstacles, environment topology, and direction of travel. It also generalizes to hippocampal coding in non-spatial tasks. Beyond

the hippocampus, we argue that entorhinal grid cells[14], which fire periodically over space, encode a low-dimensional decomposition of the predictive map, useful for stabilizing the map and discovering subgoals.

Results

The successor representation

We consider the problem of RL in a Markov decision process consisting of the following elements[15]: a set of states (e.g., spatial locations), a set of actions, a transition distribution $P(s'|s,a)$ specifying the probability of transitioning to state s' from state s after taking action a, a reward function $R(s)$ specifying the expected immediate reward in state s, and a discount factor $\gamma \in [0,1]$ that down-weights distal rewards. An agent chooses actions according to a policy $\pi(a|s)$ and collects rewards as it moves through the state space. The value of a state is defined formally as the expected discounted cumulative future reward under policy π:

$$V(s) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \gamma^t R(s_t) \mid s_0 = s \right], \tag{1}$$

where s_t is the state visited at time t. Our focus here is on policy evaluation (computing V). In our simulations we feed the agent the optimal policy; in the Supplemental Methods we discuss algorithms for policy improvement. To simplify notation, we assume implicit dependence on π and define the state transition matrix T, where $T(s,s') = \sum_a \pi(a|s) P(s'|s,a)$.

The value function can be decomposed into the inner product of the reward function with a predictive state representation known as the successor representation (SR)[4], denoted by M:

$$V(s) = \sum_{s'} M(s,s') R(s'), \tag{2}$$

The SR encodes the expected discounted future occupancy of state s' along a trajectory initiated in state s:

$$M(s,s') = \mathbb{E}\left[\textstyle\sum_{t=0}^{\infty} \gamma^t \mathbb{I}(s_t = s') \mid s_0 = s \right], \tag{3}$$

where $\mathbb{I}(\cdot) = 1$ if its argument is true, and 0 otherwise.

An estimate of the SR (denoted $\hat{M}$) can be incrementally updated using a form of the temporal difference learning algorithm[4,16]. After observing a transition $s_t \to s_{t+1}$, the estimate is updated according to:

$$\hat{M}_{t+1}(s,s') = \hat{M}_t(s_t,s') + \eta \left[\mathbb{I}(s_t = s') + \gamma \hat{M}_t(s_{t+1},s') - \hat{M}_t(s_t,s') \right], \tag{4}$$

where η is a learning rate (unless specified otherwise, $\eta = 0.1$ in our simulations). The form of this update is identical to the temporal difference learning rule for value functions[15], except that in this case the reward prediction error is replaced by a *successor prediction error* (the term in brackets). Note that these prediction errors are distinct from state prediction errors used to update an estimate of the transition function[17]; the SR predicts not just the next state but a superposition of future states over a possibly infinite horizon. The transition and SR functions only coincide when $\gamma = 0$.

The SR combines some of the advantages of model-free and model-based algorithms. Like model-free algorithms, policy evaluation is computationally efficient with the SR. However, factoring the value function into a state dynamics SR term and a reward term confers some of the flexibility usually associated

with model-based methods. Separate terms for state dynamics and reward permit rapid recomputation of new value functions when reward is changed independently of state dynamics, as demonstrated in Fig. 1. As the animal toggles between hunger and thirst, or when food is redistributed about the environment, the animal can immediately recompute a new value function based on its expected state transitions. A model-free agent would have to relearn value estimates for each location in order to make value predictions, while a model-based agent would need to aggregate the results of time-consuming searches through its model[1,4]. This advantage is demonstrated in Fig. S1, in which we demonstrate that while changing the reward function completely disrupts model free learning of a value function in a 2-step tree maze, SR learning can quickly adjust. Thus, the SR combines the efficiency of model-free control with some of the flexibility of model-based control.

For an agent trying to optimize expected discounted future reward, two states that predict similar successor states are necessarily similarly valuable, and can be safely grouped together[18]. This makes the SR a good metric space for generalizing value. Since adjacent states will frequently lead to each other, the SR will naturally represent adjacent states similarly and therefore be smooth over time and space in spatial tasks. Since the SR is well defined for any Markov decision process, we can use the same architecture for many kinds of tasks, not just spatial ones.

Hippocampal encoding of the successor representation

We now turn to our main theoretical claim: that the SR is encoded by the hippocampus. This hypothesis is Based on the central role of the hippocampus in representing space and context[19], as well as its contribution to sequential decision making[20,21]. Although the SR can be applied to arbitrary state spaces, we focus on spatial domains where states index locations.

Place cells in the hippocampus have traditionally been viewed as encoding an animal's current location. In contrast, the predictive map theory views these cells as encoding an animal's *future* locations. Crucially, an animal's future locations depend on its policy, which is constrained by a variety of factors such as the environmental topology and the locations of rewards. We demonstrate that these factors shape place cell receptive field properties in a manner consistent with a predictive map.

To build some intuition for this idea, Fig. 2 illustrates the differences between our SR model (Fig. 2C) and two alternative models (Fig. 2A-B). As examples, we implement the three models for a 2D room containing an obstacle and for a 1D track with an established preferred direction of travel. The first alternative model is a Gaussian place field in which firing is related to the Euclidean distance from the field center (Fig. 2A), usually invoked for modelling place field activity in open spatial domains[22,23]. The second alternative model is a topologically sensitive place field in which firing is related to the average path length from the field center, where paths cannot pass through obstacles[13] (Fig. 2A). Like the topological place fields and unlike the Gaussian place fields, the SR place fields respect obstacles in the 2D environment. Since states on opposite sides of a barrier do not frequently occur nearby in time, SR place fields will tend to be active on only one side of a barrier.

On the 1D track, the SR place fields skew opposite the direction of travel. This backward skewing arises because upcoming states can be reliably predicted further in advance when traveling repeatedly in a particular direction. Neither of the control models provide ways for a directed behavioral policy to interact with state representation, and therefore cannot show this effect. Evidence for predictive skewing comes from experiments in which animals traveled repeatedly in a particular direction along a linear track[24,25]. In Fig. 3, we explain how a future-oriented representation evokes a forward-skewing representation in the population at any given point in time but implies that receptive fields for any individual cell should skew backwards. In order for a given cell to fire predictively, it must begin firing before its encoded state is visited, causing a backward-skewed receptive field. Figure 4 compares the predicted and experimentally

observed backward skewing, demonstrating that the model captures the qualitative pattern of skewing observed when the animal has a directional bias.

Consistent with the SR model, experiments have shown that place fields become distorted around barriers[26–28]. In Figure 5, we explore the effect of placing obstacles in a Tolman detour maze on the SR place fields and compare to experimental results obtained by Alvernhe *et al.*[28]. When a barrier is placed in a maze such that the animal is forced to take a detour, the place fields engage in "local remapping." Place fields near the barrier change their firing fields significantly more than those further from the barrier (Fig. 5A-C). When barriers are inserted, SR place fields change their fields near the path blocked by the barrier and less so at more distal locations where the optimal policy is unaffected (Fig. 5D-F). This locality is imposed by the discount factor.

The SR model can be used to explain how hippocampal place field clustering depends on factors such as reward, punishment, and boundaries (Fig. 6). The center of mass of place fields near the wall under a random walk policy will retreat away from the walls of a rectangular room (Fig. 6A). For any model of place cell firing, the centers of mass will be biased away from boundaries, since no firing will be recorded outside of the room[26]. However, this bias should be more pronounced under the SR model because successor states will also necessarily lie within the environment, shifting the fields asymmetrically away from walls. In fragmented, multi-compartment environments, this has the effect of causing place fields within the same compartment to cluster (Fig. 6B,C). We explore the implications of this clustering in more detail toward the end of this section, accompanied by the illustration in Fig. S2.

The SR model predicts that firing fields centered near rewarded locations will expand to include the surrounding locations and increase their firing rate under the optimal policy, as has been observed experimentally[29,30]. The animal is likely to spend time in the vicinity of the reward, meaning that states with or near reward are likely to be common successors. SR place fields in and near the rewarded zone will cluster because it is likely that states near the reward were anticipated by other states near the reward (Fig. 6D,E). For place fields centered further from the reward, the model predicts that fields will skew opposite the direction of travel toward the reward, due to the effect illustrated in Fig. 3: a state will only be predicted when the animal is approaching reward from some more distant state. Given a large potentially rewarded zone or a noisy policy, these somewhat contradictory effects are sufficient to produce clustering of place fields near the rewarded zone, as has been observed experimentally in certain tasks[31,32] (Fig. 6D,E), We predict that punished locations will induce the opposite effect, causing fields near the punished location to spread away from the rarely-visited punished locations (Fig. 6F).

In addition to the influence of experimental factors, changes in parameters of the model will have systematic effects on the structure of SR place fields. Motivated by data showing a gradient of increasing field sizes along the hippocampal longitudinal axis[33,34], we explored the consequences of modifying the discount factor γ in Figure S2. Hosting a range of discount factors along the hippocampal longitudinal axis provides a multi-timescale representation of space. It also circumvents the problem of having to assume the same discount parameter for each problem or adaptively computing a new discount. Another consequence is that larger place fields reflect the community structure of the environment. A gradient of discount factors might therefore be useful for decision making at multiple levels of temporal abstraction[18,35,36].

An appealing property of the SR model is that it can be applied to non-spatial state spaces. Fig. 7A-D shows the SR embedding of an abstract state space used in a study by Schapiro and colleagues[18,37]. Human subjects viewed sequences of fractals drawn from random walks on the graph while brain activity was measured using fMRI. We compared the similarity between SR vectors for pairs of states with pattern similarity in the hippocampus. The key experimental finding was that hippocampal pattern similarity mirrored the community structure of the graph: states with similar successors were represented similarly[37]. The SR model recapitulates this finding, since states in the same community tend to be visited nearby in

time, making them predictive of one another (Fig. 7E-G).

To demonstrate further how the SR model can integrate spatial and temporal coding in the hippocampus, we simulated results from a recent study[38] in which subjects were asked to navigate among pairs of locations to retrieve associated objects in a virtual city (8A). Since it was possible to "teleport" between certain location pairs, while others were joined only by long, winding paths, spatial Euclidean distance was decoupled from travel time. The authors found that objects associated with locations that were nearby in either space or time increased their hippocampal pattern similarity (Fig. 8B). Both factors (spatial and temporal distance) had a significant effect when the other was regressed out (Fig. 8C). The SR predicts this integrated representation of spatiotemporal distance: when a short path is introduced between distant states, such as by a teleportation hub, those states come predict one another.

Dimensionality reduction of the predictive map by entorhinal grid cells

Because the firing fields of entorhinal grid cells are spatially periodic, it was originally hypothesized that grid cells might represent a Euclidean spatial metric to enable dead reckoning[8,14]. Other theories have suggested that these firing patterns might arise from a low-dimensional embedding of the hippocampal map[5,23,39]. Combining this idea with the SR hypothesis, we argue that grid fields reflect a low-dimensional eigendecomposition of the SR. A key implication of this hypothesis is that grid cells will respond differently in environments with different boundary conditions.

The boundary sensitivity of grid cells was recently highlighted by a study that manipulated boundary geometry[40]. In square environments, different grid modules had the same alignment of the grid relative to the boundaries (modulo 60°, likely due to hexagonal symmetry in grid fields), whereas in a circular environment grid field alignment was more variable, with a qualitatively different pattern of alignment (Fig. 9A-C). Krupic *et al.* performed a "split-halves" analysis, in which they compared grid fields in square versus trapezoidal mazes, to examine the effect of breaking an axis of symmetry in the environment (Fig 9D,E). They found that moving the animal to a trapezoidal environment, in which the left and right half of the environment had asymmetric boundaries, caused the grid parameters to be different on the two sides of the environment[40]. In particular, the spatial autocorrelegrams – which reveal the layout of spatial displacement at which the grid field repeats itself – were relatively dissimilar over both halves of the trapezoidal environment. The grid fields in the trapezoid could not be attributed to linearly warping the square grid field into a trapezoid, raising the question of how else boundaries could interact with grid fields.

According to the SR eigenvector model, these effects arise because the underlying statistics of the transition policy changes with the geometry. We simulated grid fields in a variety of geometric environments used by Krupic and colleagues (Fig. 9F-H). In agreement with the empirical results, the orientation of eigenvectors in the circular environment tend to be highly variable, while those recorded in square environments are almost always aligned to either the horizontal or vertical boundary of the square (Fig. 9G,J). The variability in the circular environment arises because the eigenvectors are subject to the rotational symmetry of the circular task space. SR eigenvectors also emulate the finding that grids on either side of a square maze are more similar than those on either side of a trapezoid, because the eigenvectors capture the effect of these irregular boundary conditions on transition dynamics.

Another main finding of Krupic et al.[40] was that when a square environment is rotated, grids remain aligned to the boundaries as opposed to distal cues. SR eigenvectors inherently reproduce this effect, since a core assumption of the theory is that grid firing is anchored to state in a transition structure, which is itself constrained by boundaries.

A different manifestation of boundary effects is the fragmentation of grid fields in a hairpin maze[41]. Consistent with the empirical data, SR eigenvector fields tend to align with the arms of the maze, and

frequently repeat across alternating arms (Figure 10)[41]. While patterns at many timescales can be found in the eigenvector population, those at alternating intervals are most common and therefore replicate the checkerboard pattern observed in the experimental data (Fig. S8).

To further explore how compartmentalized environments could affect grid fields, we simulated a recent study[42] that characterized how grid fields evolve over several days' exposure to a multi-compartment environment (Fig. 11). While grid cells initially represented separate compartments with identical fields (repeated grids), several days of exploration caused fields to converge on a more globally coherent grid (Fig. 11D-F). With more experience, the grid regularity of the fields simultaneously decreased, as did the similarity between the grid fields recorded in the two rooms (Fig. 11C). The authors conclude that grid cells will tend to a regular, globally coherent grid to serve as a Euclidean metric over the full expanse of the enclosure.

Our model suggests that the fields are tending not toward a globally *regular* grid, but to a predictive map of the task structure, which is shaped in part by the global boundaries but also by the multi-compartment structure. We simulated this experiment by initializing grid fields to a local eigenvector model, in which the animal has not yet learned how the compartments fit together. After the SR eigenvectors have been learned, we relax the constraint that representations be the same in both rooms and let eigenvectors and the SR be learned for the full environment. As the learned eigenvectors converge, they increasingly resemble a global grid and decreasingly match the predictions of the local fit (Fig. 11H-L; Fig. S10). As with the recorded grid cells, the similarity of the fields in the two rooms drops to an average value near zero (Fig. 11I). They also have less regular grids compared to a single-compartment rectangular enclosure, explaining the drop in grid regularity observed by Carpenter *et al.* as the grid fields became more "global"[42]. Since separating barriers between compartments perturb the task topology from an uninterrupted 2D grid.

A normative motivation for invoking low-dimensional projections as a principle for grid cells is that they can be used to smooth or "regularize" noisy updates of the SR. When the projection is based on an eigendecomposition, this constitutes a form of *spectral regularization*[43]. For example, a smoothed version of the SR can be obtained by reconstructing the SR from its eigendecomposition using only low-frequency (high eigenvalue) components, thereby filtering out high-frequency noise (see Methods). Importantly, the regularization is topologically sensitive, meaning that smoothing respects boundaries of the environment. This property is not shared by regularization using a Fourier decomposition (Fig. S3). The regularization hypothesis is consistent with data suggesting that although grid cell input is not required for the emergence of place fields, place field stability and organization depends crucially on input from grid cells[44–46]. These eigenvectors also provide a useful partitioning of the task space, as discussed in the following section.

Subgoal discovery using grid fields

In structured environments, planning can be made more efficient by decomposing the task into subgoals, but the discovery of good subgoals is an open problem. The SR eigenvectors can be used for subgoal discovery by identifying "bottleneck states" that bridge large, relatively isolated clusters of states, and group together states that fall on opposite sides of the bottlenecks[47,48]. Since these bottleneck states are likely to be traversed along many optimal trajectories, they are frequently convenient waypoints to visit. Navigational strategies that exploit bottleneck states as subgoals have been observed in human navigation[49]. It is also worth noting that accompanying the neural results displayed in Fig. 7, the authors found that when subjects were asked to parse sequences of stimuli into events, stimuli found at topological bottlenecks were frequent breakpoints[18].

The formal problem of identifying these bottlenecks is known formally as the k-way normalized min-cut problem. An approximate solution can be obtained using spectral graph theory[50]. First, the top $\log k$ eigenvectors of a matrix known as the graph Laplacian are thresholded such that negative elements of

each eigenvector go to zero and positive elements go to one. Edges that connect between these two labeled groups of states are "cut" by the partition, and nodes adjacent to these edges are a kind of bottleneck subgoal. The first subgoals that emerge will be the cut from the lowest-frequency eigenvector, and these subgoals will approximately lie between the two largest, most separable clusters in the partition (see Supplemental Methods for more detail). A prioritized sequence of subgoals is obtained by incorporating increasingly higher frequency eigenvectors that produce partition points nearer to the agent.

The SR shares its eigenvectors with the graph Laplacian (see Supplemental Methods)[5], making SR eigenvectors equally suitable for this process of subgoal discovery. We show in Fig. S4 that the subgoals that emerge in a 2-step decision task and in a multicompartment environment tend to fall near doorways and decision points: natural subgoals for high-level planning. It is worth noting that SR matrices parameterized by larger discount factors γ will project predominantly on the large-spatial-scale grid components. The relationship between more temporally diffuse, abstract SRs, in which states in the same room are all encoded similarly (Fig. S2), and the subgoals that join those clusters can therefore be captured by which eigenvalues are large enough to consider.

The fact that large SR fields project predominantly onto eigenvectors with large spatial scales, whereas smaller SR fields project more strongly onto finer scale grid fields, is consistent with the smooth longitudinal gradient in connectivity between MEC and hippocampus[34]. Hippocampal cells with larger place fields are more densely wired to the entorhinal cells with larger spatial scales in their grids, and vice versa. It has also been shown experimentally that entorhinal lesions impair performance on navigation tasks and disrupt the temporal ordering of sequential activations in hippocampus while leaving performance on location recognition tasks intact[45,51]. This suggests a role of grid cells in spatial planning, and encourages us to speculate about a more general role for grid cells in hierarchical planning.

Discussion

The hippocampus has long been thought to encode a cognitive map, but the precise nature of this map is elusive. The traditional view that the map is essentially spatial[7,8] is not sufficient to explain some of the most striking aspects of hippocampal representation, such as the dependence of place fields on an animal's behavioral policy and the environment's topology. We argue instead that the map is essentially *predictive*, encoding expectations about an animal's future state. This view resonates with earlier ideas about the predictive function of the hippocampus[20,52–54]. Our main contribution is a formalization of this predictive function in a reinforcement learning framework, offering a new perspective on how the hippocampus supports adaptive behavior.

Our theory is connected to earlier work by Gustafson and Daw[13] showing how topologically-sensitive spatial representations recapitulate many aspects of place cells and grid cells that are difficult to reconcile with a purely Euclidean representation of space. They also showed how encoding topological structure greatly aids reinforcement learning in complex spatial environments. Earlier work by Foster and colleagues[12] also used place cells as features for RL, although the spatial representation did not explicitly encode topological structure. While these theoretical precedents highlight the importance of spatial representation, they leave open the deeper question of why particular representations are better than others. We showed that the SR naturally encodes topological structure in a format that enables efficient RL.

The work is also related to work done by Dordek *et al.*[23], who demonstrated that gridlike activity patterns from principal components of the population activity of simulated Gaussian place cells. As we mentioned in the Results, one point of departure between empirically observed grid cells data and SR eigenvector account is that in rectangular environments, SR eigenvector grid fields can have different

spatial scales aligned to the horizontal and vertical axis (see Fig. S8)[14]. In grid cells, the spatial scales tend to be approximately constant in all directions unless the environment changes[55]. The principal components of Gaussian place field activity are mathematically related to the SR eigenvectors, and naturally also have grid fields that scale independently along the perpendicular boundaries of a rectangular room. However, Dordek *et al.* found that when the components were constrained to have non-negative values and the constraint that components be orthogonal was relaxed, the scaling became uniform in all directions and the lattices became more hexagonal[23]. This suggests that the difference between SR eigenvectors and recorded grid cells is not fundamental to the idea that grid cells are applying a spectral dimensionality reduction. Rather, additional constraints such as non-negativity are required.

The SR can be viewed as occupying a middle ground between model-free and model-based learning. Model-free learning requires storing a look-up table of cached values estimated from the reward history[1,56]. Should the reward structure of the environment change, the entire look-up table must be re-estimated. By decomposing the value function into a predictive representation and a reward representation, the SR allows an agent to flexibly recompute values when rewards change, without sacrificing the computational efficiency of model-free methods[4]. Model-based learning is robust to changes in the reward structure, but requires inefficient algorithms like tree search to compute values[1,15].

Certain behaviors often attributed to a model-based system can be explained by a model in which predictions based on state dynamics and the reward function are learned separately. For instance, the *context preexposure facilitation effect* refers to the finding that contextual fear conditioning is acquired more rapidly if the animal has the chance to explore the environment for several minutes before the first shock[57]. The facilitation effect is classically believed to arise from the development of a conjunctive representation of the context in the hippocampus, though areas outside the hippocampus may also develop a conjunctive representation in the absence of the hippocampus, albeit less efficiently[58]. The SR provides a somewhat different interpretation: over the course of preexposure, the hippocampus develops a *predictive* representation of the context, such that subsequent learning is rapidly propagated across space. Figure S5 shows a simulation of this process and how it accounts for the facilitation effect.

Recent work has elucidated connections between models of episodic memory and the SR. Specifically, Gershman *et al.* demonstrated that the SR is closely related to the Temporal Context Model (TCM) of episodic memory[16,19]. The core idea of TCM is that items are bound to their temporal context (a running average of recently experienced items), and the currently active temporal context is used to cue retrieval of other items, which in turn cause their temporal context to be retrieved. The SR can be seen as encoding a set of item-context associations. The connection to episodic memory is especially interesting given the crucial mnemonic role played by the hippocampus and entorhinal cortex in episodic memory. Howard and colleagues[59] have laid out a detailed mapping between TCM and the medial temporal lobe (including entorhinal and hippocampal regions).

Spectral graph theory provides insight into the topological structure encoded by the SR. We showed specifically that eigenvectors of the SR can be used to discover a hierarchical decomposition of the environment for use in hierarchical RL. Mahadevan *et al.* demonstrated that the related Laplacian eigenvectors are useful as a representational basis for approximating value functions, dubbing these eigenvectors "protovalue functions"[60]. Spectral analysis has frequently been invoked as a computational motivation for entorhinal grid cells (e.g., by Krupic and colleagues[61]). The fact that any function can be reconstructed by sums of sinusoids suggests that the entorhinal cortex implements a kind of Fourier transform of space. However, Fourier analysis is not the right mathematical tool when dealing with spatial representations in a topologically structured environment, since we do not expect functions to be smooth over boundaries in the environment. This is precisely the purpose of spectral graph theory: Instead of being maximally smooth over Euclidean space, the eigenvectors of the graph Laplacian embed the smoothest

approximation of a function that respects the graph topology[60].

In conclusion, the SR provides a unifying framework for a wide range of observations about the hippocampus and entorhinal cortex. The multifaceted functions of these brain regions can be understood as serving a superordinate goal of prediction.

Methods

Task simulation

Environments were simulated by discretizing the plane into points, and connecting these points along a triangular lattice. The adjacency matrix A was constructed such that $A(i,j) = 1$ wherever it is possible to transition between states i and j, and 0 otherwise. The transition probability matrix T was defined so that $T(i,j)$ is the probability of transitioning from state i to j. Under a random walk policy, the transition probability distribution is uniform over allowable transitions.

SR computation

To solve for the successor representation, we used the convergence of the geometric sum of transition matrices, T, discounted by $\gamma \in [0,1]$:

$$M = \sum_{t=0}^{\infty} \gamma^t T^t = (I - \gamma T)^{-1} \tag{5}$$

To simulate the long-term effect of rewards and punishments in the environment, the optimal policy was found using value iteration[15], and the SR was computed analytically under this policy. When demonstrating learning dynamics, the SR was estimated using the TD-learning (Fig. 11, Supp. Fig. 1, Supp. Fig. 3). Noise was injected into the location signal by adding uniform random noise with mean 0 to the state indicator vector.

Eigenvector computation and Spectral Regularization

For Figure 11, eigenvectors were computed incrementally using a Candid Covariance-free Incremental PCA (CCIPCA), an algorithm that efficiently implements stochastic gradient descent to compute principal components[62] (eigenvectors and principal components are equivalent in this and many domains). All eigenvectors were thresholded because firing rates cannot be negative. Spectral regularization was implemented by reconstructing the SR from the truncated eigendecomposition (Fig. S3). Details can be found in the Supplemental Methods.

In generating the grid cells shown, we assume random walk policy, which is the maximum entropy prior for policies (see[63] for why maximum entropy priors can be good priors for regularization). However, since the learned eigenvectors are sensitive to the sampling statistics, our model predicts that regions of the task space more frequently visited would come to be over-represented in the grid space (see Figure S6 for examples).

Quantifying place and grid fields

To quantify place field clustering, center of mass (CoM) of SR place fields was computed by summing the locations of firing, weighted by the firing rate at that location (normalized so that the total firing summed to 1):

$$\mathrm{CoM}(s) = \frac{\sum_{s'} M(s,s')\mathbf{p}(s')}{\sum_{s'} M(s,s')}, \tag{6}$$

where $\mathbf{p}(s')$ is the (X,Y) coordinate of the place field centered at state s'.

Grid field quantifications paralleled the analyses of Krupic *et al.*[40]: an ellipse was fit to the 6 peaks closest to the central peak, "orientation" refers to the orientation of the main axes (a,b). "Correlation" always refers to the Pearson correlation, "spatial correlation" refers to the Pearson correlation computed over points in space (as opposed to points in a vector), and spatial autocorrelation refers to the 2D auto-convolution.

To measure similarity between halves of the environment in Figure 9, we 1) computed the spatial autocorrelation for each half, 2) selected a circular window in the center of the autocorrelation, and 3) computed the correlation between autocorrelations of the two halves in the window. This paralleled the analysis taken by Krupic *et al.*[40] and provides a measure of grid similarity across halves of the environment. The circular window is used to control for the fact that the boundaries of the square and trapezoid in the two halves of the respective environments differ.

In evaluating our simulations of the grid fields reported by Carpenter *et al.*[42] (Fig. 11), the local model consisted of the set of 2D Fourier components bounded by the size of the compartment and the global model consisted of the set of 2D Fourier components bounded by the size of the environment. "Model fit" was measured for each eigenvector by finding maximum correlation over all model components between the eigenvector and model component.

References

1. Daw, N. D., Niv, Y. & Dayan, P. Uncertainty based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience* **8**, 1704–1711 (2005).
2. Tolman, E. C. Cognitive maps in rats and men. *Psychological Review* **55**, 189–208 (1948).
3. Schultz, W., Dayan, P. & Montague, P. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
4. Dayan, P. Improving generalization for temporal difference learning: The successor representation. *Neural Computation* **5**, 613–624 (1993).
5. Stachenfeld, K. L., Botvinick, M. & Gershman, S. J. Design principles of the hippocampal cognitive map. In *Advances in Neural Information Processing Systems 27*, 2528–2536 (MIT Press, 2014).
6. O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Research* **34**, 171–175 (1971).
7. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Oxford: Clarendon Press, 1978).
8. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M. B. Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience* **7**, 663–678 (2006).
9. Muller, R. U., Stead, M. & Pach, J. The hippocampus as a cognitive graph. *The Journal of General Physiology* **107**, 663–694 (1996).
10. Penny, W., Zeidman, P. & Burgess, N. Forward and backward inference in spatial cognition. *PLoS Computational Biology* **9**, e1003383 (2013).
11. Rueckert, E., Kappel, D., Tanneberg, D., Pecevski, D. & Peters, J. Recurrent spiking networks solve planning tasks. *Scientific reports* **6** (2016).
12. Foster, D., Morris, R. & Dayan, P. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* **10**, 1–16 (2000).

13. Gustafson, N. J. & Daw, N. D. Grid cells, place cells, and geodesic generalization for spatial reinforcement learning. *PLoS Computational Biology* **7**, e1002235 (2011).

14. Hafting, T., Fyhn, M., Molden, S., Moser, M. B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).

15. Sutton, R. & Barto, A. *Reinforcement Learning: An Introduction* (MIT Press, 1998).

16. Gershman, S., Moore, C., Todd, M., Norman, K. & Sederberg, P. The successor representation and temporal context. *Neural Computation* **24**, 1553–1568 (2012).

17. Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**, 585–595 (2010).

18. Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B. & Botvinick, M. M. Neural representations of events arise from temporal community structure. *Nature neuroscience* **16**, 486–492 (2013).

19. Howard, M. & Kahana, M. A distributed representation of temporal context. *Journal of Mathematical Psychology* **46**, 269–299 (2002).

20. Lisman, J. & Redish, A. D. Prediction, sequences and the hippocampus. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **364**, 1193–1201 (2009).

21. Pfeiffer, B. E. & Foster, D. J. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature* **497**, 74–79 (2013).

22. Burgess, N., Recce, M. & O'Keefe, J. A model of hippocampal function. *Neural Networks* **7**, 1065–1081 (1994).

23. Dordek, Y., Meir, R. & Derdikman, D. Extracting grid characteristics from spatially distributed place cell inputs using non-negative PCA. *eLife* (2015).

24. Mehta, M. R., Barnes, C. A. & McNaughton, B. L. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences* **94**, 8918–8921 (1997).

25. Mehta, M., Quirk, M. & Wilson, M. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* **25**, 707–715 (2000).

26. Muller, R. U. & Kubie, J. L. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *The Journal of Neuroscience* **7**, 1951–1968 (1987).

27. Skaggs, W. & McNaughton, B. Spatial firing properties of hippocampal CA1 populations in an environment containing two visually identical regions. *The Journal of Neuroscience* **18**, 8455–8466 (1998).

28. Alvernhe, A., Save, E. & Poucet, B. Local remapping of place cell firing in the Tolman detour task. *European Journal of Neuroscience* **33**, 1696–1705 (2011).

29. Fenton, A., Zinyuk, L. & Bures, J. Place cell discharge along search and goal-directed trajectories. *European Journal of Neuroscience* **12**, 3450 (2001).

30. Kobayashi, T., Tran, A., Nishijo, H., Ono, T. & Matsumoto, G. Contribution of hippocampal place cell activity to learning and formation of goal-directed navigation in rats. *Neuroscience* **117**, 1025–35 (2003).

31. Hollup, S., Molden, S., Donnett, J., Moser, M. & Moser, E. Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *Journal of Neuroscience* **21**, 1635–1644 (2001).

32. Hok, V. *et al.* Goal-related activity in hippocampal place cells. *The Journal of Neuroscience* **27**, 472–82 (2007).

33. Kjelstrup, K. *et al.* Finite scale of spatial representation in the hippocampus. *Science* **321**, 140–143 (2008).

34. Strange, B., Witter, M., Lein, E. & Moser, E. Functional organization of the hippocampal longitudinal axis. *Nature Reviews Neuroscience* **15**, 655–669 (2014).

35. Sutton, R. Td models: Modeling the world at a mixture of time scales. In *Proceedings of the 12th International Conference on Machine Learning* (1995).

36. Modayil, J., White, A. & Sutton, R. Multi-timescale nexting in a reinforcement learning robot. *arXiv:1112.1133 [cs]* (2011).

37. Schapiro, A. C., Turk-Browne, N., Norman, K. & Botvinick, M. Statistical learning of temporal community structure in the hippocampus. *Hippocampus* **26**, 3–8 (2016).

38. Deuker, L., Bellmund, J., Schröder, T. & Doeller, C. An event map of memory space in the hippocampus. *eLife* **5**, e16534 (2016).

39. Franzius, M., Sprekeler, H. & Wiskott, L. Slowness and sparseness lead to place, head-direction, and spatial-view cells. *PLoS Computational Biology* **3**, 3287–3302 (2007).

40. Krupic, J., Bauza, M., Burton, S., Barry, C. & O'Keefe, J. Grid cell symmetry is shaped by environmental geometry. *Nature* **518**, 232–235 (2015).

41. Derdikman, D. *et al.* Fragmentation of grid cell maps in a multicompartment environment. *Nature Neuroscience* **12**, 1325–1332 (2009).

42. Carpenter, F., Manson, D., Jeffery, K., Burgess, N. & Barry, C. Grid cells form a global representation of connected environments. *Current Biology* **25**, 1176–1182 (2015).

43. Mazumder, R., Hastie, T. & Tibshirani, R. Spectral regularization algorithms for learning large incomplete matrices. *Journal of Machine Learning Research* **11**, 2287–2322 (2010).

44. Bonnevie, T. *et al.* Grid cells require excitatory drive from the hippocampus. *Nature Neuroscience* **16**, 309–317 (2013).

45. Hales, J. *et al.* Medial entorhinal cortex lesions only partially disrupt hippocampal place cells and hippocampus-dependent place memory. *Cell Reports* **9**, 893–901 (2014).

46. Muessig, L., Hauser, J., Wills, T. & Cacucci, F. A developmental switch in place cell accuracy coincides with grid cell maturation. *Neuron* **86**, 1167–1173 (2015).

47. Şimşek, Ö., Wolfe, A. & Barto, A. Identifying useful subgoals in reinforcement learning by local graph partitioning. In *Proceedings of the 22nd International Conference on Machine Learning*, 816–823 (ACM, 2005).

48. Solway, A. *et al.* Optimal behavioral hierarchy. *PLoS Computational Biology* **559** (2014).

49. Ribas-Fernandes, J. *et al.* A neural signature of hierarchical reinforcement learning. *Neuron* **71**, 370–379 (2011).

50. Shi, J. & Malik, J. Normalized cuts and image segmentation. In *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, 888–905 (IEEE, 2000).

51. Schlesiger, M. *et al.* The medial entorhinal cortex is necessary for temporal organization of hippocampal neuronal activity. *Nature Neuroscience* **18**, 1123–1132 (2015).

52. Blum, K. & Abbott, L. A model of spatial map formation in the hippocampus of the rat. *Neural Computation* **8**, 85–93 (1996).

53. Levy, W. B., Hocking, A. B. & Wu, X. Interpreting hippocampal function as recoding and forecasting. *Neural Networks* **18**, 1242–1264 (2005).

54. Buckner, R. L. The role of the hippocampus in prediction and imagination. *Annual Review of Psychology* **61**, 27–48 (2010).

55. Barry, C., Hayman, R., Burgess, N. & Jeffery, K. Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience* **10**, 682–684 (2007).

56. Dolan, R. J. & Dayan, P. Goals and habits in the brain. *Neuron* **80**, 312–25 (2013).

57. Fanselow, M. From contextual fear to a dynamic view of memory systems. *Trends in Cognitive Sciences* **14**, 7–15 (2010).

58. Wiltgen, B. J., Sanders, M. J., Anagnostaras, S., Sage, J. & Fanselow, M. S. Context fear learning in the absence of the hippocampus. *The Journal of Neuroscience* **26**, 5484–5491 (2006).

59. Howard, M., Fotedar, M., Datey, A. & Hasselmo, M. The temporal context model in spatial navigation and relational learning: toward a common explanation of medial temporal lobe function across domains. *Psychological Review* **112**, 75–116 (2005).

60. Mahadevan, S. & Maggioni, M. Proto-value functions: A Laplacian framework for learning representation and control in markov decision processes. *Journal of Machine Learning Research* **8**, 2169–2231 (2007).

61. Krupic, J., Burgess, N. & O'Keefe, J. Neural representations of location composed of spatially periodic bands. *Science* **337**, 853–857 (2012).

62. Weng, J., Zhang, Y. & Hwang, W. Candid covariance-free incremental principal component analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**, 1034–1040 (2003).

63. Bialek, W. *Biophysics: Searching for Principles* (Princeton University Press, 2012).

Acknowledgments

We are grateful to Tim Behrens, Ida Mommenejad, and Kevin Miller for helpful discussions, and to Alexander Mathis and Honi Sanders for comments on an earlier draft of the paper. This research was supported by the NSF Collaborative Research in Computational Neuroscience (CRCNS) Program Grant IIS-120 7833 and The John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agencies.

Author contributions statement

All authors conceived the model and wrote the manuscript. Simulations were carried out by K.S.

Additional information

The authors declare no competing financial interests.

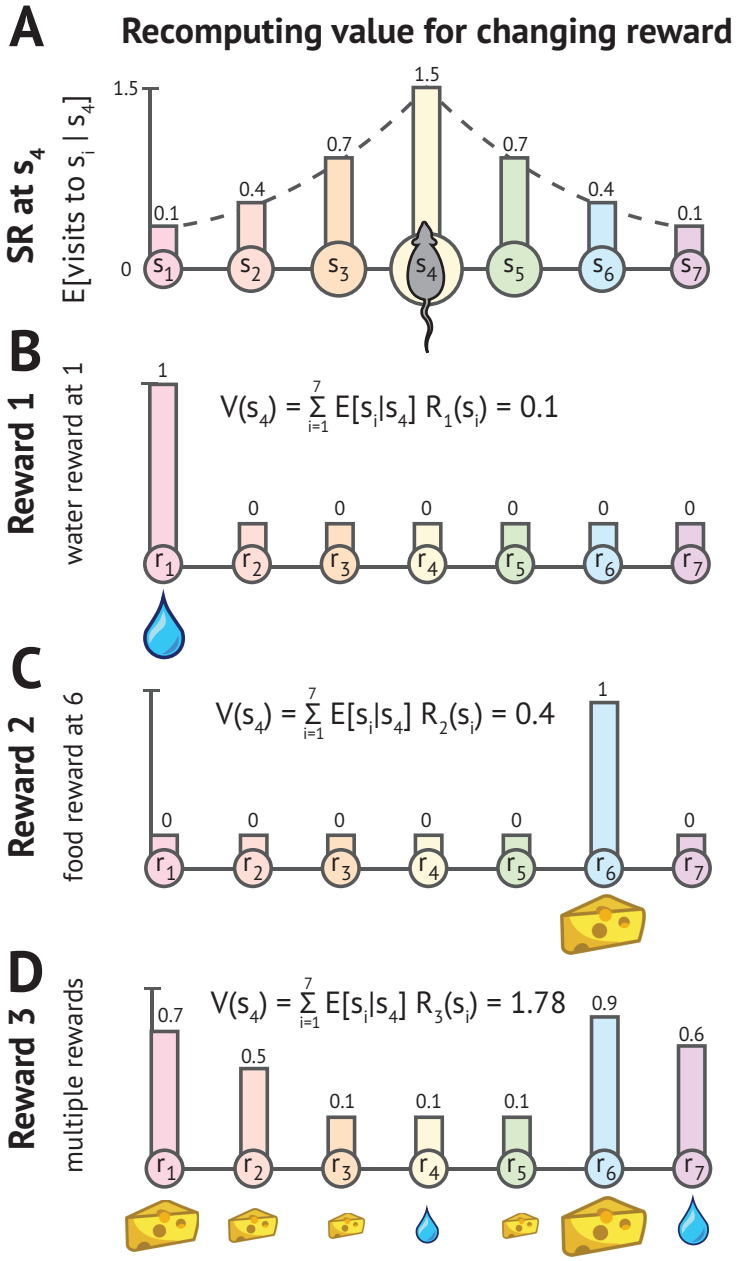

Figure 1. *Updating value following change in reward.* Since the representations of state and reward are decoupled, value functions can be rapidly recomputed for new reward functions without changing the SR. Panels A-D show how the value of s_4 changes under different reward functions.

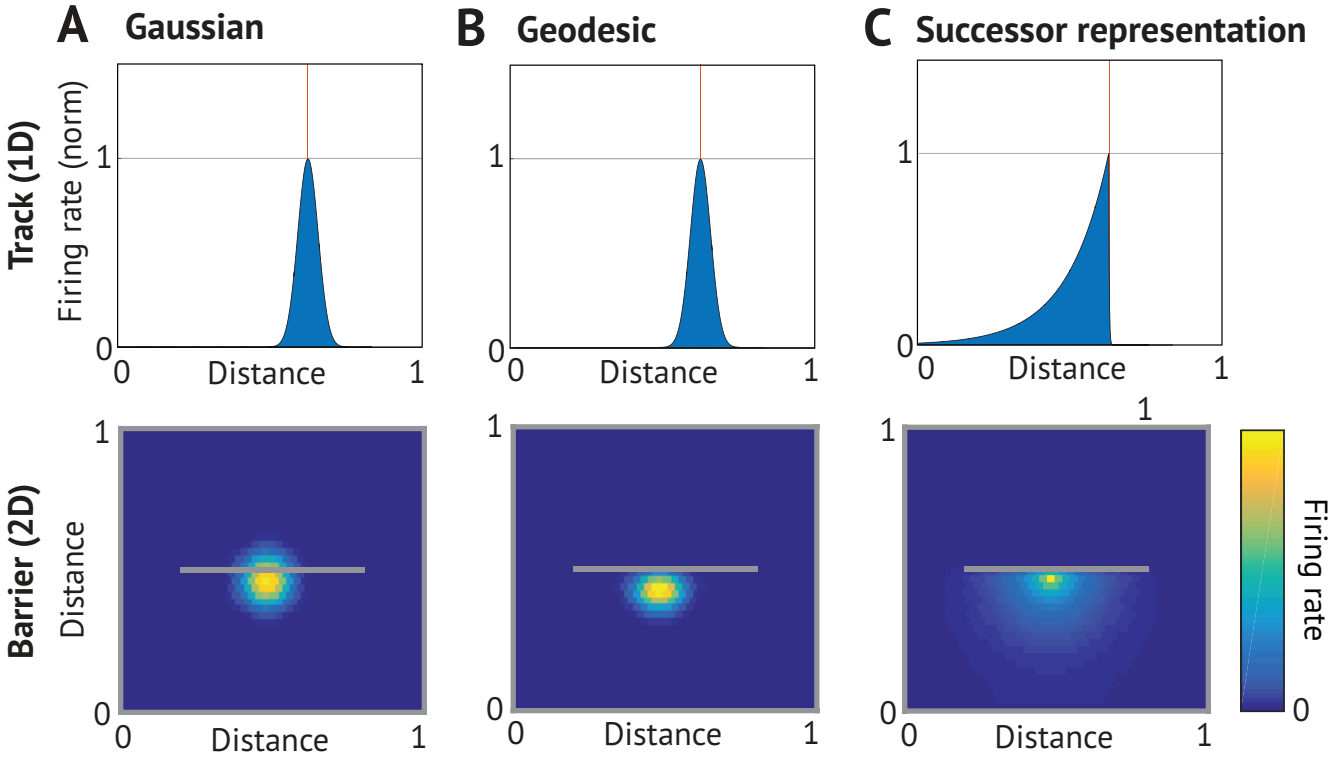

Figure 2. *Comparison of place cell models.* (Top) 1D track with left-to-right preferred direction of travel, red line marking field center; (bottom) 2D environment with a barrier indicated by gray line. (A) *Gaussian place field.* Firing of place cells decays with Euclidean distance from the center of the field regardless of experience and environmental topology. (B) *Topological place field.* Firing rate decays with geodesic distance from the center of the field, which respects boundaries in the environment but is invariant to the direction of travel[13]. (C) *SR place field.* Firing rate is proportional to the discounted expected number of visits to other states under the current policy. On the directed track, fields will skew opposite the direction of motion to anticipate the upcoming successor state. Since the policy will not permit traversing walls, successor fields warp around obstacles.

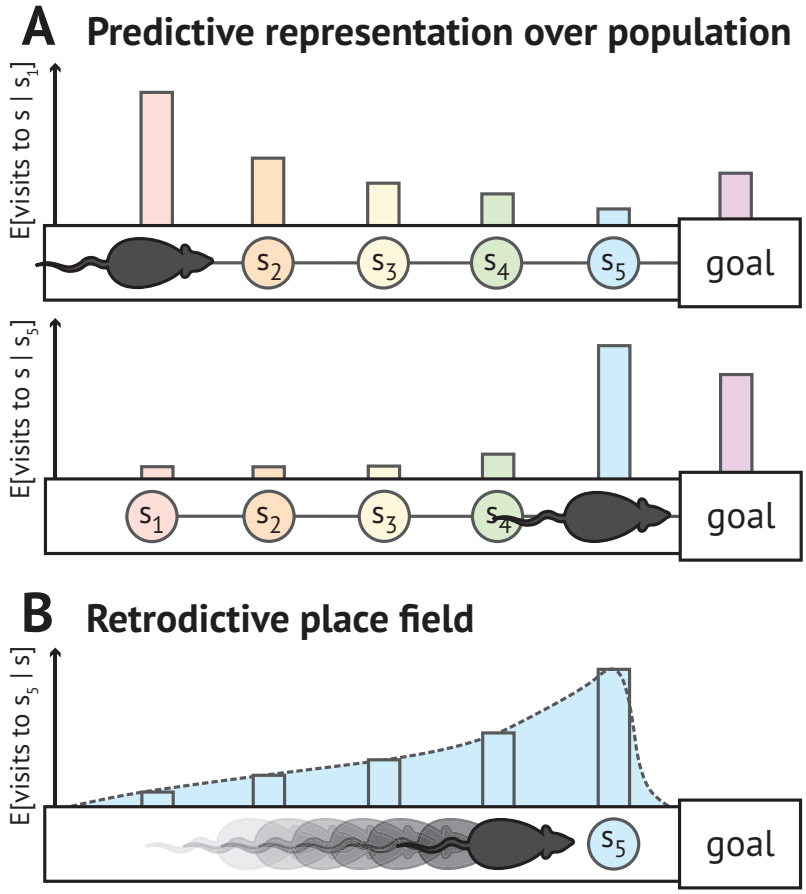

Figure 3. *Illustration of retrodictive cells.* (A) A neural population encodes a prospective representation such that the firing rate of each cell is proportional to the discounted expected number of times its preferred state will be visited in the future. This population code is skewed toward upcoming states. Each colored bump represents the firing rate of a different place field located along the track. (B) The place field for a single cell skews retrodictively toward past states that predict the cell's preferred state. When the blue state is visited, it becomes automatically associated with all past states that predicted it. This automatically assigns credit for upcoming reward to preceding states.

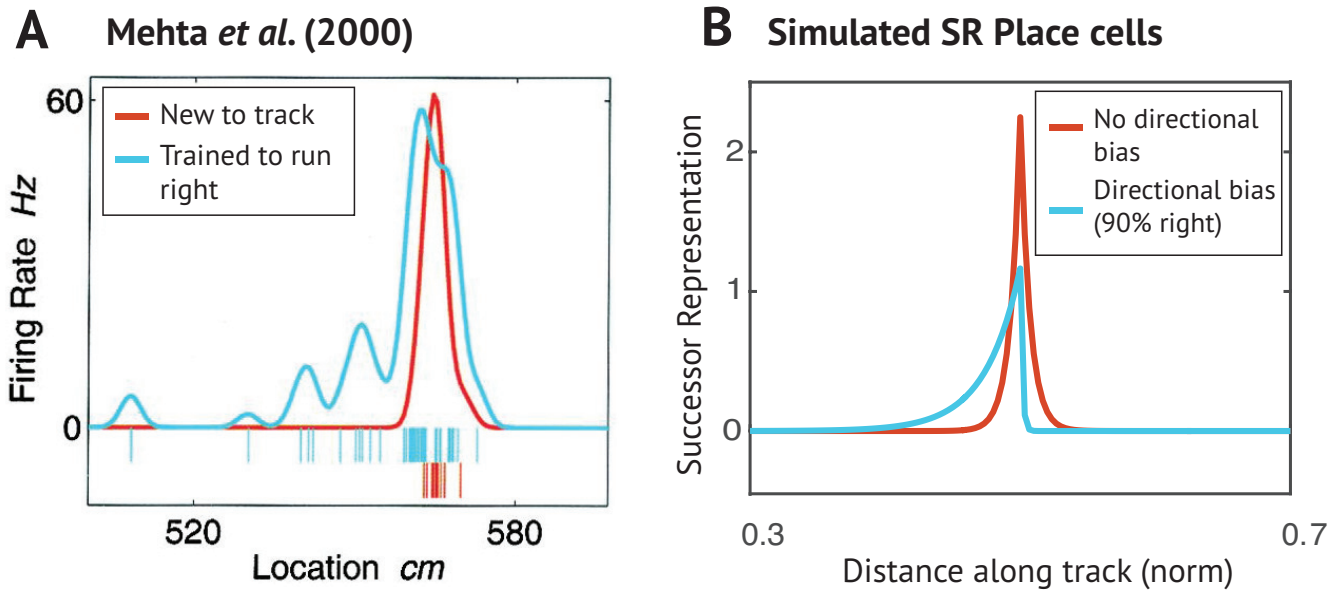

Figure 4. *Predictive skewing of place fields.* (A) As a rat is trained to run repeatedly in a preferred direction along a narrow track, initially symmetric place cells (red) begin to skew (blue) opposite the direction of travel[25]. (B) When transitions in either direction are equally probable, SR place fields are symmetric (red). Under a policy in which transitions to the right are more probable than to the left, simulated SR place fields skew opposite the direction of travel toward states predicting the preferred state (blue).

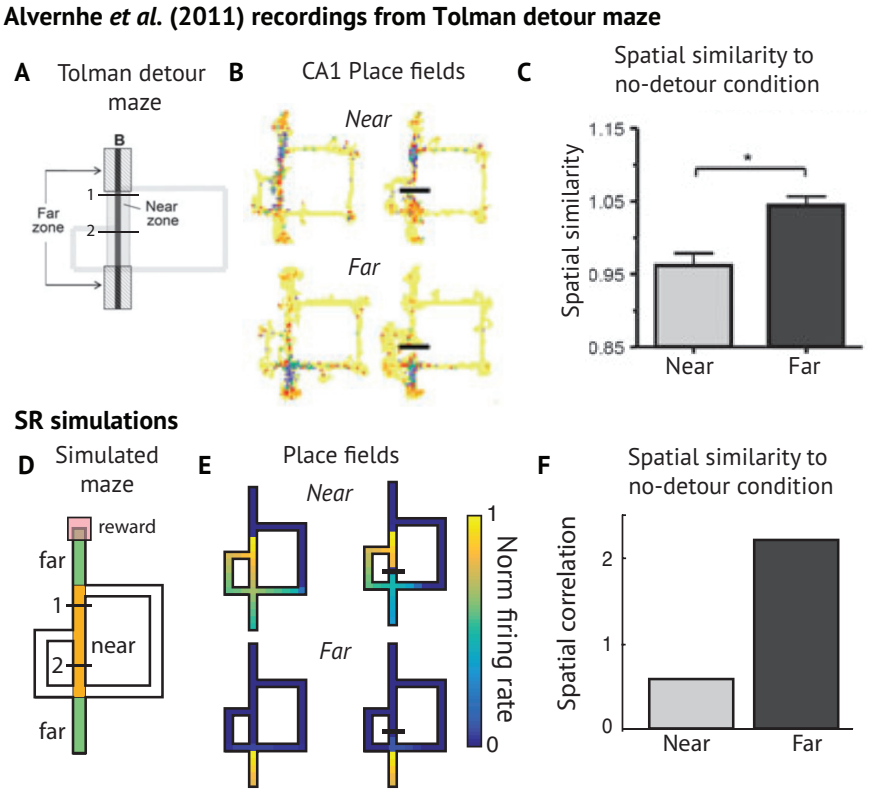

Figure 5. *Place fields near detour.* (A) Maze used by Alvernhe and colleagues[28] for studying how place cell firing is affected by the insertion of barriers in a Tolman detour maze. Reward is delivered at location B. "Near" and "Far" zones are defined. In "early" and "late" detour conditions, a clear barrier blocks the shortest path, forcing the animal to take the take the short detour to the left or the longer detour to the right. (B) Example CA1 place fields recorded from a rat navigating the maze. (C) Over the population, place fields near the barrier changed their shape, while the rest remained unperturbed. This is shown by computing the spatial correlation between place field activity maps with and without barriers present. (D) The environment used to simulate the experimental results. (E) Example SR place fields near to and far from the barrier, before and after barrier insertion. (F) When barriers are inserted, SR place fields change their fields near the path blocked by the barrier and less so at more distal locations where policy is unaffected. The effect is more pronounced in the early detour condition because the detour appears closer to the start.

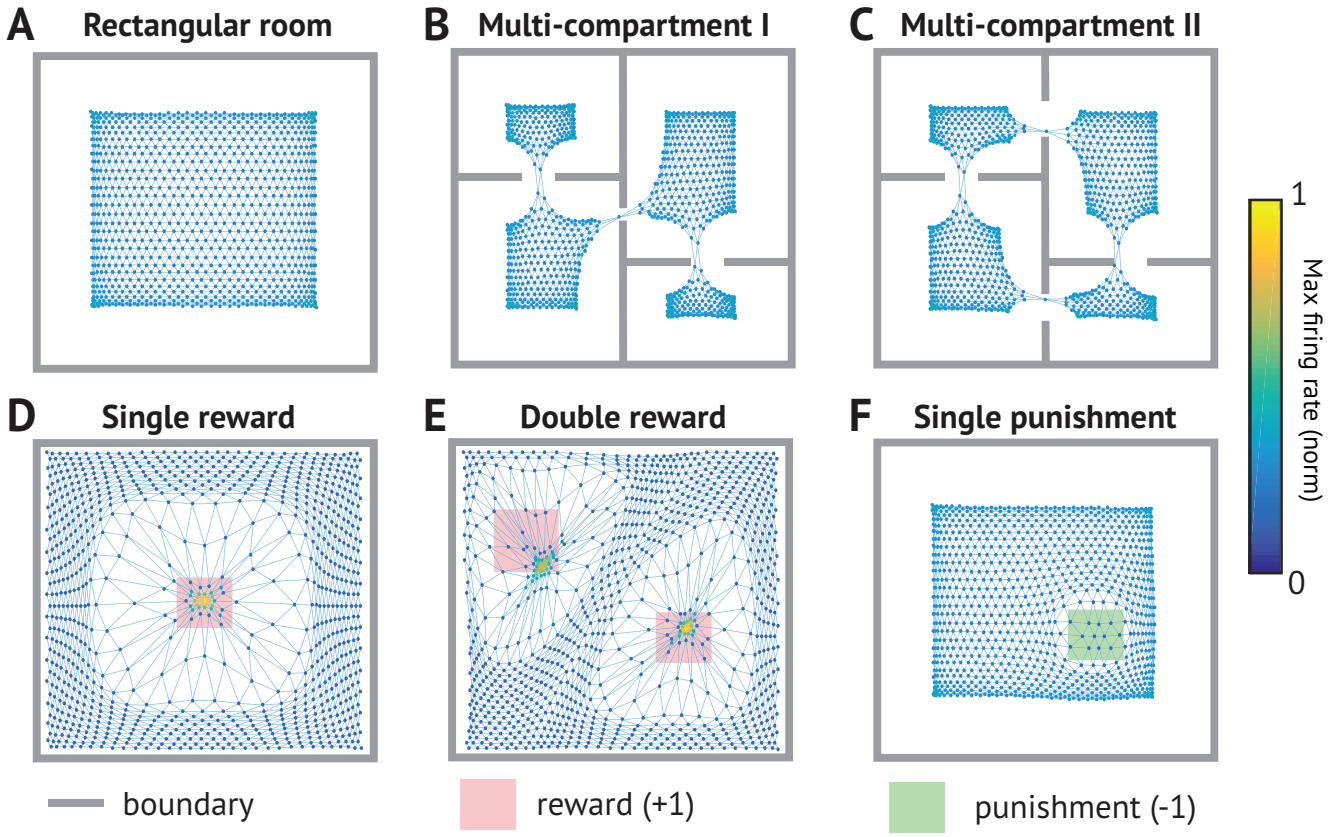

Figure 6. *Clustering of place fields.* Points in each plot mark centers of mass (CoMs) of SR place fields. CoMs are skewed towards locations with higher visitation probability. Each point is colored according to maximum firing rate of its field (the maximum number of expected visitations). (A) Under a random walk, CoMs will be uniformly arranged near the center. Near boundaries, CoMs shift towards the center of the room because the boundaries repel the animal's motion. (B,C) Similarly, CoMs tend to cluster within compartments of a multi-compartment environment. (D,E) When rewards are introduced, nearby place fields shift their CoMs toward the rewarded locations (pink), because the animal tends to traverse those locations. In contrast, other fields shift their CoMs away from the rewarded location, because they tend only to be visited when the animal approaches from a more distal location en route to the reward. (F) Punishments (green) repel the animal, and accordingly shift the CoMs of nearby place fields away from the punished location.

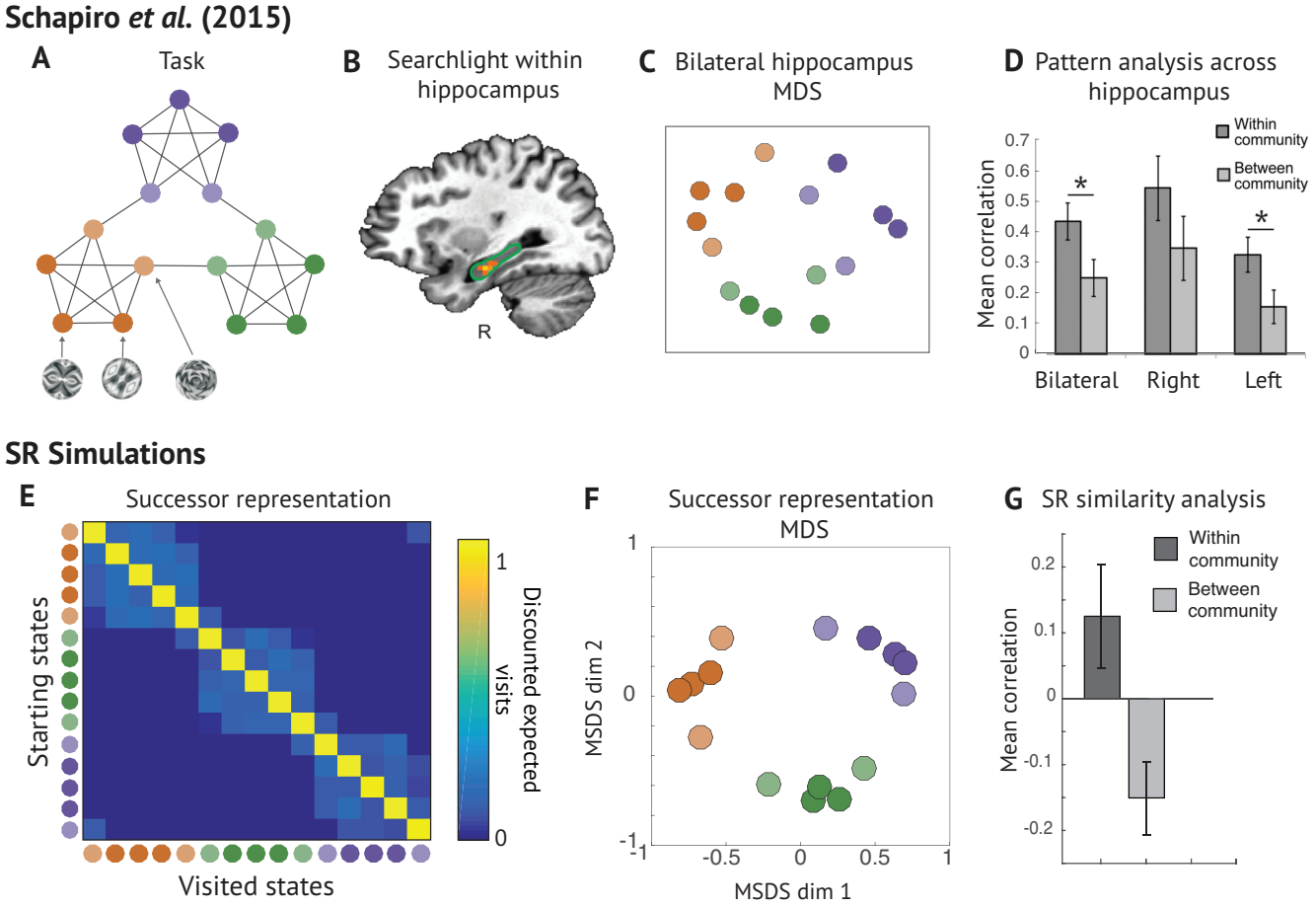

Figure 7. *Hippocampal representations in non-spatial task.* (A) Schapiro *et al.*[37] showed subjects sequences of fractal stimuli drawn from the task graph shown, which has clusters of interconnected nodes (or "communities"). Nodes of the same color fall within the same community, with the lighter colored nodes connecting to adjacent communities. (B) A searchlight within hippocampus showed a stronger within-community similarity effect in anterior hippocampus. (C, D) States within the same cluster had a higher degree of representational similarity in hippocampus, and multidimensional scaling (MDS) of the hippocampal BOLD dissimilarity matrix captured the community structure of the task graph[37]. (E) The SR matrix learned on the task. The block diagonal structure means that states in the same cluster predict each other with higher probability. (F) Multidimensional scaling of dissimilarity between rows of the SR matrix reveals the community structure of the task graph. (G) Consistent with this, the average within-community SR state similarity is consistently higher than the average between-community SR state similarity.

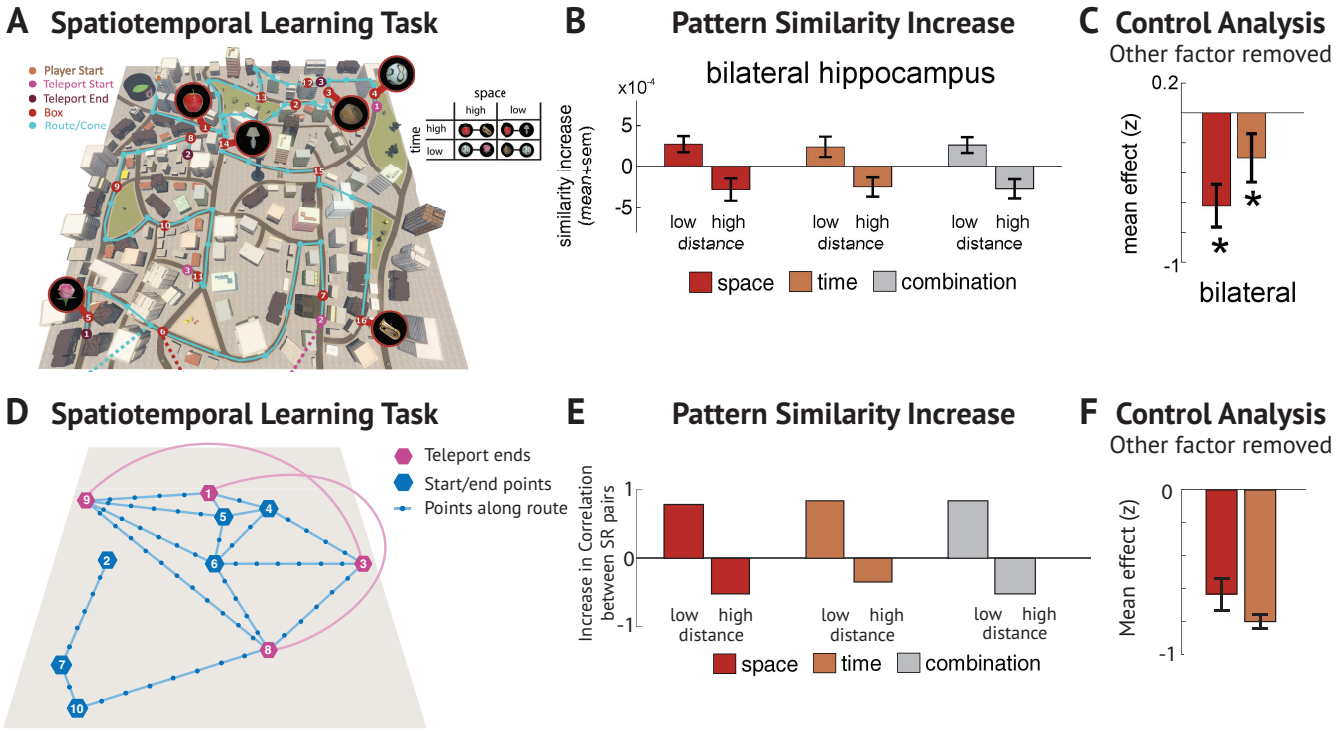

Figure 8. *Hippocampal representations in spatio-temporal task.* (A) Deuker *et al.*[38] trained subjects on a spatio-temporal navigation task. Subjects were told to objects scattered about the map. It is possible to take a "teleportation" shortcut between certain pairs of states (pink and purple), and other pairs of states are sometimes joined only by a long, winding path. Nearness in time is therefore partially decoupled from nearness in space. (B) The authors find significant increase in hippocampal representational similarity between nearby states and a decrease for distant states. This effect holds when states are nearby in space, time, or both. (C) Since spatial and temporal proximity are correlated, the authors controlled for the each factor and measured the effect of the remaining factor on the residual. (D-F) Simulation of experimental results in panels A-C.

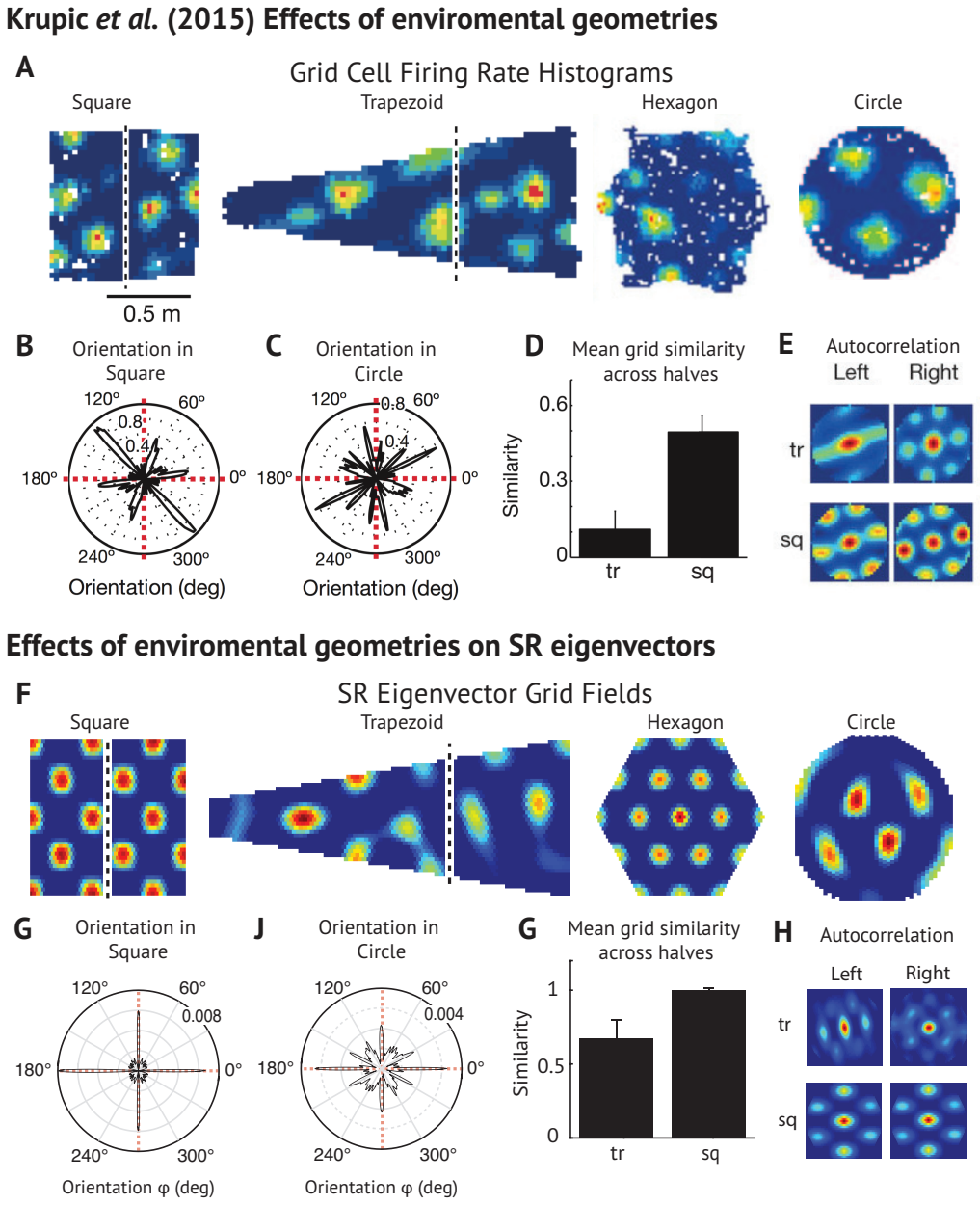

Figure 9. *Grid fields in geometric environments.* (A) Grid fields recorded in a variety of geometric environments[40]. Grid fields in trapezoid and square environments are split at the dividing line shown for split-halves analysis. (B,C) Grid fields in the square environment had more consistent orientations with respect to boundaries and distal cues than in the square environment. (D) While grid fields tend to be similar on both halves of a square (sq) environment, they tend to be less similar across halves of the irregular trapezoidal (tr) environment. (E) Autocorrelograms for different halves of trapezoidal and square environments in circular windows used for split-halves anal. (F-H) Simulations of experimental results in panels A-E.

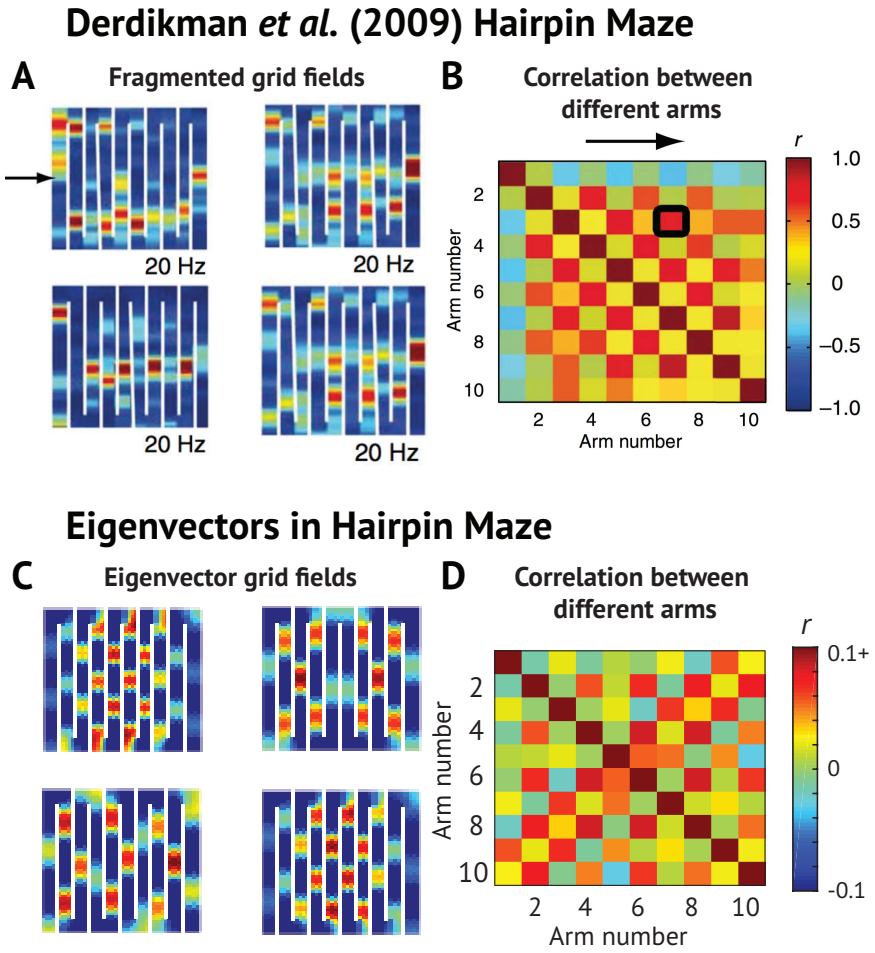

Figure 10. *Grid fragmentation in compartmentalized maze.* (A) Barriers in the hairpin maze cause grid fields to fragment repetitively across arms[41]. (B) Spatial correlation between activity in different arms. The checkerboard pattern emerges because grid fields frequently repeat themselves in alternating arms. (C-D) Simulations of the experimental results in panels A-B.

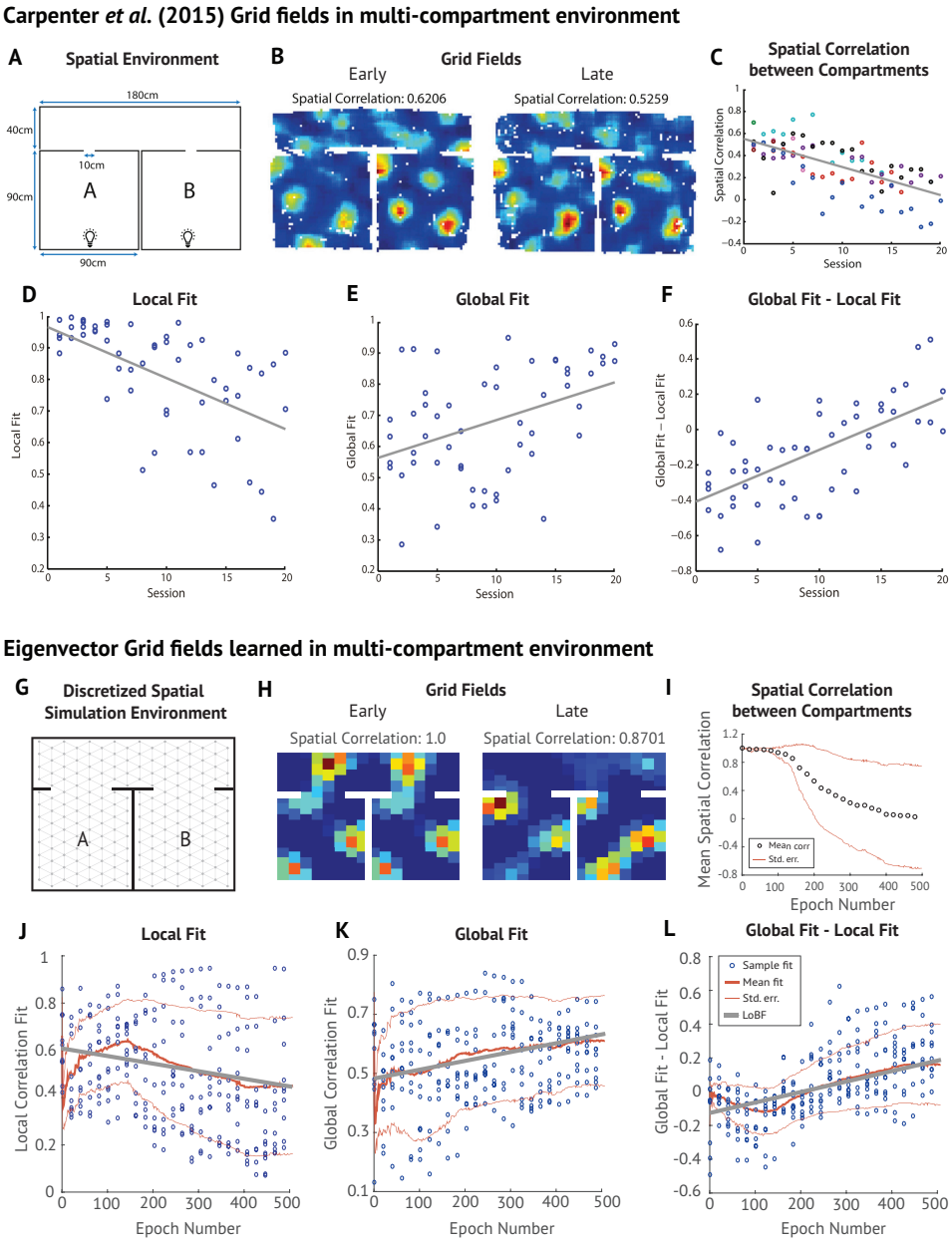

Figure 11. *Grid fields in multi-compartment environment.* (A) Multi-compartment environment employed by Carpenter and colleagues[42]. (B) Example grid fields early and late in training. (C) Spatial correlation between grid fields in compartments A and B across sessions. (D-F) To explain this decline in inter-compartment similarity, Carpenter and colleagues fit a local model (grid constrained to replicate between the two compartments) and a global model (single continuous grid spanning both compartments). They found that the local fit decreased across sessions, while the global fit increased, and correspondingly the difference between the two models increased. (G-L) Simulation of experimental results in panels A-F. In I-J, the blue circles indicate individual samples, the thick red line denotes the mean, the thin red lines denote one standard deviation from the mean, and the thick gray lines are lines of best fit.

Supplementary information for "The hippocampus as a predictive map"

Kimberly L. Stachenfeld, Matthew M. Botvinick, Samuel J. Gershman

Learning eigenvectors by stochastic gradient descent

The problem of computing the K largest eigenvectors of the successor representation is equivalent to choosing the embedding matrix Y that minimizes the following cost function:

$$C(Y) = \sum_s \sum_{s'} E(y_s, y_{s'}) T(s, s'), \tag{1}$$

where $y_s \in \mathbb{R}^K$ is the embedding of state s and $E(y_s, y_{s'}) = \gamma ||y_s - y_{s'}||^2$. Intuitively, this cost function favors embeddings that are similar for states that are likely to be visited sequentially. The optimal embedding can be viewed as the solution to a kernel PCA problem with M as the kernel.

The cost function can also be viewed as an expectation under the Markov chain induced by the transition function T:

$$C(Y) = \mathbb{E}_{s' \sim T(s,\cdot)} \left[E(y_s, y_{s'}) \right]. \tag{2}$$

We can therefore minimize the cost function online by sampling transitions ($s \to s'$) from the Markov chain and stochastically following the gradient:

$$y_i^{n+1} = y_i^n - \alpha^n \nabla_{y_i} E(y_s^n, y_{s'}^n), \tag{3}$$

where α^n is the step-size (learning rate) at time n, and the gradient is given by:

$$\nabla_{y_i} E(y_s^n, y_{s'}^n) = 2\eta\gamma \frac{y_s^n - y_{s'}^n}{||y_s^n - y_{s'}^n||}, \tag{4}$$

$$\eta = \begin{cases} 1 & \text{if } i = s \\ -1 & \text{if } i = s' \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

In addition, the embedding vectors are subject to a normalization constraint such that $\|y_s\| = 1$ for every state.

Smoothing with Spectral Regularization

For Fig. S3, spectral regularization was implemented by reconstructing the SR, M_t, from the truncated k-dimensional eigendecomposition at each timestep (Fig. S3). This means that we set projections along all but the eigenvectors with the k highest eigenvalues to 0. To be more precise, the SR M can be eigendecomposed as:

$$M = U \Lambda U^T \tag{6}$$

Where U is a matrix with the eigenvectors of M as its columns and Λ is a diagonal matrix with eigenvalues along the diagonal. The SR M can be approximately reconstructed using only the first k eigenvectors and eigenvalues, which we will denote as U_k and Λ_k respectively.

$$M^k = U_k M_t^k U_k^T \tag{7}$$

We implemented spectral regularization online by projecting M_t onto the eigenvectors of M and then reconstructing M_t^k from the k components with largest eigenvectors:

$$M_t^k = U_k^T U_k M_t^k U_k^T U_k \tag{8}$$

Partitioning the state space into subgoals with normalized min-cut

The eigendecomposition of the SR can be written as:

$$M = U\Lambda U^T \tag{9}$$

where eigenvectors and eigenvalues are ordered such that $\lambda_i \geq \lambda_{i+1}$.

The normalized min-cut can be approximated by taking the second eigenvector, u_2, and thresholding it such that states i for which $u_{2i} < 0 - \varepsilon$ go to 0, and states for which $u_{2i} > 0 + \varepsilon$ go to 1 for some very small ε. States such that $-\varepsilon \leq u_{2i} \leq \varepsilon$ will go to 0.5 and will comprise your subgoals. These subgoals will separate nodes so that there are as few edges as possible going between nodes labeled 0 and 1, while keeping these groups as large as possible[1]. Ideally, ε will be sufficiently small that this selects only the nodes that connect the partitioned groups.

Partial group membership can be implemented by rescaling the eigenvector to the [0,1] range rather than thresholding. Subgoals then can be found at the states nearest 0.5.

An increasingly fine partition of the task space can emerge by including more thresholded eigenvectors. Thresholding eigenvectors $2-k$ will return a maximum of 2^{k-1}. Eigenvectors can be added until a sufficiently fine partition is obtained.

An iterative method is to use the initially identified subgoal to partition the environment, and then compute a new set of eigenvectors over the partitioned subspace. This recursive cutting leads to a segmentation of the task, but requires multiple computations of a first eigenvector (although the complexity of the problem is reduced by half each time).

Traditionally, the eigenvectors of the normalized graph Laplacian, a matrix frequently invoked in spectral graph theory[2], are used in spectral methods such as this one (rather than the eigenvectors of the SR). However, the eigenvectors of the normalized graph Laplacian matrix and the random walk SR matrix are approximately equivalent in cases where most nodes have the same degree, such as in spatial domains. We sketch a very informal proof below.

Given an undirected graph with symmetric weight matrix W and given D is a diagonal degree matrix with $D(s,s) = \sum_{s'} W(s,s')$, the graph Laplacian is given by $L = D - W$. The normalized graph Laplacian is given by $\mathscr{L} = I - D^{-1/2} W D^{-1/2}$.

For a random walk over this graph, the transition matrix is given by $T = D^{-1}W$. If ϕ is an eigenvector of the matrix $I - \gamma T$ and the degree is constant for all nodes, then ϕ is an un-normalized eigenvector of the normalized graph Laplacian. In a spatial domain, where the degree is constant everywhere except boundaries (almost everywhere as the discretization of the environment approaches continuity), the eigenvectors of $I - \gamma T$ will approach those of the normalized Laplacian. Since the SR, M, is equal to $(I - \gamma T)^{-1}$ and a matrix shares its eigenvectors with its inverse, the eigenvectors of the SR will also be similar to the eigenvectors of the normalized graph Laplacian.

Parameters

The parameters used for each simulation are described below, organized by figure (both main text and supplement).

Figure 1: Updating value following change in reward

A. Discount $\gamma = 0.75$.

B. Discretized to 15×15 grid. Discount $\gamma = 0.998$.

Figure 2: Comparison of place cell models

Tracks discretized to 500 states. 2D barrier environments discretized with 40×40. Parameters described below as fraction of boundary length, so 0.5 refers to a half of the track or wall length.

A. $\mu_{\text{track}} = 0.75$, $\sigma_{\text{track}} = 0.04$; $\mu_{\text{barr}} = \langle 0.50, 0.45 \rangle$, $\sigma_{\text{barr}} = 0.066$.
B. $\mu_{\text{track}} = 0.75$, $\sigma_{\text{track}} = 0.04$; $\mu_{\text{barr}} = (0.50, 0.45)$, $\sigma_{\text{barr}} = 0.066$.
C. $x_{\text{track}}(s) = 0.75$, $\gamma_{\text{track}} = 0.084$; $x_{\text{barr}}(s) = (0.50, 0.45)$, $\gamma_{\text{barr}} = 0.13$.

Figure 3: Illustration of retrodictive place fields

None.

Figure 4: Predictive skewing of place fields[3]

Discretized to 300 states. $\gamma = 0.9$, $p_{\text{right}} = 0.66$, $p_{\text{left}} = 0.34$.

Figure 5: Place fields near detour[4]

Discretized maze enclosed by 20×10 box, channel width = 1, discount $\gamma = 0.95$, softmax inverse temperature parameter $\beta = 5$.

Figure 6: Clustering of place fields

Discretization 32×30, discount $\gamma = 0.99$.
A. Random walk policy.
B. Random walk policy with no transition through boundaries.
C. Random walk policy with no transition through boundaries.
D. Rewarded location = $[14, 18] \times [13, 17]$, softmax policy given inverse temperature parameter $\beta = 1$.
E. Rewarded location 1 = $[20, 24] \times [10, 14]$, rewarded location 2 = $[5, 10] \times [20, 25]$, softmax policy given inverse temperature parameter $\beta = 1$.
F. Punished location 1 = $[20, 24] \times [10, 14]$, softmax policy given inverse temperature parameter $\beta = 1$.

Figure 7: Hippocampal representations in non-spatial task[5]

SR learned by TD learning with 10^4 step random walk, learning rate $\eta = 0.01$, discount $\gamma = 0.9$.

Figure 8: Hippocampal representations in spatio-temporal task[6]

Discount $\gamma = 0.98$. Similarity between states computed as the correlation between pairs of successor representations.

Spatial pairs
Low distance: [5,1; 5,4; 10,7; 6,5; 4,1]
High distance: [9,3; 10,1; 10,4; 7,3; 10,3]
Temporal pairs
Low distance: [5,1; 5,4; 8,1; 9,3; 4,1]
High distance: [5,2; 4,2; 6,2; 3,2; 9,2]
Both pairs
Low distance: [5,1; 5,4; 4,1; 6,5; 10,7]
High distance: [5,2; 6,2; 4,2; 3,2; 9,2]

SR learned by TD learning with 10^4 step random walk, learning rate $\eta = 0.01$, discount $\gamma = 0.9$.

Figure 9: Grid cells in different geometric environments

Square. Discretization 40×42. Eigenvector 37 shown. Split in half down the slightly longer edge (between 21 and 22) for the split halves analysis.
Hexagon. Diameter (distance between opposite vertices) 40. Eigenvector 51 shown.
Circle. Diameter 40. Eigenvector 31 shown.

Trapezoid. Base 40, height 80, top 8. Eigenvector 52 shown. Split two thirds of the way between the short edge and long edge of the trapezoid (between 53 and 54) for the split halves analysis to match results to Krupic *et al.* (2015).

Histograms. Includes all first 120 eigenvectors for which ellipses could be fit to the 6 central peaks of the spatial autocorrelation (defined as the 6 peaks nearest the center). This eliminated fields with very large scales (not enough peaks) or fields where the fit ellipse was a degenerate line (all peaks in a row). To compute grid similarity, the spatial autocorrelation was computed for both halves. The circular window used to compare autocorrelations had radius of 15 (since the dimensions of the autocorrelation are twice that of the grid field, the window never went off the edge of the autocorrelation map). The autocorrelation outside this window was set to zero for similarity comparisons. "Similarity" between two halves of a grid field refers to the Pearson correlation coefficient of the windowed autocorrelations on each half of the maze.

Figure 10: Grid fragmentation in compartmentalized maze[7]

Discretization 40×40, 10 alternating channels of width 4. Eigenvectors (top row) 32, 13, (bottom row) 14, 22 shown. Eigenvectors 20 through 120 used for similarity matrix to exclude low-frequency, non-grid eigenvectors.

Figure 11: Grid fields in multi-compartment environment[8]

Discretization 16×12 with the two square compartments each discretized at 8×8, the connecting passageway at 16×4, and the opening to the passageway 6 nodes wide.

Local model initialization. Corresponding states in the different rooms were mapped to the same index in the adjacency matrix to simulate local state aliasing as the initial condition. SR discount $\gamma = .9$, SR learning rate $\alpha = 0.1$ for the first 500 pre-training epochs and 0.01 for the next 500, SGD learning rate 0.0005, 1000 epochs total of 50 timesteps each.

Global model initialization. The SR was initialized to the local SR model learned, then constraint that corresponding states in the different rooms be the same was removed so that the rooms could slowly differentiate. SR discount $\gamma = .9$, SR learning rate $\alpha = 0.01$, SGD learning rate 0.0005, 1000 epochs total of 50 timesteps each.

Figure S1: Model Free versus SR

Channel width 1, channel lengths 4. Discount $\gamma = 0.98$, learning rate $\eta = 0.01$, noise $\varepsilon \in [-0.05, -0.05]$, 100 epochs of 500 timesteps for each reward change, random walk policy.

Figure S2: Multiscale representations

Discretization 20×20. Multicompartment environments have same discretization parameters as described for Figure S1. Points mapped to 2D using multidimensional scaling.

Figure S3: Spectral Regularization

Discretization 12×12, discount $\gamma = 0.95$, learning rate $\eta = 0.1$, noise $\varepsilon \in [-0.1, -0.1]$, 20 epochs of 500 timesteps, $k = \frac{3}{4} n_{\text{states}} = 108$ components used for spectral and Fourier regularization.

Figure S4: Subgoals for hierarchical RL

Circles marking subgoals were placed where the eigenvector with the corresponding color was 0. Lines were drawn when a whole contour was equal to zero.

A. Multicompartment environment I. Discretized at 20×20. Doorways placed at: (10,10), (5,8), (5,14). Doorway width 2.

B. Multicompartment environment II. Discretized at 20×20. Doorways placed at: (10,4), (10,16), (5,8), (5,14). Doorway width 2.
C. 2-Step tree maze. Channel width 5, first channel length 10, other channel lengths 5.

Figure S5: Context Preexposure Facilitation
Discretization 10×10, punished location $R(5,5) = -1$, discount $\gamma = .9$; 10 simulation trials shown, 2×10^4 timesteps.

Figure S6: SR Place fields
See parameters for Figure 5. Shown are place cells for the following states in the $32 \times 30 = 960$ state environment: 178 464 422 611 418 363 728 755 665 303 907 33 266 45 94 789.

Figure S7: Effects of rewarded policy on SR eigenvectors
Discretization 32×30, rewarded location $R(16,15) = 1$, discount $\gamma = .98$, softmax inverse temperature parameter $\beta = 3$ for optimal policy. First 64 eigenvectors shown.

Figure S8: Additional grid fields
See parameters for Figures 7, 8, and 9.

Figure S9: Additional grid fields
See parameters for Figure 8.

Figure S10: Ground truth and learned eigenvectors
See parameters for Figure 9.

References

1. Shi, J. & Malik, J. Normalized cuts and image segmentation. In *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, 888–905 (IEEE, 2000).

2. Belkin, M. & Niyogi, P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In Dietterich, T., Becker, S. & Ghahramani, Z. (eds.) *Advances in Neural Information Processing Systems 14*, 585–591 (MIT Press, 2002).

3. Mehta, M., Quirk, M. & Wilson, M. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron* **25**, 707–715 (2000).

4. Alvernhe, A., Save, E. & Poucet, B. Local remapping of place cell firing in the Tolman detour task. *European Journal of Neuroscience* **33**, 1696–1705 (2011).

5. Schapiro, A. C., Turk-Browne, N., Norman, K. & Botvinick, M. Statistical learning of temporal community structure in the hippocampus. *Hippocampus* **26**, 3–8 (2016).

6. Deuker, L., Bellmund, J., Schröder, T. & Doeller, C. An event map of memory space in the hippocampus. *eLife* **5**, e16534 (2016).

7. Derdikman, D. *et al.* Fragmentation of grid cell maps in a multicompartment environment. *Nature Neuroscience* **12**, 1325–1332 (2009).

8. Carpenter, F., Manson, D., Jeffery, K., Burgess, N. & Barry, C. Grid cells form a global representation of connected environments. *Current Biology* **25**, 1176–1182 (2015).

Supplemental Figures

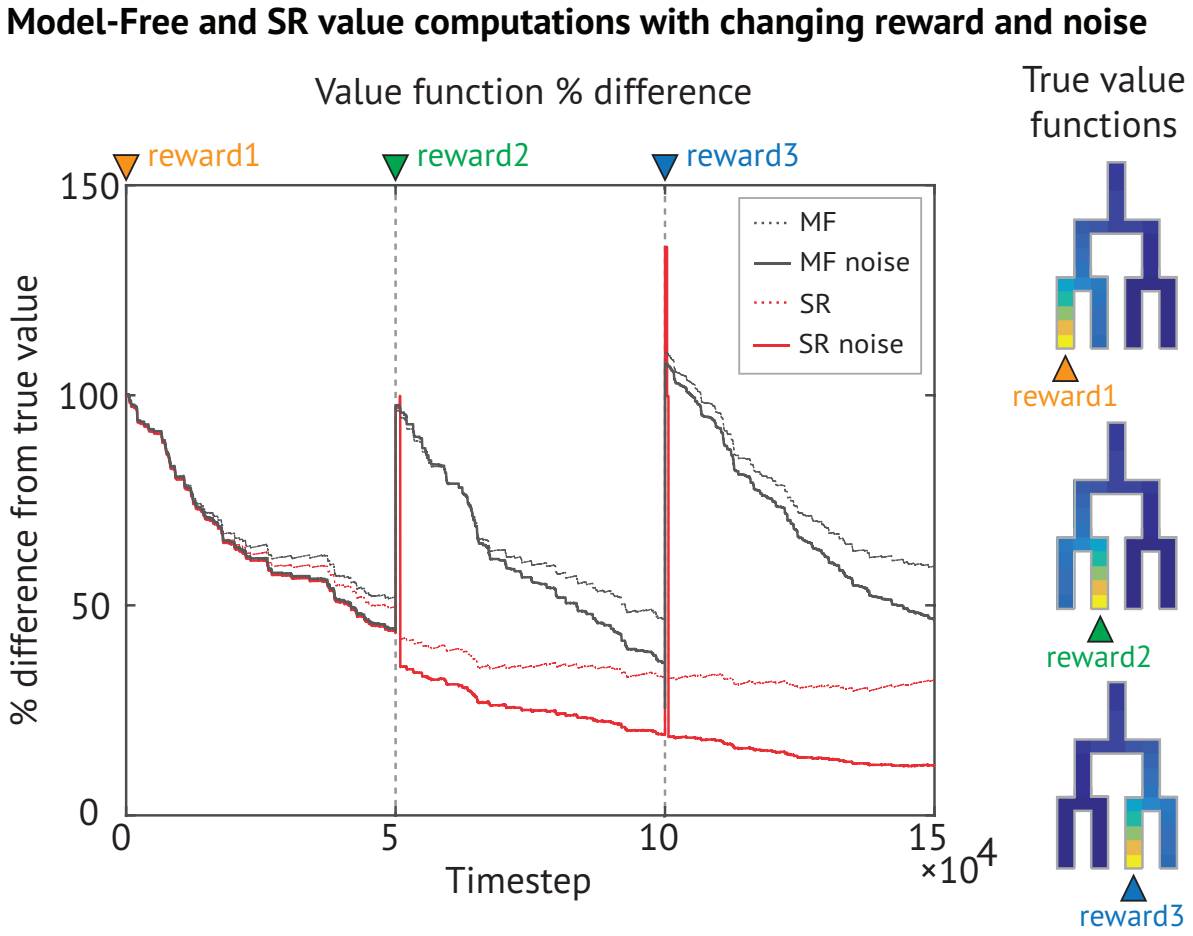

Figure S1. (A) Comparison between learning a value function with model-free (MF) TD learning and with SR TD learning under a random walk policy with changing reward location. For an MF agent, the value function must be entirely relearned each time the reward changes location. For an SR agent, the error will jump with reward change but quickly drop as soon as the new reward is found. Furthermore, the SR provides a slight advantage over MF learning when there is noise in the state signal. The error is more than 100% when the animal has neither learned the current value function nor unlearned the previous value function. The value functions corresponding to each reward location are shown on the right.

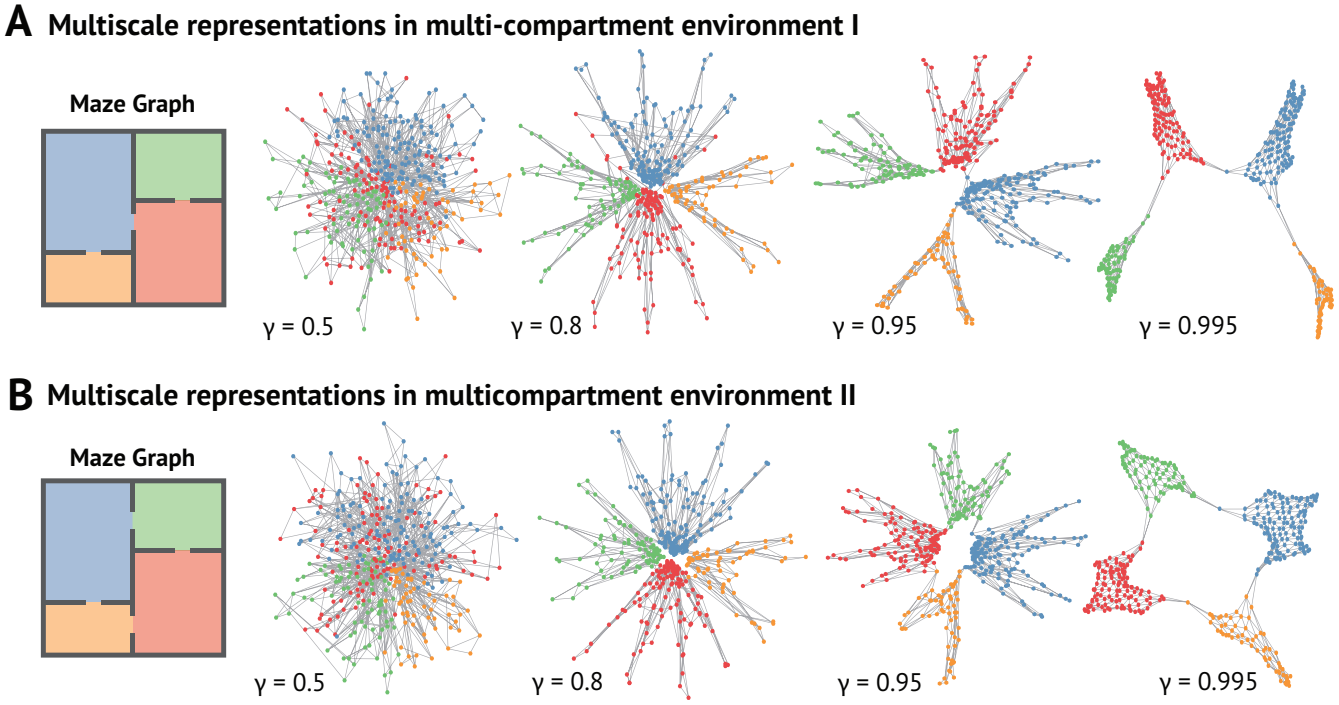

Figure S2. Multi-dimensional scaling projections of the SRs for different states in multicompartment environments. The two environments feature the same compartments connected in different ways. These plots show a 2D projection that maximally preserves Euclidean distances among SR vectors for each state. From left to right, we increase the discount factor γ used to compute the SR. The larger planning horizon ($\gamma = 0.995$) causes states that predict each other on long time scales to cluster. This exposes the long-timescale, low-dimensional organization of the environment. The shorter time horizons capture With small values of the discount parameter, the macroscale structure of the maze is not readily apparent in the low dimensional embedding; rather, states in the same compartment are represented distinctly. Simultaneously representing predictive representations over a range of planning horizons gives rise to a multiscale, hierarchical embedding of states. We can thus visualize how a range of discounts expressed along the hippocampal longitudinal axis may enable hierarchical planning and planning at different timescales.

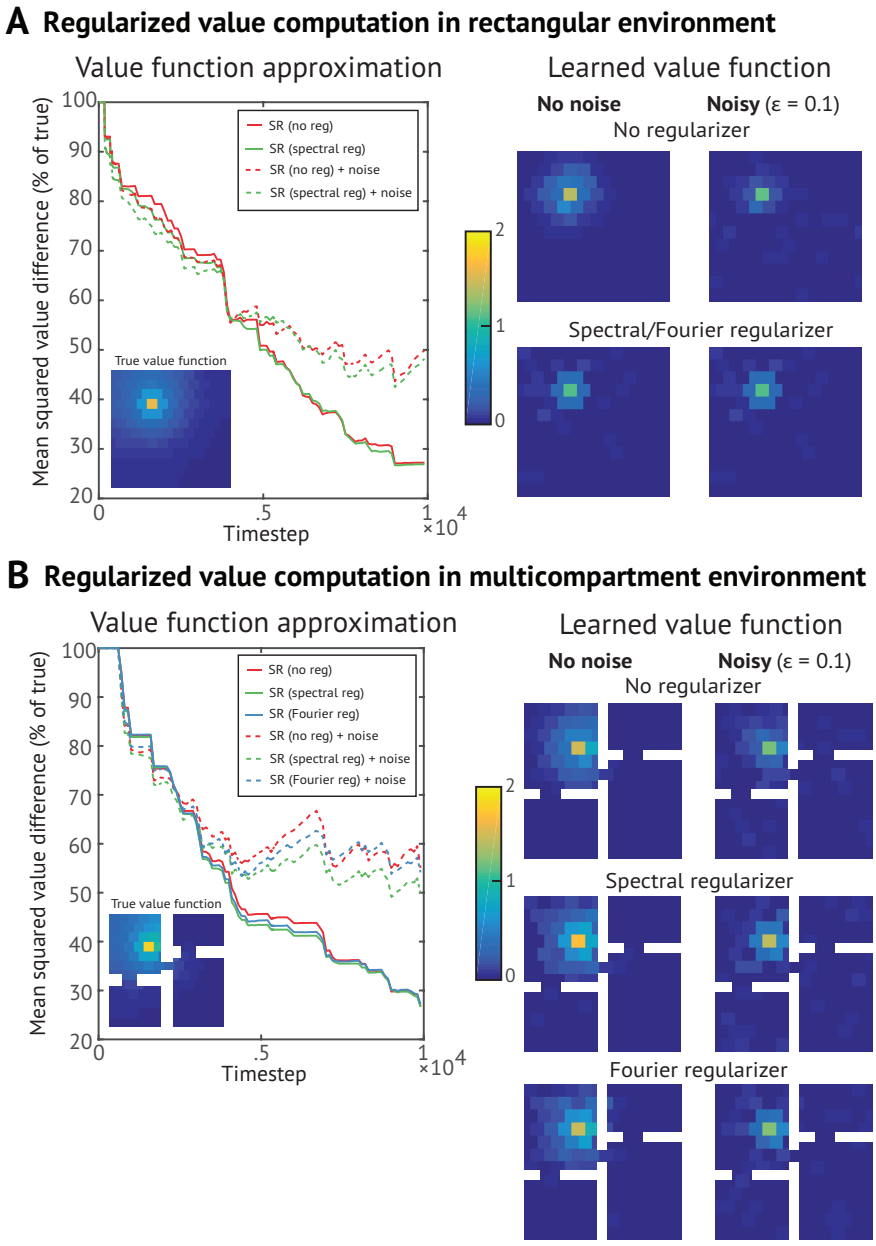

Figure S3. (A) Comparison between SR TD learning of the value function with and without spectral regularization in a rectangular environment. The value converges slightly faster with the smoothing effects of the regularizer, especially when noise is injected into the state signal. On the right are value functions for each condition. We can see that the regularized value functions are smoother. (B) Comparison between SR TD learning of the value function with different types of regularization in a multi-compartment environment. We once again see that learning is facilitated by spectral regularization, especially in the noisy condition. We also explored the effects of using a Fourier regularizer that is insensitive to the topology of the particular environment. This provides an improvement over no regularization, but is still inferior to spectral regularization.

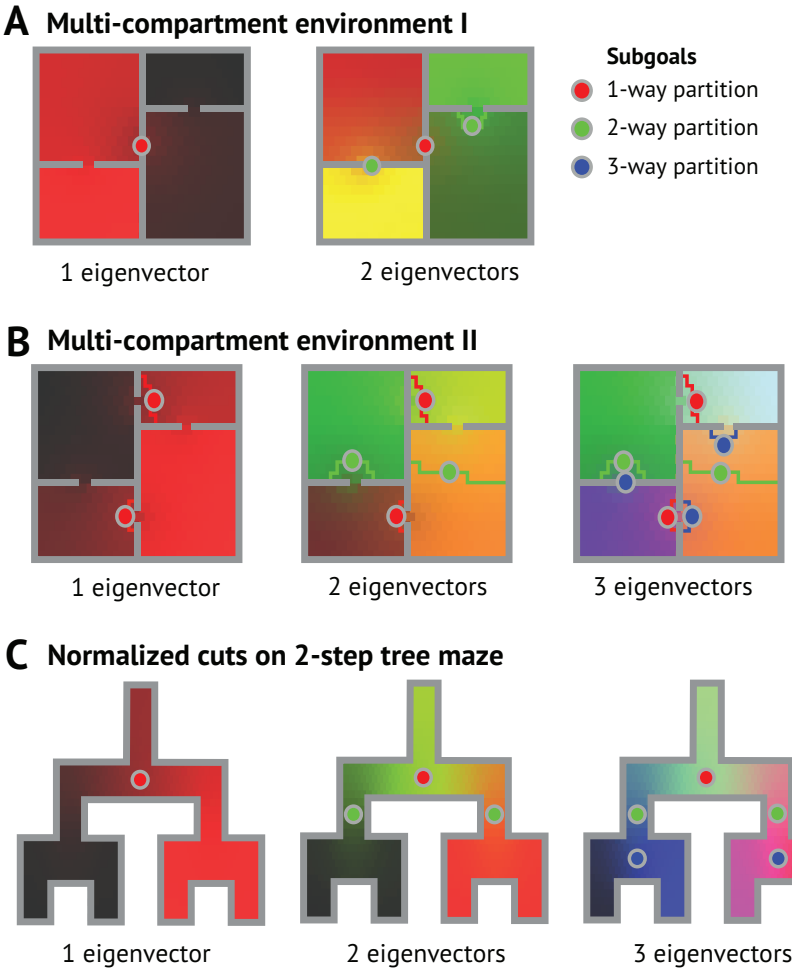

Figure S4. *Subgoals for hierarchical RL.* (A) The eigenvectors with largest eigenvalues can be used to partition the task space, with boundaries between the partitions producing useful subgoals. In compartmentalized environments, subgoal partitions tend to fall at or near doorways. The environment is colored by setting the RGB values of each point in the maze to the corresponding value of the 1st, 2nd, and 3rd eigenvectors used in the illustrated partition. (B) Different environment topologies give rise to different subgoals. The partition is not as clean when compartments are connected in a loop instead of hierarchically (as in A), but the doorways are still identified by the top 3 eigenvectors. (C) Extracted subgoals correspond to major decision points in a 2-step maze.

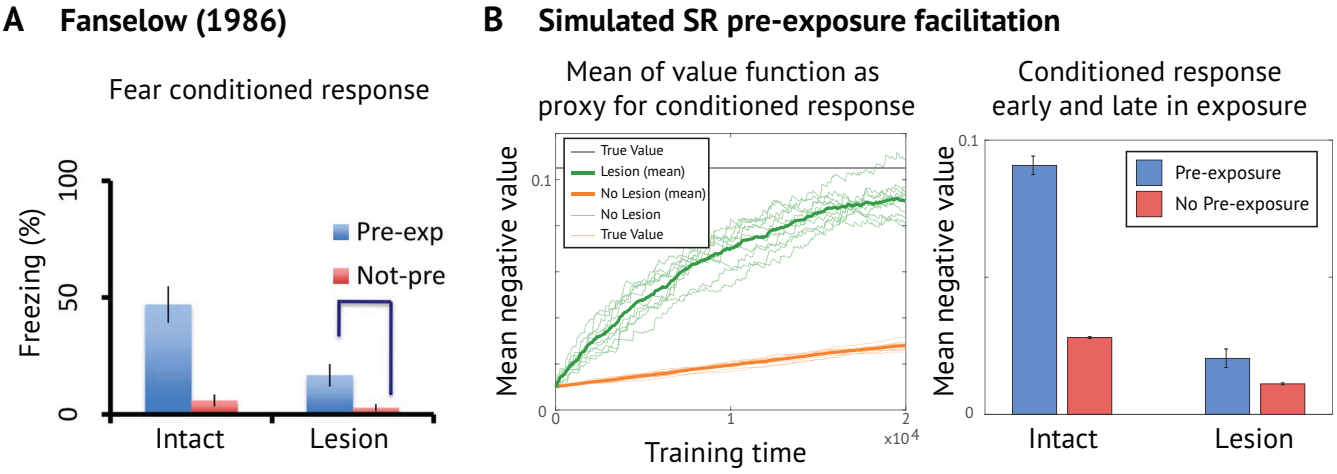

Figure S5. (A) Fear-related freezing response in rats following fear conditioning has been shown to be stronger if the animal is "pre-exposed" to the conditioning environment. This effect is much weaker following hippocampal lesions. (B) Under an SR interpretation, exploring the environment allows the hippocampus to learn a predictive representation. Since value is computed by multiplying the SR by the reward function, an under-developed SR will produce an underestimate of the negative value in the room. As the animal learns a model of how states in the environment predict each other, the value signal can be approximated more fully. Model-free learning does not predict this, since the reward signal cannot propagate through the environment without experience preceding the time of shock.

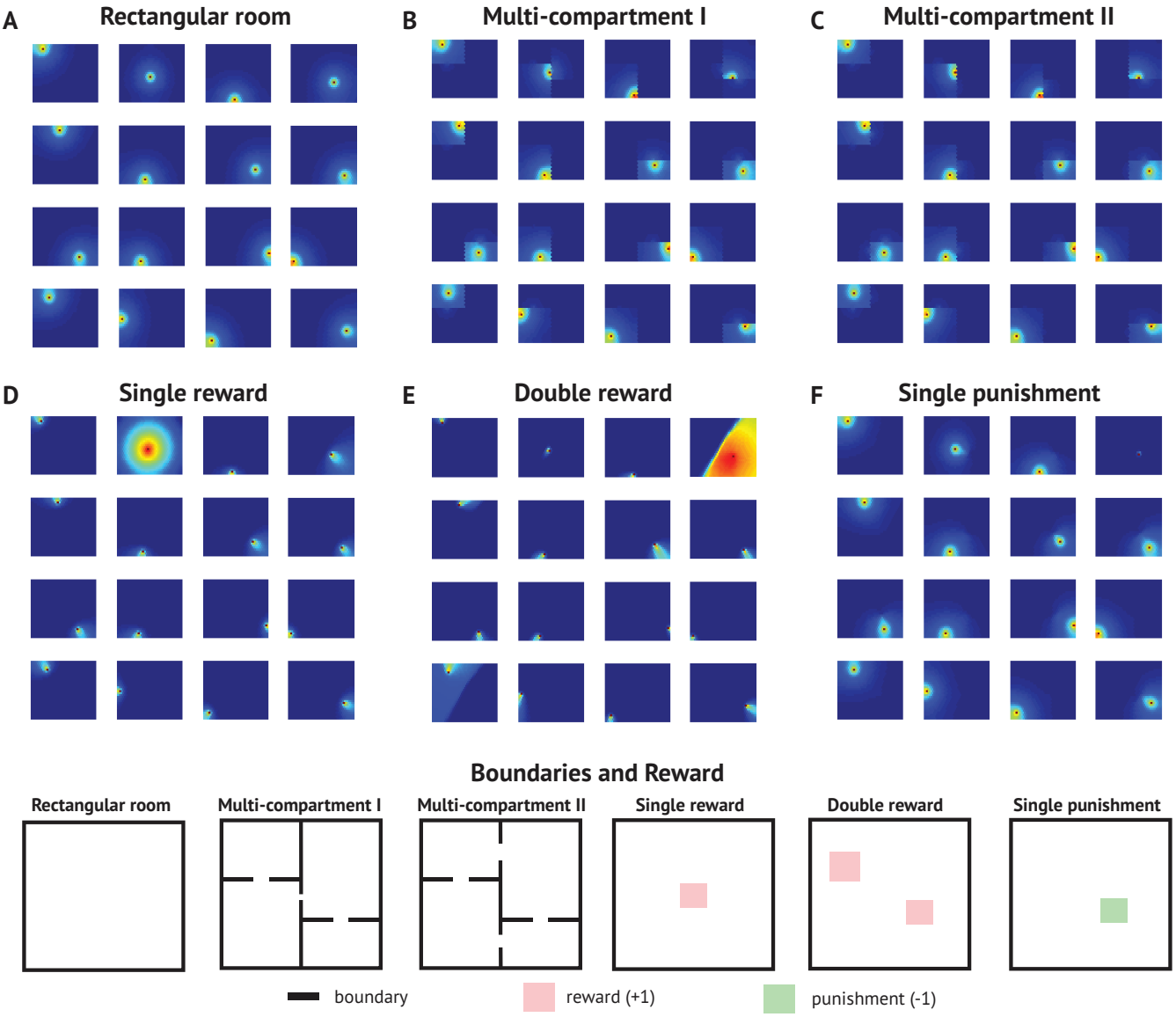

Figure S6. Example SR place fields corresponding to environments and reward configurations described in Figure 5 of main document. (A) SR fields are radially symmetric, gradually decaying circles under a random walk. (B,C) SR fields are constrained to remain within the compartments of the environment. (D,E) Place fields near the rewarded locations swell to include the many states that predict them, those further away skew backwards. (F) Most place fields are unaffected, those near the punished region skew slightly towards it to account for fleeing from the punished area.

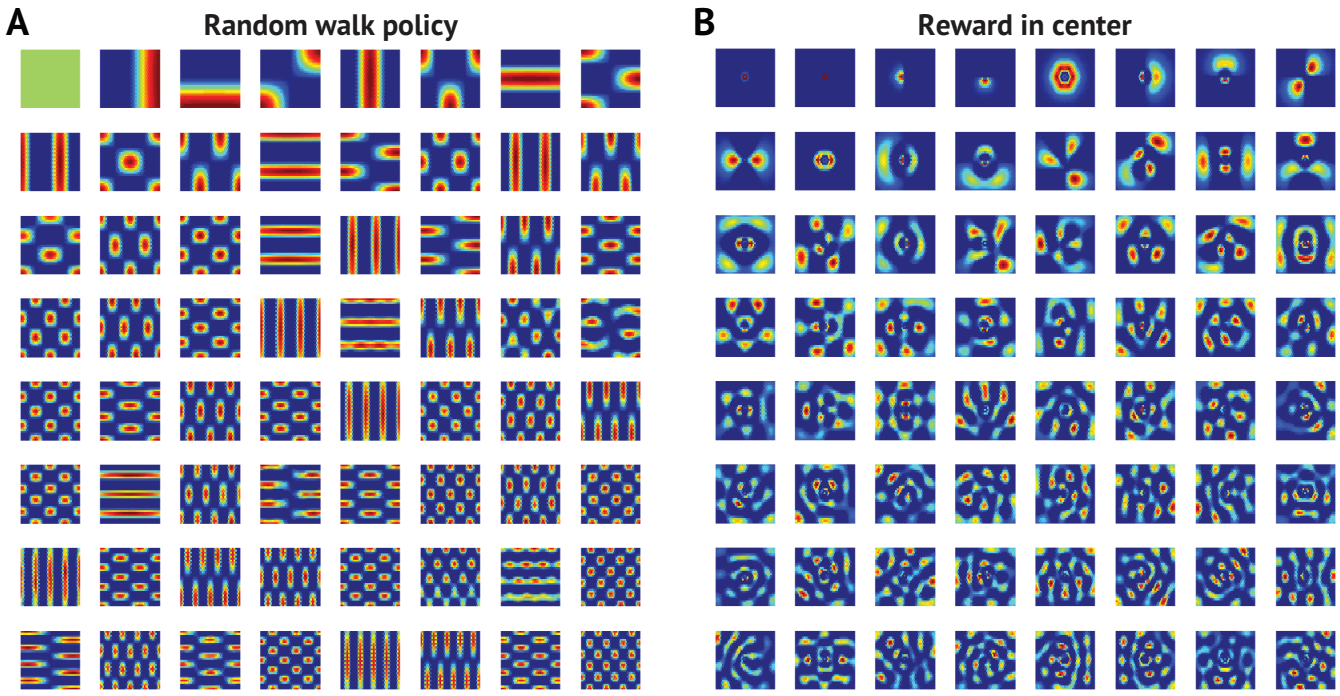

Figure S7. (A) SR eigenvectors under random walk policy and (B) under optimal softmax policy with reward in center of room. The attractor state induced in the policy by the rewarded location warps the SR eigenvectors.

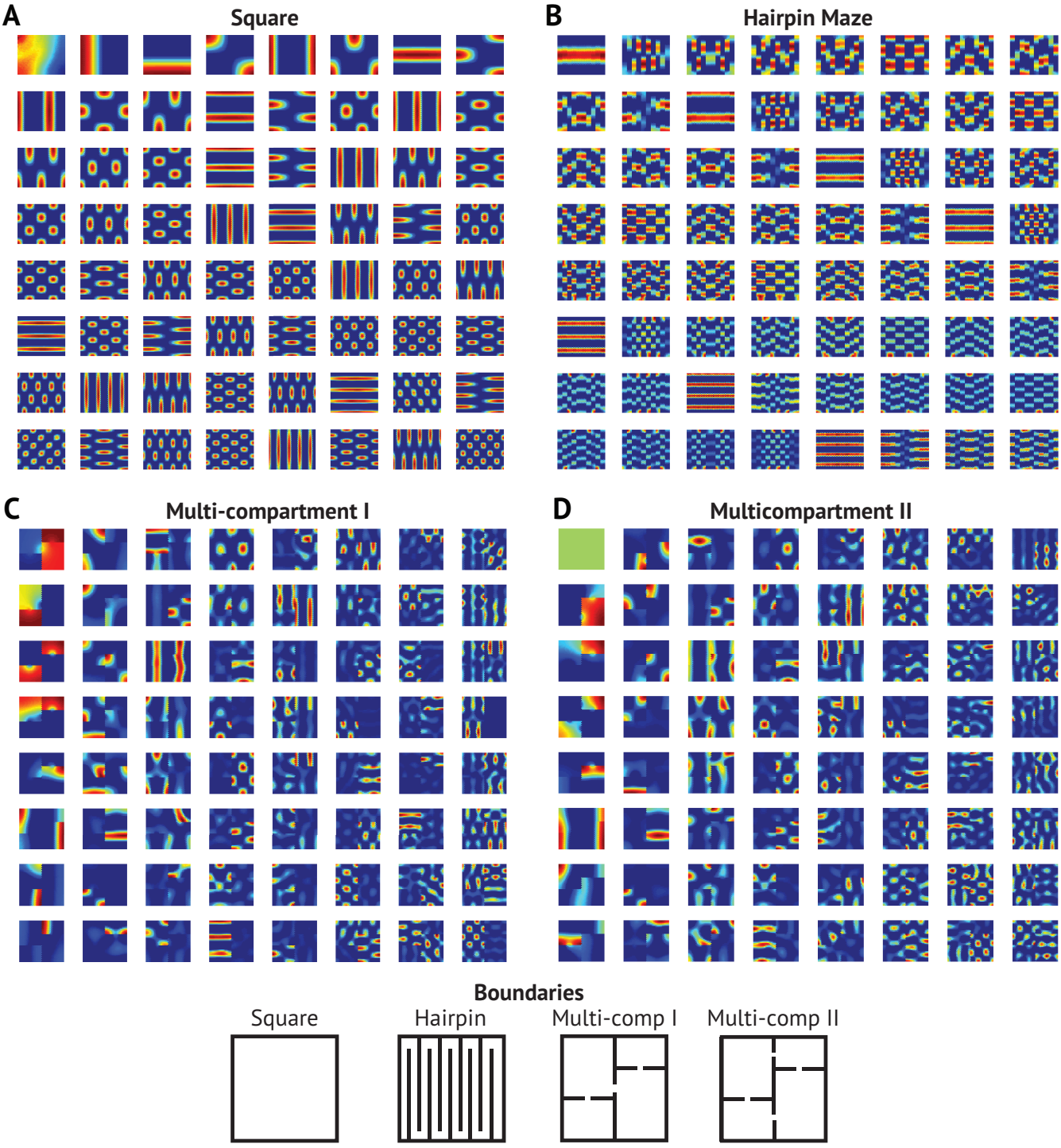

Figure S8. Additional eigenvector grids fields in rectangular and multi-compartment environments (see Fig. 9, 10 in main document and Fig. S3, S4).

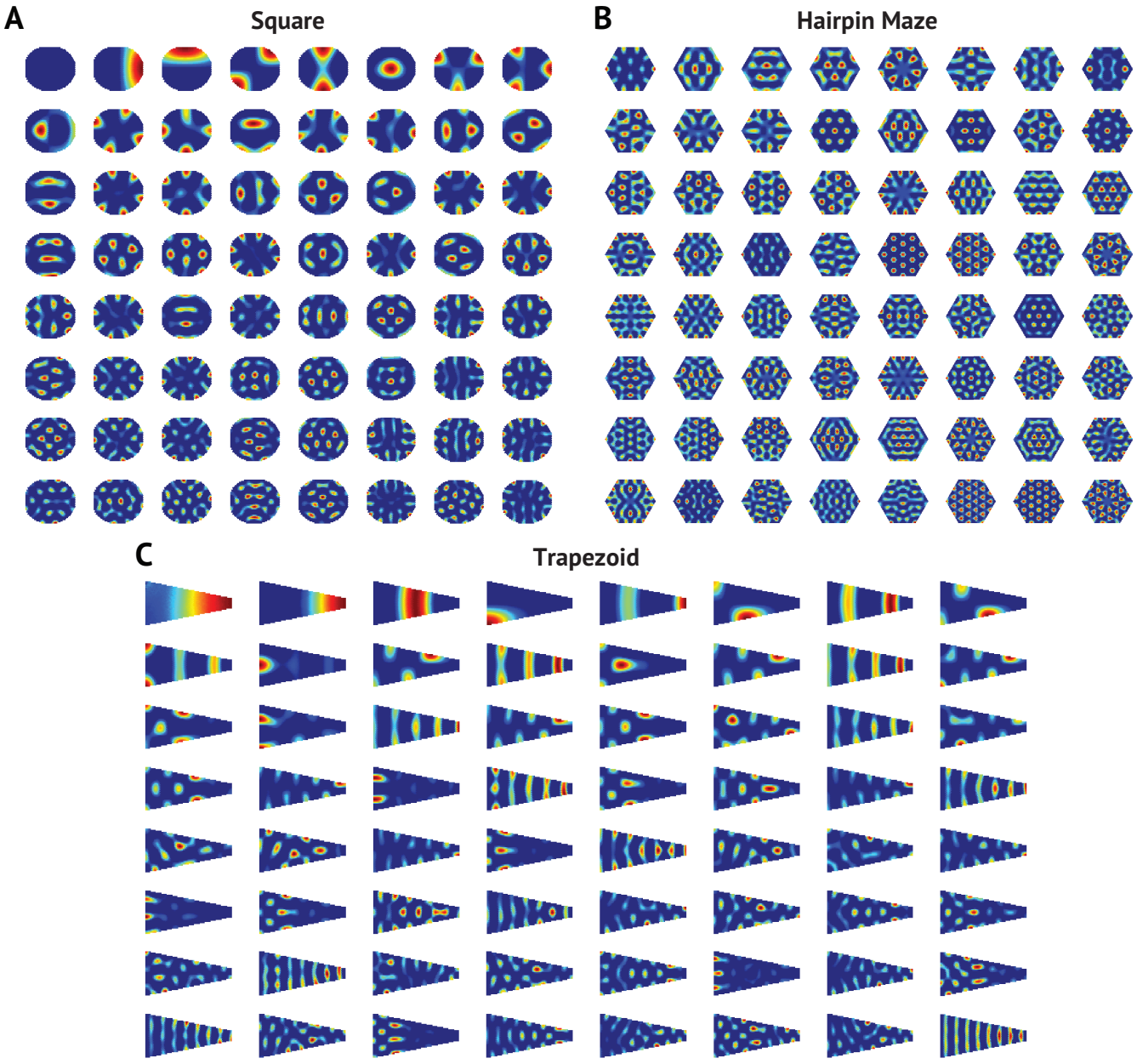

Figure S9. Additional eigenvector grid fields in non-rectangular geometric environments (see Fig. 9 in main document).

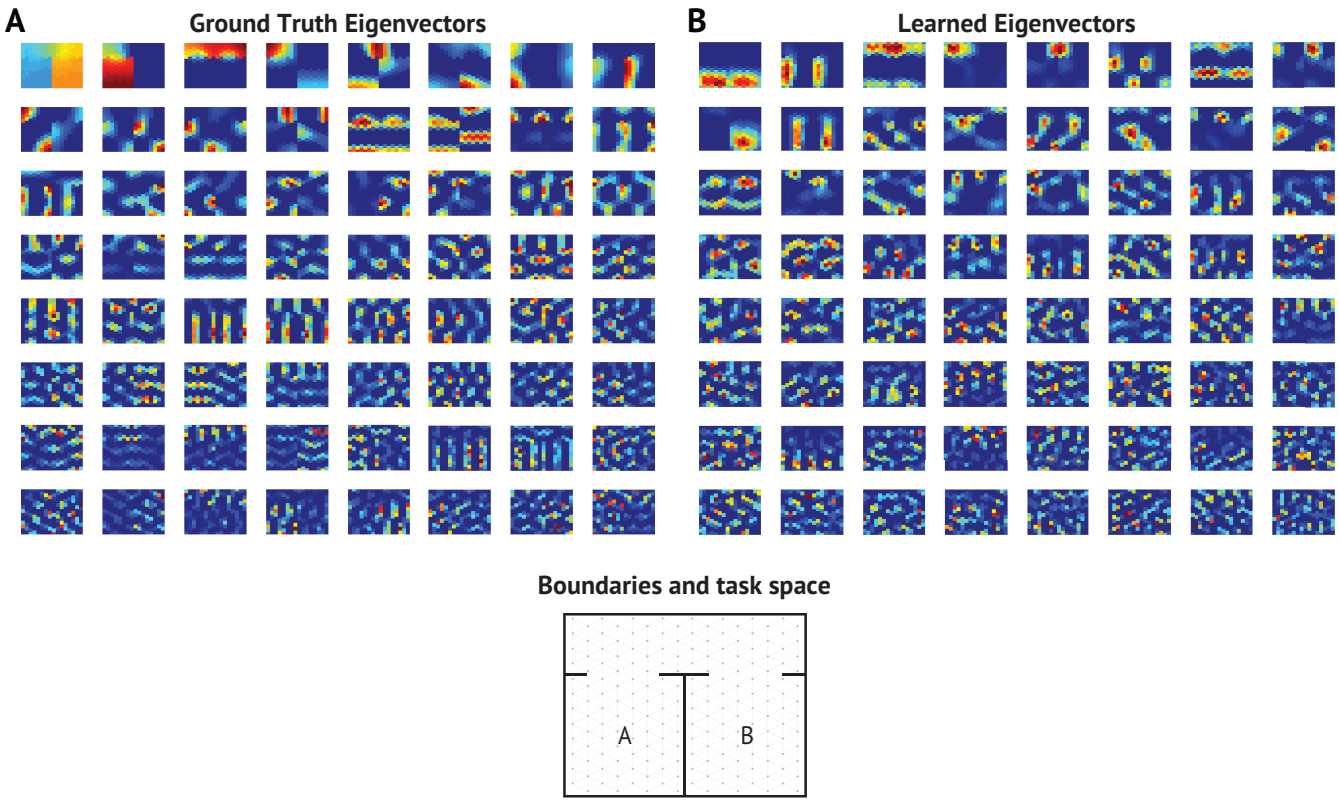

Figure S10. Additional (A) ground truth SR eigenvector grid fields and (B) learned SR eigenvector grid fields for two compartment environment described in Fig. 11 in the main text.