Flexible navigation with neuromodulated cognitive maps

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Animals naturally form personalized cognitive maps to support efficient navigation and goal-directed behavior. In the brain, the CA1 subregion of the hippocampus plays a key role in this process, hosting spatially tuned neurons that adapt based on the behavioral context and internal states. Computational models of this ability include labeled graphs with locally specified spatial information, which avoid global metric structure, and deep neural networks trained on spatial tasks that exhibit emergent spatial tuning. However, these approaches often struggle to model one-shot adaptive mapping and typically rely on plasticity rules that lack biological plausibility.

We propose a neural architecture inspired by place-cell dynamics that enables rapid on-the-fly construction of cognitive maps during exploration of novel environments. The model relies on velocity inputs and grid cell modules to generate spatial representations and integrates neuromodulatory signals responsive to boundaries and rewards. Learning combines synaptic plasticity, lateral inhibition, and modulatory gating of place-cell activity. For reward-driven navigation, the agent uses a graph-based algorithm to plan paths on the emergent cognitive map, treating place cells as nodes in a locally structured graph.

We tested the model on different environments, achieving high sample efficiency and solving tasks in a single episode, for which usual RL agents require thousands of training steps. This performance advantage arises from biologically inspired inductive biases embedded in the model architecture. In simulation, the agent adapts to dynamic reward locations and changes in the environment layout. Ablation experiments and analysis of neuromodulated place cells reveals task-dependent changes in tuning field size and spatial density, aligning with experimental findings from hippocampal recordings. These results highlight the promise of biologically grounded computation and locally structured graph representations for flexible and data-efficient cognitive mapping.

1 Introduction

Survival in complex environments requires efficient navigational strategies. From desert ants to humans, successful wayfinding—navigating toward goals that are not directly visible depends on emergent internal spatial representations, known as cognitive maps [1, 2]. Understanding how these maps are constructed from ongoing experiences, and how they can be exploited for flexible goal-directed navigation remains an active area of research in both neuroscience and reinforcement learning (RL).

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

The hippocampus (HP) and entorhinal cortex (EC) serve as the primary neural substrates for spatial representation in the brain. They containing specialized neurons that encode spatial and contextual information—including grid, border, speed, and place cells [3, 4, 5]. Place cells in the CA1 region have attracted particular interest due to their convergence of inputs from periodically tuned grid cells, the CA3 region, and the lateral EC [6, 7, 8, 9]. This strategic integration of diverse spatial and contextual signals suggest CA1 place cells may play a critical role in the formation and maintenance of cognitive maps [10].

Another important component are neuromodulators. Their actions include modulation of neuronal dynamics, for instance adjusting the synaptic strength, tuning place fields [11, 12, 13]. Further, neuromodulators such as dopamine transmit reward signals reshaping place cell tuning [14, 15, 16, 17], support novelty detection [12], and encode prediction errors [18, 13], particularly via LEC inputs [19, 20]—mechanisms closely related to reinforcement learning principles.

Traditional cognitive map theories propose multiple strategies for spatial navigation based on map-like representations. Route learning encodes paths as sequences of action–position pairs, but it limited in scalability and generalization, especially at route intersections [21, 22, 23]. In contrast, survey maps, which rely on Euclidean geometry, offer greater flexibility [22, 24]; however, their strong geometric assumptions often conflict with neural and behavioral evidence pointing to geometric distortions and topological biases in spatial neural representations [25, 26, 27, 28]. As a middle ground, labeled graphs encode landmarks and transitions within a topological network, enabling vector-like operations, planning, and prediction [29, 30, 31].

Computational models have captured some of these aspects individually, showing new ways the brain might use for addressing spatial navigation tasks. Early work proposed that the hippocampus encodes spatial position and direction [32], while topological graph models based present scalability challenges [23]. More recent approaches draw inspiration from predictive coding and reinforcement learning, including successor representations and the Tolman-Eichenbaum Machine, which generalize across spatial and relational tasks while mimicking biological neural activity patterns [33, 34, 35]. Path integration models trained on velocity inputs give rise to spatial-like receptive fields [36, 37, 38]. Others incorporate reward-driven Hebbian plasticity modulated by neuromodulators [39]. However, these architectures mostly fail to unify these ingredients into a biologically grounded system that at the same time learns a map of the environment online without relying on an external coordinate system, and flexbility perform goal-directed navigation.

In this work, we present a biologically inspired model of cognitive map formation that integrates place cell representations, neuromodulatory signals, and graph-based spatial computations. Our aim is to demonstrate an architecture capable of building a content-rich topological map of the environment on the fly, and leveraging it for efficient, goal-directed navigation—without requiring offline training.

Critically, neuromodulators play a central role, as they form scalar fields over the map [40], drive local Hebbian plasticity in response to sensory updates [41, 18, 42], and support the formation and adaptation of reward-modulated neural representations used for planning [43, 44, 45, 46, 34]. We show how this system dynamically adapts to environmental changes and how neuromodulation shapes place field allocation and remapping [47, 48], linking cognitive flexibility to underlying physiological mechanisms.

The remainder of the paper is organized as follows: Section 2 details the model and experimental setup; Section 3 presents results; Section 4 discusses broader implications and future directions.

2 Methods

We propose a model of cognitive map formation driven by an agent's experience within a closed environment.

The architecture operates with minimal external inputs—limited to binary reward and collision signals—as illustrated in Figure 1**a**. Instead of relying on exteroceptive cues, spatial representations emerge from idiothetic information, i.e., the agent's internal perception of self-motion [49], consistent with prior path integration frameworks. Concretely, we use the agent's ground-truth velocity vector, *i.e.* its actual displacement within the environment, as the primary navigational signal, reflecting the integration of inertial and proprioceptive cues observed in biological systems [50, 51]. Since no visual information is used, the agent is effectively navigating in the dark.

Place Cell Formation The primary spatial representation is formed by a set of grid cell modules, each encoding a periodic tiling of 2D space, which directly maps to a toroidal manifold $\mathbf{T}^2$ (Fig. 1**b**). Departing from traditional grid cell modeling approaches [52, 53], we generate population activity directly via Gaussian tuning over the torus, continuously updated using the agent's velocity vector—an approach used in prior work [8].

The grid cell population vector $\mathbf{u}^{\mathrm{GC}}$ is forwarded to a place cell network with initially zeroed synaptic weights. When no place cell is sufficiently active for a given input, a silent unit is randomly selected and imprinted with the current grid activity pattern. To enforce representational sparsity and tuning specificity, lateral inihibition is implemented by comparing the cosine similarity between the new weight vector and any existing one and a threshold $\theta^{\mathrm{PC}}_{\mathrm{inh}}$.

Each place cell's activation is computed via a bounded cosine similarity function, determining its corresponding place field (Fig. 1**d**). Further implementation details, including lateral inhibition and recurrent connectivity, are provided in the Appendix.

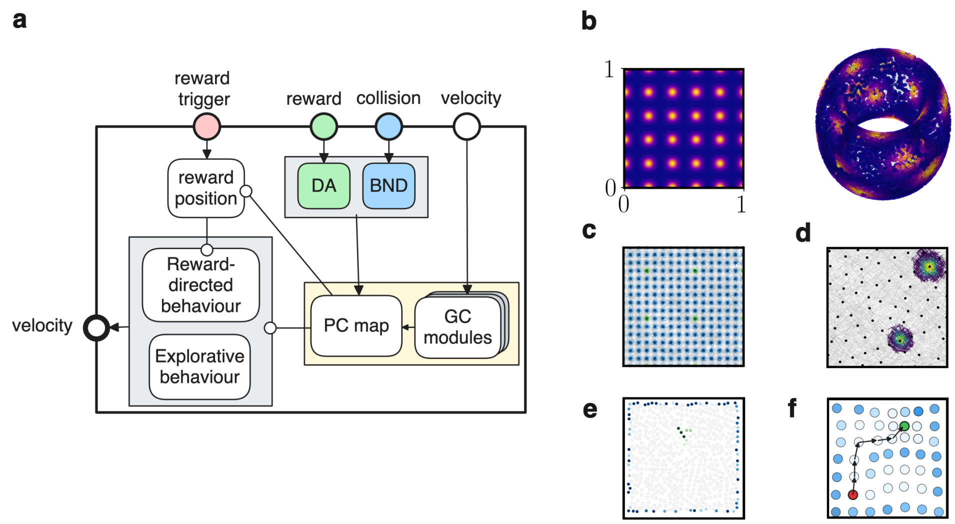

Figure 1: **Model layout and spatial representations** - **a**: *the full architecture of the model, consisting of three main sensory input, targeting the two modulators and the cognitive map module, and the executive components, represented by a policy module, two behavioral programs and a reward receiver*. **b**: *a module of grid cells defined in a bounded square space of length 1, and an activity representation of their receptive field over a torus*. **c**: *the neural activity of a grid cell module from a random trajectory; in blue the repeating activity of all cell, while in green the activity of only one, highlighting the periodicity in space*. **d**: *the distribution in space of the place cells centers, together with the activity of two cells showing the size of their place field*. **e**: *neuromodulation activity over the place cells map, with in blue the cells tagged by the collision modulation, and in green the ones targeted by reward modulation*. **f**: *the place cells layer can be regarded as a graph with values assigned to each node according to the modulation strength; a path-finding algorithm can then be used to connect any two nodes taking into account the node values*.

Neuromodulation Neuromodulators deliver event signals: rewards, denoted DA (for dopamine), and boundary collisions, denoted BND. They are driven by binary inputs and are defined through a leaky variable with exponential decay.

To remain resilient to environmental changes (e.g., moving rewards), the model uses a predictive mechanism to correct keep internal representations updated. Each modulator k updates synaptic weights to place cells through Hebbian plasticity:

$$\Delta \mathbf{W}^{\mathrm{k}} = \eta^{\mathrm{k}} \mathbf{u}^{\mathrm{PC}} \left(\mathrm{v}^{\mathrm{k}} - \mathbf{W}^{\mathrm{k}} \right) \tag{1}$$

The term in brackets can be regarded as an error, implementing a simple form of predictive coding and is inspired by temporal-difference learning [54], aligning with evidence that neuromodulatory systems signal prediction errors and update beliefs [55, 34, 56].

Weight vectors are constrained to remain non-negative. Reward modulation tags cells near rewarded locations, while boundary modulation builds a representation of environmental edges. These scalar fields form the core of the cognitive map (Fig. 1**e**). See Appendix for full learning rules and parameter settings.

Modulation of Place Fields We further tested whether neuromodulators could directly alter spatial tuning. Place fields were dynamically shifted and resized based on recent salience signals.

Following a salient event (reward or collision), place field centers were displaced in grid cell space with magnitude scaled by the neuromodulator v^{k} and proximity to the event:

$$\Delta \mathbf{W}_{\mathrm{i}}^{\mathrm{GC,PC}} = \mathrm{c}^{\mathrm{k}} \mathrm{v}^{\mathrm{k}} \varphi_{\sigma^{\mathrm{k}}} (\mathbf{u}^{\mathrm{GC}} - \mathbf{W}_{\mathrm{i}}^{\mathrm{GC,PC}}) \tag{2}$$

Here, φ is a Gaussian function with width σ^{k}, and c^{k} a scaling factor. This rule is inspired by BTSP plasticity [16, 47], which shifts CA1 place fields following salient experiences. This action was applied only to recently active cells, namely with an activity trace greater than a threshold θ^{k}. Futher, lateral inhibition prevents field overlap during remapping.

In addition to dislocation, field size was modulated by scaling the gain of recently active neurons. Such mechanism allows neuromodulators to transiently enhance or suppress spatial sensitivity for specific cells. This modulation rule involves the gain β_{i} of each cell being adjusted proportionally to its activity trace m_{i}, a reference gain constant $\bar{\beta}$, and a modulatory scaling variable:

$$\beta_{\mathrm{i}} = \mathrm{c}_{\mathrm{a}}^{\mathrm{k}} \mathrm{m}_{\mathrm{i}} \bar{\beta} + (1 - \mathrm{m}_{\mathrm{i}}) \bar{\beta} \tag{3}$$

where $\mathrm{c}_{\mathrm{a}}^{\mathrm{k}}$ is a scaling gain parameter, for which a value of 1 signifies that no modulation takes place.

Policy and Behavior To evaluate the model's utility in navigation, we implemented a simple policy toggling between exploration and reward-seeking behavior, depending on an external reward trigger and the internal map.

Exploration consisted of two possibile strategied: a random walk, for purely stochastic movements, and periodic goal-directed navigation towards a random but visited location, aimed at preventing stagnation. In contrast, exploitation—defined as reward-directed navigation—involved identifying the reward location within the cognitive map. This location corresponded to the average position of DA-modulated place cells, reflecting mechanisms such as hippocampal replay and value-based navigation [57, 58, 59]. A graph-based pathfinding algorithm was then used to compute the route from the current location to the reward. In this graph, place cells served as nodes, while synaptic connections acted as edges. Additionally, a cost function was introduced over the graph nodes, assigning lower values to BND-modulated cells. This was designed to discourage proximity to the environment's walls, as shown in plot 1**f**.

Table 1: Comparison of neural network models for spatial navigation and representation

Model	Architecture	Training method	Ext. C.
Banino et al. [36]	LSTM + linear layers + CNN	BPTT and deep RL, supervised	Yes
Cueva et al. [38]	RNN + linear layers	Hessian-free algorithm with regularization	Yes
Sorcher et al. [60]	RNN + linear layers	Backpropagation with regularization	Yes
Whittington et al. [35]	Attractor network and deep networks	Backpropagation and Hebbian learning	No
de-Cothi et al. [34]	Successor representation	TD-learning + eligibility traces	Yes
Brozsko et al. [61]	Spike Response Model	Online modulated Hebbian plasticity	Yes
Ours	Rate layers	Online neuromodulated plasticity	No

Model	Task	Input	Output
Banino et al.	Path integration, goal navigation	Velocity, visual input, reward	PC, HDC
Cueva et al.	Path integration	Velocity	Position
Sorcher et al.	Path integration	Velocity	PC
Whittington et al.	Relational graph knowledge	Observation and action	Observation
de-Cothi et al.	Planned navigation	Observation	–
Brozsko et al.	Goal navigation	Position, reward	Action
Ours	Goal navigation	Velocity, reward, collision	Action

Note: PC = Place Cells, HDC = Head Direction Cells, Ext. C. = External Spatial Coordinates

Comparison with previous architectures Several previous computational models show structural and conceptual similarities with the present work. A prominent category among them employs deep neural networks—often with recurrent components—and relies on gradient-based learning strategies such as backpropagation through time. These models typically require multiple training episodes or large datasets for convergence. In contrast, our model adopts biologically inspired, synaptically local plasticity rules, and requires only a single training episode for adaptation.

Other models utilize spiking neurons [61] or explicit neural representations [34], and incorporate online learning rules more closely aligned with ours. These models also focus more directly on goal-directed navigation, in contrast to purely path integration tasks. However, both of these rely on external spatial coordinates to represent current position. Our model instead constructs an internal coordinate system by integrating its own velocity output, enabling endogenous spatial tracking.

Naturalistic task The model was evaluated on a biologically inspired navigation benchmark involving exploration and goal-seeking behavior in closed environments. Performance was measured as the total number of rewards collected over multiple trials.

Optimization of the model hyper-parameters was carried out using the evolutionary Covariance-Matrix Adaptation strategy (CMA-ES) [62] with a population size of 90. The choice of the hyper-parameters to evolved took into account the type of dynamcis for which it was difficult to select pre-defined values; the number of evolved variables was 10. The use of such method derived from the impractility of other optimization algorithms, also given the non-differentiability of several dynamic. Viable alternatives were based on reinforcement learning, Bayesian and grid search, but evolution was more appealing for exploration and visualization of the parameter space exploration. For evaluating the individuals it was used a multi-objective fitness function: maximization of the reward count, and minimization of the collision count from the time the reward position was discovered.

3 Results

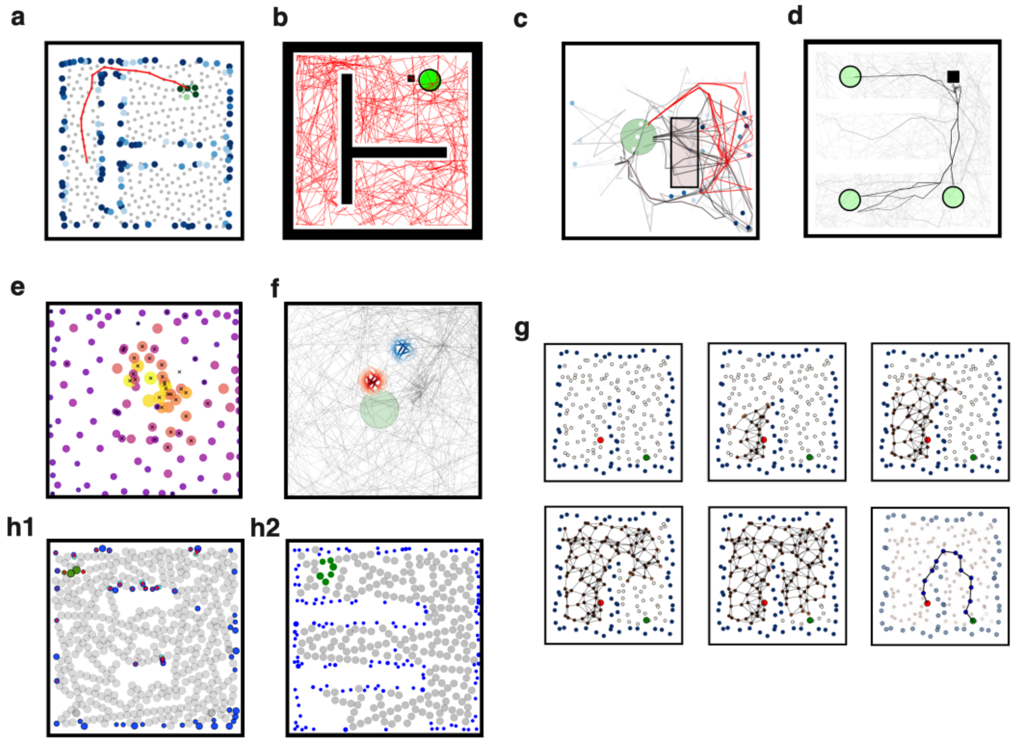

Figure 2: **Cognitive maps and performance results** - **a**: *a cognitive map over a space, together with the plan (red line) to reach a target location from a starting position.*. **b**: *the same enviroment but with the reward (green circle), trajectory (red line), agent position (black square).* - **c**: *plot of trajectories before (black) and after (red) the insertion of a wall (rectangle) between the starting and goal positions, the wall can also be spotted from the boundary cells in blue* - **d**: *trajectories for multiple trials with the agent starting at the same position (black square) but with the reward location (green circles) periodically moving* - **e**: *place cells centers with circle size and color proportional to their node degree; cells with a black cross changed their place field location over time* - **f**: *place fields of the same cell before and after several relocation of its center following reward events* - **g**: *visualization of part of the path-finding algorithm, propagation of an activity wave through the place cells network from top-left to bottom-center, and the calculated path visualization in the bottom-right.* - **h1**-**h2**: *place cells centers with size inversely proportional to their gain value; in blue boudary cells with the highest average gain, in green reward cells with second smallest gain, the others in grey.*

Performance in wayfinding Our primary aim was to evaluate the formation of the cognitive map through neuromodulation in terms of the performance of the goal navigation in different environments. The best model resulting from evolution reached solid navigation and adaptation skills. The agent was able to visit a significant portion of the environment during exploration and use neuromodulation to produce useful spatial representations.

The left panel of Fig. 2**a**-**b** displays place cells associated with collisions and reward events, signaling boundaries (in blue) and reward (in green) locations. The overlap of these two representations and the non-modulated place cells (in grey) is what we refer to as a cognitive map, since these are the main sources of spatial and contextual information used during planned navigation, whose path is depicted as a gray line. The right panel instead portrays the actual environment with walls (black), reward location (green), and multiple trajectories (red). During exploration, the main areas were visited until the reward position was located and the goal-directed navigation dominated, as highlighted by the density of the path lines. Considering the position of the walls and corners, the layout of this environment does not always make the target locations visible, as it is a non-convex area and therefore can be classified as wayfinding [63]. The challenge of not being able to use straight lines is overcome by the graph approach using local data and the consideration of boundary place cells, allowing the agent to plan accordingly. In addition, considering the Gaussian receptive fields and the approximately homogenous distribution of place field centers supported by lateral inhibition, the calculated path accounting for node-length also implicitly minimizes effective path-length, although not necessarily exactly. Figure 2**g** visualizes part of the path-finding process.

In general, this result confirms the ability of the model to focus on navigation and obstacle avoidance. However, it is worth nothing that not all simulations resulted in a reward being found in the first place, due to the randomness of the exploratory process; this effect was more pronounced in environment with more walls and narrow passages.

Detour task The planning ability and the plastic nature of the cognitive map should provide resilience against unexpected changes in the environment layout. In order to verify this we implemented a detour experiment. Initially, the agent was familiarized with a square environment with the reward in the middle and starting always from the same position. Then, a wall was placed in between the starting position and the reward, therefore forrcing new trajectory for reaching it. As expected, the agent was able to form a representation of the new obstacle and calculating new paths around it, succeeding the task. In Fig. 2**c** they are shown the trajectories before and after the wall placement, and it is manifesed the ability of detour in the new layout.

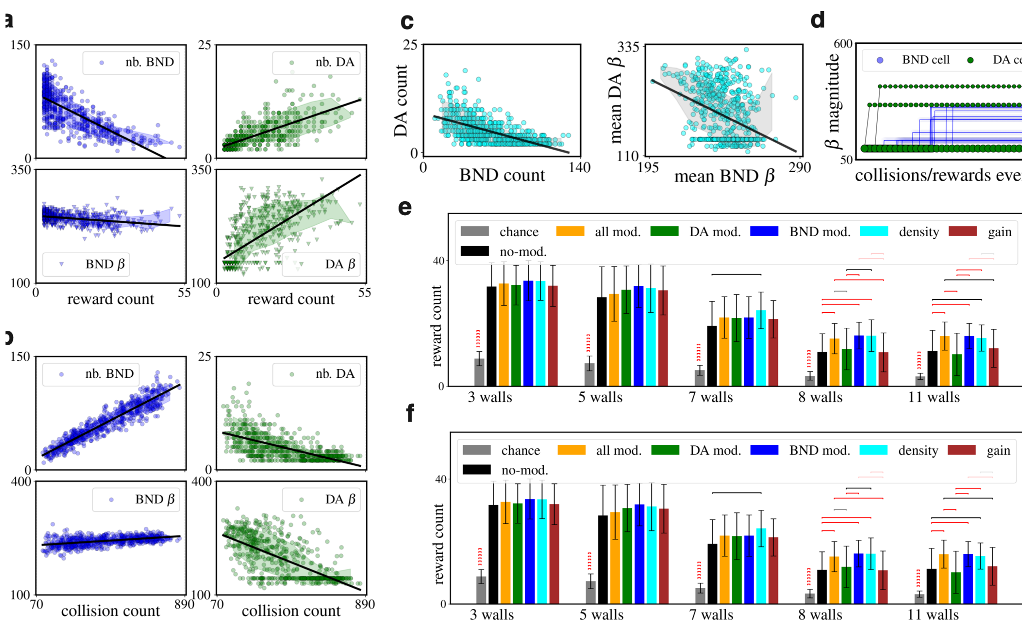

Figure 3: **Cognitive maps and performance results** - **a**: *effect of reward count on reward and boundary modulated cells (green and blue respectively), both in total count (top row) and average gain modulation magnitude (bottom row); simulation of 160 independent runs.* - **b**: *similar plot but with respect to collision count.* - **c**: *relation between count of reward and boundary modulated cells, and between gain modulation magnitude.* - **d**: *gain magnitude of boundary and reward cells over sequences of collision and reward events respectively* - **e**: *performance comparison for the same base model but with different levels of ablation: completely without modulatiory actions, all modulation enabled, only DA-related modulation, only BND-related modulation, only place cells density modulation, only activation gain modulation. Pair-wise t-test over 128 iterations and Bonferroni correction*

Adaptive goal representation through sensory error Then, we tested the adaptability to environmental changes. In this scenario, the reward object was moved after being fetched a fixed number of times. Here, the difficulty was to unlearning previous locations and discovering new ones, in a protocol similar to [39]. In Fig. 2**d** is reported the set of trajectories over many trials with the reward displaced in three possible locations. The agent was capable of planning behavior, as earlier, but also exploring and finding the new rewards, as shown by the density of lines. Whenever a goal path resulted in a failed prediction, the DA-based sensory error weakened the association between the place cells and the reward signal, leading to an extinction of its representation at that location.

This result validates the resilience of the model to changing sensory expectations, in this case the reward position.

Modulation of place field size The construction of model is such that the experience of environmental events can impact the neuronal properties of the generated place cells. In particular, collision and reward events have the effect of affecting the neural activation gain β of BND and DA-modulated cells through an hyperparameter . The hyperparameter values $\mathrm{c}_{\mathrm{a}}^{\mathrm{BND}}$, $\mathrm{c}_{\mathrm{a}}^{\mathrm{DA}}$ that yielded the best results were both larger that 1., 4.6, 4.4 respectively, meaning a shrinking of field size. In Fig. 2**h2**-**h2** are showed the cognitive maps with relative place field sizes for two enviroments, showcasing the differences between boundary, reward, and non-modulated cells. In addition, Fig. 2**h1** indicates the place cells that underwent re-location of their fields with a red lines representing the displacement vector. The distributon of these re-location vector is biased towards the rewarding area, a similar observation to Fig. 2**e** in which the cells involved by modulation of field positions are markered with a black cross. In Fig. 3**d** is showed the evolution of the gain magnitude for a sample of boundary cells and reward cells over their corresponding modulatory events. Notable is the possible decrease in value, case that occurs when a cell is active but the modulation is absent and the gain thus is pushed down towards the baseline b$\bar{\text{e}}$ta. This feature can be considered another adaptation property.

Effect of modulation on performance Lastly, we investigated the effect of modulating the density of place cells and the size of the field. The goal position was fixed, but the agent was randomly relocated after fetching; performance was defined as the total number of reward counts within a time window.

Our working hypothesis is that these experience-driven neuronal changes would improve the quality of the cognitive map and be reflected in navigational abilities. The assessment of this claim was conducted by comparing variants of the model, obtained by progressively ablating the reward (DA) and collision (BND) modulatory actions, as well as the density and gain mechanisms separately. All models were run in five different environments differing by number of internal walls, from 2 to 11, for a total of 64 independent simulation repetitions done for each single case.

The statistical results are shown in Fig. 3**e**. All models performed above chance, but the main finding is the confirmation of the importance of neuronal modulation, as revealed by the statistical difference of most modulation-powered models with respect to the one with all modulation disabled.

An additional finding was that density modulation is significantly the primary action behind the behavioral improvements and alone does not perform worse than the model with all actions together.

Figure 2**d** showcases the distribution of place cells with the circle size and color representing the node degree, which aligns discretely with the position of reward and the density. Furthermore, in Fig. 2**e** the place field of one cell is shown, before and after several reward occurrences and consequent center relocation.

Taken together, these findings support the hypothesis of practical utility of direct modulation of place-field structure for active navigation, even in these simple settings.

Convergence of evolved hyperparameters The use of an evolutionary algorithm for selecting the model dynamics had the convenience of providing a distribution of hyper-parameters over a population. In Figure 4, each dot represent the the value for one individual plotted with respect to its average reward count. A marked convergence is visible for most variables, with some displaying a quasi-bimodal distribution. The hyper-parameters directly involved in the neural activation, such as the parameters of the activation function β, α and the trace time constant τ^{PC}, display an elevated degree of clustering. Further, their specific values are such that the model is brought to develop a dense place cell network with, due to the weak lateral inhibition given the high threshold $\theta_{\mathrm{rep}}^{\mathrm{PC}}$, and low overlapping of fields, due to the steep gain β.

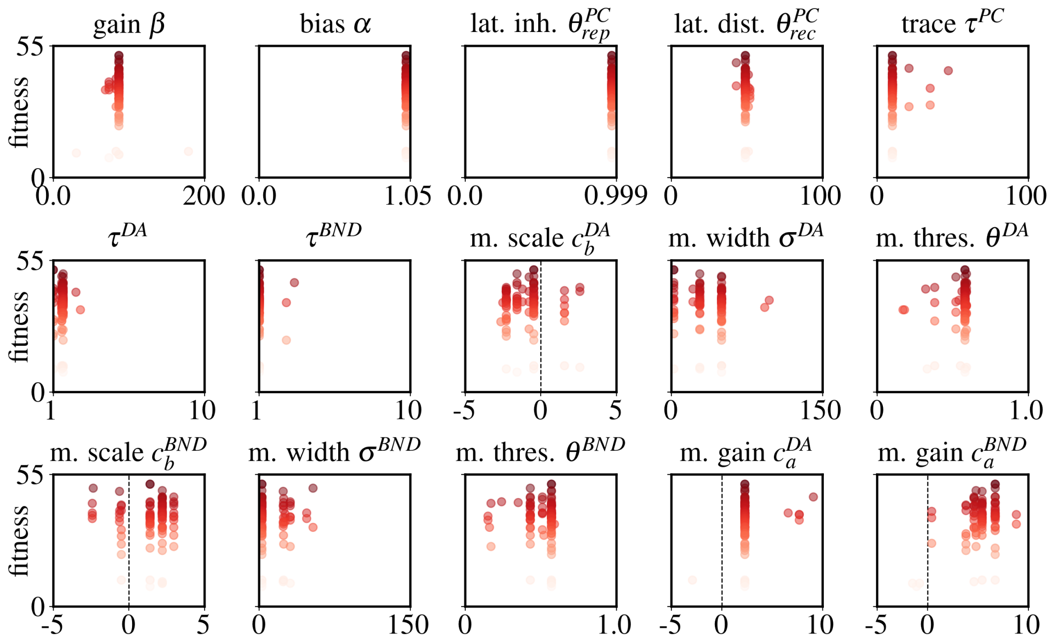

Figure 4: **Distribution of evolved hyper-parameters** - *Results relative to the last generation, from a run with population size of 90 individuals. The hyper-parameters are: neural activation gain β, neural acitvation bias α, lateral inhibition threshold θ^{PC}_{inh}, lateral distance threshold θ^{PC}_{rec}, activity trace time constant τ^{PC}, reward modulation scale c_b^{DA}, reward modulation spread c_b^{DA}, reward modulation threshold θ^{DA}, boundary modulation scale c_b^{BND}, boundary modulation spread c_b^{BND}, boundary modulation threshold θ^{BND}, reward gain modulation c_a^{DA}, and boundary gain modulation c_a^{BND}.*

As for modulation, there is a strong tendency of increasing the magnitude of the gain for both reward c_a^{DA} and collision c_a^{BND} events, with the effect of further reducing the size the active place fields. This result is in the direction of experimental observation showing shrinking of place fields near objects and walls [64, 65].

Instead, concerning the modulation of the place cells centers, there is more diversity in the adopted action. In fact, it can be both of pulling towards and pushing away from the current agent position, with the latter being more frequent in regard of collision modulation. These findings align, at least partially, with previous work about place fields remapping [66] and reward-induced changes in the preferred tuning location of hippocampal place cells [67, 68, 69, 16].

In summary, over the course of generations evolution converged to specific neural dynamics that are reminiscent of patterns observed in biological neuronal systems.

4 Discussion

Exploration and planning in the known and past environment are essential behaviors of animals, directly affecting their success in world understanding and goal reaching.

An important element behind these abilities is the formation of a map of their surroundings as they make new experiences, known as a cognitive map. Numerous speculations have been made about the shape and neural foundations of such an object, varying in the types of modeling assumptions and experimental support.

The contribution of the present work was to propose a rate network model, inspired by the CA1 hippocampal region [10]. We used grid cells together with synaptic plasticity as a mechanism to develop information-rich representations based on place cells updated through experience, grouped with common perspectives on cognitive maps [70]. In the spirit of minimizing the geometric assumptions in the neural space, we treated the generated place network as a topological graph, with sensory information added locally through the action of neuromodulators. This idea aligned with the concept of a *labeled graph* [71, 26], however, it is also true that no metric violations were possible in these settings.

The tasks we applied the agent to consisted of an exploratory and exploitatory phase, in which it was tasked to plan and reach reward positions. For simplicity, the first stage relied on a random walk process, as it was outside the scope of this work. This choice had the side effect that the reward was not always discovered, leading to the formation of incomplete maps, and thus impairing performance. However, this issue was limited in frequency.

The simulation results validated the model, showing the expected emergence of cognitive maps and their encoding of information collected during the experience. The online nature of the formation of the locations on the map aligns with the idea of using only idiothetic velocity input, as in path integration [24, 72, 73]. Previous work followed a similar direction using recurrent networks, but required extensive gradient-based training [60, 38, 51]. Another important difference is that our resulting neural network was composed solely of place cells, although neuromodulated, and no other types of neuron were present. This distinction is justified by the partially different task structure, which did not involve supervised learning and did not receive visual information as in [36]. Furthermore, our model relied on predefined grid cells layers, which constituted a strong and sufficient inductive bias, and did not have to be learned from scratch.

An additional relevant aspect is also the consideration of the place cell layer as an explicit graph data structure, on which the path-planning and decision-making algorithm was applied. The adoption of this level of description lead to robustness and flexibility, enabling effective navigation in all tested environments, which varying in layout complexity. Nevertheless, this approach did act as another clear inductive bias, which lifted the need to learn an approximation of it through network dynamics and even more differently tuned neurons.

Adaptability was tested by occasionally moving the reward position, leading to the generation of an internal prediction error that was used to update its representation on the map. The agent was proved capable of unlearning previous associations, returning to exploration, and memorizing new reward locations. This behavioral protocol is similar to previous work [61], in which dopaminergic and cholinergic activity was utilized within a Hebbian plasticity rule to strengthen or weaken reward-associated spatial representations. However, alternatively to exploiting neuromodulators with opposite valence, we followed a preditive coding framework, a direction linked to hippocampal representations [34, 74] and explored various computational approaches [75, 76, 77]. This choice departed from our focus on using operations on the cognitive map itself by simulating future sensory experiences and learning from feedback. In fact, neuromodulation has been long associated with this functionality [40], especially dopamine [14, 43, 46, 18].

Lastly, the hypothesis of the relevance of the active modulation of the neuronal properties of place cells was corroborated by simulating ablation experiments. These tests reported a significant impact of altering the place cells density on the total count of collected rewards. In general, these results are consistent with the experimental observations of alteration of place cells following reward events [16, 78], in particular in terms of increased clustering of cells [79, 80], reminiscent of changes in firing rate after contextual changes [66, 67].

Concerning the modulation of place fields, there is significant experimental evidence of their alternation during reward events [68, 81, 69], some reporting shrinkage near reward objects [64], and boundaries [65]. The coupling with higher local density could be explained by better optimization of the cell distribution for goal representation and planning [82]. However, in our settings, the fields become enlarged, especially in the direction of the target, although the performance improvements were not tested significantly. A possible explanation can be the simplicity of our reward, which was solely defined as an area of space. The lack of rich non-spatial features thus did not require the place cells to code for smaller spatial variation. Therefore, enlargement might have improved the stability of the representation, marking the nodes associated with rewards more solidly, given the stochasticity of its delivery. Further, the graph-path algorithm utilized the strength of the DA-modulated connections for determining the goal representation; stronger fields inherently developed stronger weights, making planning more reliable. Although these findings are limited within the limits of our simulation protocol, there have been experimental observations of elongation of place fields along trajectories over meaningful experiences [83, 84].

In conclusion, this work showed a possible architecture for coupling emergent spatial representations with neuromodulated plasticity to achieve an experience-driven cognitive map. The reliance on a few spatial and algorithmic inductive biases, grid cells, and a planning algorithm supports the idea of a label graph for goal navigation. Future work can investigate the application to other spatial domains, such as motor control and three-dimensional navigation. In addition, a richer input feature can be added, such as visual information [85], as well as new neuromodulators that encode different sensory dimensions or internally generated signals.

Acknowledgements & Statements
The authors declare no competing interests.
The code is publicly available and can be found at `https://github.com/iKiru-hub/PCNN`.
This research was funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement Nº 945371 and the University of Oslo.
The research presented in this paper has benefited from the Experimental Infrastructure for Exploration of Exascale Computing (eX3), which is financially supported by the Research Council of Norway under contract 270053.

References

[1] Reginald Golledge, Dan Jacobson, Rob Kitchin, and Mark Blades. Cognitive Maps, Spatial Abilities, and Human Wayfinding. *GEOGRAPHICAL REVIEW OF JAPAN SERIES B*, 73:93–104, December 2000.

[2] Russell A. Epstein and Lindsay K. Vass. Neural systems for landmark-based wayfinding in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1635):20120533, February 2014.

[3] Francesca Sargolini, Marianne Fyhn, Torkel Hafting, Bruce L. McNaughton, Menno P. Witter, May-Britt Moser, and Edvard I. Moser. Conjunctive Representation of Position, Direction, and Velocity in Entorhinal Cortex. *Science*, 312(5774):758–762, May 2006.

[4] Emilio Kropff, James E. Carmichael, May-Britt Moser, and Edvard I. Moser. Speed cells in the medial entorhinal cortex. *Nature*, 523(7561):419–424, July 2015.

[5] Trygve Solstad, Edvard I. Moser, and Gaute T. Einevoll. From grid cells to place cells: A mathematical model. *Hippocampus*, 16(12):1026–1031, 2006.

[6] Daniel Bush, Caswell Barry, and Neil Burgess. What do grid cells contribute to place cell firing? *Trends in Neurosciences*, 37(3):136–145, March 2014.

[7] Torsten Neher, Amir Hossein Azizi, and Sen Cheng. From grid cells to place cells with realistic field sizes. *PLOS ONE*, 12(7):e0181618, July 2017.

[8] Tianyi Li, Angelo Arleo, and Denis Sheynikhovich. *Modeling Place Cells and Grid Cells in Multi-Compartment Environments: Hippocampal-Entorhinal Loop as a Multisensory Integration Circuit*. April 2019.

[9] Olesia M. Bilash, Spyridon Chavlis, Cara D. Johnson, Panayiota Poirazi, and Jayeeta Basu. Lateral entorhinal cortex inputs modulate hippocampal dendritic excitability by recruiting a local disinhibitory microcircuit. *Cell Reports*, 42(1):111962, January 2023.

[10] Flavio Donato, Anja Xu Schwartzlose, and Renan Augusto Viana Mendes. How Do You Build a Cognitive Map? The Development of Circuits and Computations for the Representation of Space in the Brain. *Annual Review of Neuroscience*, 46(Volume 46, 2023):281–299, July 2023.

[11] John E. Lisman and Anthony A. Grace. The Hippocampal-VTA Loop: Controlling the Entry of Information into Long-Term Memory. *Neuron*, 46(5):703–713, June 2005.

[12] Adrian J. Duszkiewicz, Colin G. McNamara, Tomonori Takeuchi, and Lisa Genzel. Novelty and Dopaminergic Modulation of Memory Persistence: A Tale of Two Systems. *Trends in Neurosciences*, 42(2):102–114, February 2019.

[13] Wolfram Schultz, Peter Dayan, and P. Read Montague. A Neural Substrate of Prediction and Reward. *Science*, 275(5306):1593–1599, March 1997.

[14] Kimberly A. Kempadoo, Eugene V. Mosharov, Se Joon Choi, David Sulzer, and Eric R. Kandel. Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proceedings of the National Academy of Sciences*, 113(51):14835–14840, December 2016.

[15] Aude Retailleau and Thomas Boraud. The Michelin red guide of the brain: Role of dopamine in goal-oriented navigation. *Frontiers in Systems Neuroscience*, 8, March 2014.

[16] Katie C. Bittner, Aaron D. Milstein, Christine Grienberger, Sandro Romani, and Jeffrey C. Magee. Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science*, 357(6355):1033–1036, September 2017.

[17] Alexandra Mansell Kaufman, Tristan Geiller, and Attila Losonczy. A Role for the Locus Coeruleus in Hippocampal CA1 Place Cell Reorganization during Spatial Reward Learning. *Neuron*, 105(6):1018–1026.e4, March 2020.

[18] Denis Sheynikhovich, Satoru Otani, Jing Bai, and Angelo Arleo. Long-term memory, synaptic plasticity and dopamine in rodent medial prefrontal cortex: Role in executive functions. *Frontiers in Behavioral Neuroscience*, 16, January 2023.

[19] Kei M. Igarashi, Hiroshi T. Ito, Edvard I. Moser, and May-Britt Moser. Functional diversity along the transverse axis of hippocampal area CA1. *FEBS Letters*, 588(15):2470–2476, August 2014.

[20] Hiroshi T. Ito and Erin M. Schuman. Functional division of hippocampal area CA1 via modulatory gating of entorhinal cortical inputs. *Hippocampus*, 22(2):372–387, 2012.

[21] Michael Peer, Iva K. Brunec, Nora S. Newcombe, and Russell A. Epstein. Structuring Knowledge with Cognitive Maps and Cognitive Graphs. *Trends in cognitive sciences*, 25(1):37–54, January 2021.

[22] Elizabeth R. Chrastil and William H. Warren. From Cognitive Maps to Cognitive Graphs. *PLoS ONE*, 9(11):e112544, November 2014.

[23] Steffen Werner, Bernd Krieg-Brückner, and Theo Herrmann. Modelling Navigational Knowledge by Route Graphs. In Christian Freksa, Christopher Habel, Wilfried Brauer, and Karl F. Wender, editors, *Spatial Cognition II: Integrating Abstract Theories, Empirical Studies, Formal Methods, and Practical Applications*, pages 295–316. Springer, Berlin, Heidelberg, 2000.

[24] C. R. Gallistel and Audrey E. Cramer. Computations on Metric Maps in Mammals: Getting Oriented and Choosing a Multi-Destination Route. *Journal of Experimental Biology*, 199(1):211–217, January 1996.

[25] Michael Peer, Catherine Nadar, and Russell A. Epstein. The format of the cognitive map depends on the structure of the environment. *Journal of Experimental Psychology: General*, 153(1):224–240, January 2024.

[26] William H. Warren. Non-Euclidean navigation. *Journal of Experimental Biology*, 222(Suppl_1):jeb187971, February 2019.

[27] Mark Wagner. Comparing the psychophysical and geometric characteristics of spatial perception and cognitive maps. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 15(1):6–21, 2008.

[28] Rainer Rothkegel, Karl F. Wender, and Sabine Schumacher. Judging Spatial Relations from Memory. In Christian Freksa, Christopher Habel, and Karl F. Wender, editors, *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, pages 79–105. Springer, Berlin, Heidelberg, 1998.

[29] Tobias Meilinger. The Network of Reference Frames Theory: A Synthesis of Graphs and Cognitive Maps. In Christian Freksa, Nora S. Newcombe, Peter Gärdenfors, and Stefan Wölfl, editors, *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, pages 344–360, Berlin, Heidelberg, 2008. Springer.

[30] Jane X. Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z. Leibo, Remi Munos, Charles Blundell, Dharshan Kumaran, and Matt Botvinick. Learning to reinforcement learn, January 2017.

[31] Victor R. Schinazi, Daniele Nardi, Nora S. Newcombe, Thomas F. Shipley, and Russell A. Epstein. Hippocampal size predicts rapid learning of a cognitive map in humans. *Hippocampus*, 23(6):515–528, 2013.

[32] Bruno Poucet. Spatial cognitive maps in animals: New hypotheses on their structure and neural mechanisms. *Psychological Review*, 100(2):163–182, 1993.

[33] Paul Stoewer, Achim Schilling, Andreas Maier, and Patrick Krauss. Neural network based formation of cognitive maps of semantic spaces and the putative emergence of abstract concepts. *Scientific Reports*, 13(1):3644, March 2023.

[34] William de Cothi, Nils Nyberg, Eva-Maria Griesbauer, Carole Ghanamé, Fiona Zisch, Julie M. Lefort, Lydia Fletcher, Coco Newton, Sophie Renaudineau, Daniel Bendor, Roddy Grieves, Éléonore Duvelle, Caswell Barry, and Hugo J. Spiers. Predictive maps in rats and humans for spatial navigation. *Current Biology*, 32(17):3676–3689.e5, September 2022.

[35] James C. R. Whittington, Timothy H. Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy E. J. Behrens. The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell*, 183(5):1249–1263.e23, November 2020.

[36] Andrea Banino, Caswell Barry, Benigno Uria, Charles Blundell, Timothy Lillicrap, Piotr Mirowski, Alexander Pritzel, Martin J. Chadwick, Thomas Degris, Joseph Modayil, Greg Wayne, Hubert Soyer, Fabio Viola, Brian Zhang, Ross Goroshin, Neil Rabinowitz, Razvan Pascanu, Charlie Beattie, Stig Petersen, Amir Sadik, Stephen Gaffney, Helen King, Koray Kavukcuoglu, Demis Hassabis, Raia Hadsell, and Dharshan Kumaran. Vector-based navigation using grid-like representations in artificial agents. *Nature*, 557(7705):429–433, May 2018.

[37] Ben Sorscher, Gabriel C. Mel, Samuel A. Ocko, Lisa M. Giocomo, and Surya Ganguli. A unified theory for the computational and mechanistic origins of grid cells. *Neuron*, 111(1):121–137.e13, January 2023.

[38] Christopher J. Cueva and Xue-Xin Wei. Emergence of grid-like representations by training recurrent neural networks to perform spatial localization, March 2018.

[39] Zuzanna Brzosko, Susanna B. Mierau, and Ole Paulsen. Neuromodulation of Spike-Timing-Dependent Plasticity: Past, Present, and Future. *Neuron*, 103(4):563–581, August 2019.

[40] Marielena Sosa, Mark H. Plitt, and Lisa M. Giocomo. Hippocampal sequences span experience relative to rewards. *bioRxiv*, page 2023.12.27.573490, February 2024.

[41] Abdullahi Ali, Nasir Ahmad, Elgar de Groot, Marcel A. J. van Gerven, and Tim C. Kietzmann. Predictive coding is a consequence of energy efficiency in recurrent neural networks, November 2021.

[42] Jacopo Bono, Sara Zannone, Victor Pedrosa, and Claudia Clopath. Learning predictive cognitive maps with spiking neurons during behavior and replays. *eLife*, 12:e80671, March 2023.

[43] Wolfram Schultz. Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*, 18(1):23–32, March 2016.

[44] Jeffrey B. Inglis, Vivian V. Valentin, and F. Gregory Ashby. Modulation of Dopamine for Adaptive Learning: A Neurocomputational Model. *Computational brain & behavior*, 4(1):34–52, March 2021.

[45] Philippe N. Tobler, Christopher D. Fiorillo, and Wolfram Schultz. Adaptive Coding of Reward Value by Dopamine Neurons. *Science*, 307(5715):1642–1645, March 2005.

[46] Roshan Cools. Chemistry of the Adaptive Mind: Lessons from Dopamine. *Neuron*, 104(1):113–131, October 2019.

[47] Aaron D Milstein, Yiding Li, Katie C Bittner, Christine Grienberger, Ivan Soltesz, Jeffrey C Magee, and Sandro Romani. Bidirectional synaptic plasticity rapidly modifies hippocampal representations. *eLife*, 10:e73046, December 2021.

[48] André A. Fenton. Remapping revisited: How the hippocampus represents different spaces. *Nature Reviews Neuroscience*, 25(6):428–448, June 2024.

[49] Luxin Zhou and Yong Gu. Cortical Mechanisms of Multisensory Linear Self-motion Perception. *Neuroscience Bulletin*, 39(1):125–137, July 2022.

[50] Steven J. Jerjian, Devin R. Harsch, and Christopher R. Fetsch. Self-motion perception and sequential decision-making: Where are we heading? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1886):20220333, August 2023.

[51] Ian Q. Whishaw and Brian L. Brooks. Calibrating space: Exploration is important for allothetic and idiothetic navigation. *Hippocampus*, 9(6):659–667, 1999.

[52] Yuri Dabaghian. Grid Cells, Border Cells and Discrete Complex Analysis.

[53] Vemund Sigmundson Schøyen, Kosio Beshkov, Markus Borud Pettersen, Erik Hermansen, Konstantin Holzhausen, Anders Malthe-Sørenssen, Marianne Fyhn, and Mikkel Elle Lepperød. Hexagons all the way down: Grid cells as a conformal isometric map of space. *PLOS Computational Biology*, 21(2):e1012804, February 2025.

[54] Richard S Sutton and Andrew G Barto. The Reinforcement Learning Problem.

[55] P. R. Montague, P. Dayan, and T. J. Sejnowski. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16(5):1936–1947, March 1996.

[56] Seetha Krishnan, Chad Heer, Chery Cherian, and Mark E. J. Sheffield. Reward expectation extinction restructures and degrades CA1 spatial maps through loss of a dopaminergic reward proximity signal. *Nature Communications*, 13(1):6662, November 2022.

[57] Colin G. McNamara, Álvaro Tejero-Cantero, Stéphanie Trouche, Natalia Campo-Urriza, and David Dupret. Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. *Nature Neuroscience*, 17(12):1658–1660, December 2014.

[58] Frédéric Michon, Esther Krul, Jyh-Jang Sun, and Fabian Kloosterman. Single-trial dynamics of hippocampal spatial representations are modulated by reward value. *Current biology: CB*, 31(20):4423–4435.e5, October 2021.

[59] Philip Shamash and Tiago Branco. Mice identify subgoal locations through an action-driven mapping process, December 2021.

[60] Ben Sorscher, Gabriel Mel, Surya Ganguli, and Samuel Ocko. A unified theory for the origin of grid cells through the lens of pattern formation. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[61] Zuzanna Brzosko, Sara Zannone, Wolfram Schultz, Claudia Clopath, and Ole Paulsen. Sequential neuromodulation of Hebbian plasticity offers mechanism for effective reward-based navigation. *eLife*, 6:e27756, 2017.

[62] Christian Igel, Nikolaus Hansen, and Stefan Roth. Covariance Matrix Adaptation for Multi-objective Optimization. *Evolutionary Computation*, 15(1):1–28, March 2007.

[63] Tobias Meilinger, Marianne Strickrodt, and Heinrich H. Bülthoff. Qualitative differences in memory for vista and environmental spaces are caused by opaque borders, not movement or successive presentation. *Cognition*, 155:77–95, October 2016.

[64] S.N. Burke, A.P. Maurer, S. Nematollahi, A.R. Uprety, J.L. Wallace, and C.A. Barnes. The Influence of Objects on Place Field Expression and Size in Distal Hippocampal CA1. *Hippocampus*, 21(7):783–801, July 2011.

[65] Sander Tanni, William De Cothi, and Caswell Barry. State transitions in the statistically stable place cell population correspond to rate of perceptual change. *Current Biology*, 32(16):3505–3514.e7, August 2022.

[66] Michael I. Anderson and Kathryn J. Jeffery. Heterogeneous Modulation of Place Cell Firing by Changes in Context. *Journal of Neuroscience*, 23(26):8827–8835, October 2003.

[67] Inah Lee, Amy L. Griffin, Eric A. Zilli, Howard Eichenbaum, and Michael E. Hasselmo. Gradual Translocation of Spatial Correlates of Neuronal Firing in the Hippocampus toward Prospective Reward Locations. *Neuron*, 51(5):639–650, September 2006.

[68] Marianne Fyhn, Sturla Molden, Stig Hollup, May-Britt Moser, and Edvard I. Moser. Hippocampal Neurons Responding to First-Time Dislocation of a Target Object. *Neuron*, 35(3):555–566, August 2002.

[69] David Dupret, Joseph O'Neill, Barty Pleydell-Bouverie, and Jozsef Csicsvari. The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature Neuroscience*, 13(8):995–1002, August 2010.

[70] Vincent Hok, Pierre-Pascal Lenck-Santini, Sébastien Roux, Etienne Save, Robert U. Muller, and Bruno Poucet. Goal-Related Activity in Hippocampal Place Cells. *Journal of Neuroscience*, 27(3):472–482, January 2007.

[71] Toru Ishikawa and Daniel R. Montello. Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, 52(2):93–129, March 2006.

[72] Sabine Gillner and Hanspeter A. Mallot. Navigation and Acquisition of Spatial Knowledge in a Virtual Maze. *Journal of Cognitive Neuroscience*, 10(4):445–463, July 1998.

[73] Bruce L. McNaughton, Francesco P. Battaglia, Ole Jensen, Edvard I. Moser, and May-Britt Moser. Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 7(8):663–678, August 2006.

[74] Fraser Aitken and Peter Kok. Hippocampal representations switch from errors to predictions during acquisition of predictive associations. *Nature Communications*, 13(1):3294, June 2022.

[75] Manu Srinath Halvagal and Friedemann Zenke. The combination of Hebbian and predictive plasticity learns invariant object representations in deep sensory networks. *Nature Neuroscience*, pages 1–10, October 2023.

[76] Alexander Ororbia. Spiking neural predictive coding for continually learning from data streams. *Neurocomputing*, 544:126292, August 2023.

[77] Kimberly L Stachenfeld, Matthew M Botvinick, and Samuel J Gershman. The hippocampus as a predictive map. *Nature Neuroscience*, 20(11):1643–1653, November 2017.

[78] Indrajith R. Nair, Guncha Bhasin, and Dipanjan Roy. Hippocampus Maintains a Coherent Map Under Reward Feature–Landmark Cue Conflict. *Frontiers in Neural Circuits*, 16, April 2022.

[79] Valerie L. Tryon, Marsha R. Penner, Shawn W. Heide, Hunter O. King, Joshua Larkin, and Sheri J. Y. Mizumori. Hippocampal neural activity reflects the economy of choices during goal-directed navigation. *Hippocampus*, 27(7):743–758, July 2017.

[80] Hannah S Wirtshafter and Matthew A Wilson. Differences in reward biased spatial representations in the lateral septum and hippocampus. *eLife*, 9:e55252, May 2020.

[81] P.-P. Lenck-Santini, B. Rivard, R.u. Muller, and B. Poucet. Study of CA1 place cell activity and exploratory behavior following spatial and nonspatial changes in the environment. *Hippocampus*, 15(3):356–369, 2005.

[82] Pablo Scleidorovich, Jean-Marc Fellous, and Alfredo Weitzenfeld. Adapting hippocampus multi-scale place field distributions in cluttered environments optimizes spatial navigation and learning. *Frontiers in Computational Neuroscience*, 16:1039822, December 2022.

[83] Mayank R. Mehta, Carol A. Barnes, and Bruce L. McNaughton. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences*, 94(16):8918–8921, August 1997.

[84] Jangho Lee, Jeonghee Jo, Byounghwa Lee, Jung-Hoon Lee, and Sungroh Yoon. Brain-inspired Predictive Coding Improves the Performance of Machine Challenging Tasks. *Frontiers in Computational Neuroscience*, 16:1062678, 2022.

[85] John H. Wen, Ben Sorscher, Emily A. Aery Jones, Surya Ganguli, and Lisa M. Giocomo. One-shot entorhinal maps enable flexible navigation in novel environments. *Nature*, 635(8040):943–950, November 2024.

5 Appendix

5.1 Grid cell module

It is defined a correspondence between the global environment in which the agent moves, a two-dimensional Euclidean space $\mathbf{R}^2$, and a bounded local space of a grid module, corresponding to a torus.

The global velocity $\mathbf{v} = \{x, y\}$ is then mapped to a local velocity, scaled by a speed scalar s_l^{gc} specific to the grid cell module l, which determines its periodicity in space.

The choice of a toroidal space is motivated by consolidated experimental evidence of the neural space of grid cells, which are organized in modules of different sizes spanning the animal's environment. However, the shape of their firing pattern is known to be hexagonal, which corresponds to the optimal tiling of a two-dimensional plane, giving rise to a neural space lying on a twisted torus. In this work, for simplicity, we consider a square tiling and thus a square torus, without much loss of generality except for the slight increase of grid cells required for a sufficiently cover.

A grid cell module l of size N^{gc} is identified by a set of positions defined over a square centered on the origin and size of 2, such that $\{(x_i, y_i) \mid i \in N^{gc} \;\wedge\; x_i, y_i \in (-1, 1)\}$. This local square space has boundary conditions for each dimension, such that, for instance, when $x_t + s_l^{gc} \cdot v_x > 2$ the position update is taken to the other side $x_{t+1} = x_t + s_l^{gc} \cdot v_x - 2$, where s_l^{gc} is the scale of the velocity in the local space of the module l with respect to the real global agent velocity $v = \{v_x, v_y\}$. When the module is initialized, the starting positions of its cells are uniformly distributed over the square forming a lattice. When the agent is reset in a new position at the beginning of new trial, a displacement vector is applied to the last cells positions such that the mapping between the module local space and the global environment is preserved.

The firing rate vector of each cell is determined with respect to a 2D Gaussian tuning curve centered at the origin at $(0, 0)$, and it is calculated as
$r_i = \exp\left(-\frac{x_i^2+y_i^2}{\sigma_l^{gc}}\right)$, where σ_l^{gc} is the width of the tuning curve for module l. An illustration of the receptive field over a 2D environment and a toroidal space is reported in Figure 5**a**-**b**.

The final population vector of the grid cell network GC is the concatenated and flattened firing rate vector of all modules $\mathbf{u}^{GC}$.

In our model, each grid cell had a tuning width of 0.04. They were defined as 8 modules of size 36, and the relative speed scales were $\{1., 0.8, 0.7, 0.5, 0.4, 0.3, 0.2, 0.1, 0.07\}$.

5.2 Place cells

Tuning formation The tuning of a new place cell is simply defined as the current GC population vector $\mathbf{u}_t^{GC}$, and its index is that of the first silent cell, which is added to the forward weight matrix $\mathbf{W}_i^{GC,PC} \leftarrow \mathbf{u}^{GC}$.

In order to avoid overlapping of place fields, lateral inhibition is implemented. More specifically, the tuning process is aborted in case the cosine similarity of the new pattern and the old ones is greater than a threshold $\theta^{\mathrm{PC}}_{\mathrm{inh}}$.

Each cell represents a position in the GC activity space, which can be considered a node within a graph of place cells (PC). Although it is totally possible to only use the N^{GC}-dimensional tuning patterns and be agnostic about the dimensionality of the space in which the agent lives, to simplify the calculations, we mapped each pattern to 2D positions in a vector space. Then, the PC recurrent connectivity matrix is calculated with a nearest neighbors algorithm, which instead of a fixed number K of neighbors uses a lateral distance threshold $\theta^{\mathrm{PC}}_{\mathrm{rec}}$.

Activity The current firing rate of the PC population is determined by the cosine similarity between the GC input and the forward weight matrix, then passed through a generalized sigmoid $\phi(\mathrm{z}) = \left[1 + \exp(-\beta(\mathrm{z} - \alpha))\right]^{-1}$. The parameter α represents the activation threshold, or horizontal offset, while β the gain, or steepness.

$$\mathrm{u}^{\mathrm{PC}}_{\mathrm{i}} = \phi\left(\cos\left(\mathbf{u}^{\mathrm{GC}}, \mathbf{W}^{\mathrm{GC,PC}}_{\mathrm{i}}\right)\right) \tag{4}$$

It is also defined as an activity trace, which has an upper value of 1 and decays exponentially:

$$\mathrm{m}_{\mathrm{i}} = -\mathrm{m}_{\mathrm{i}}/\tau^{\mathrm{PC}} + \mathrm{u}_{\mathrm{i}} \tag{5}$$

It is used as a proxy for a memory trace.

In the model, a PC population is defined by its average place field size, determining the granularity of the representation of the place. In plot 5**b** it is illustrated an example of place cells layer tuning obtained from a continuous trajectory over a square environment.

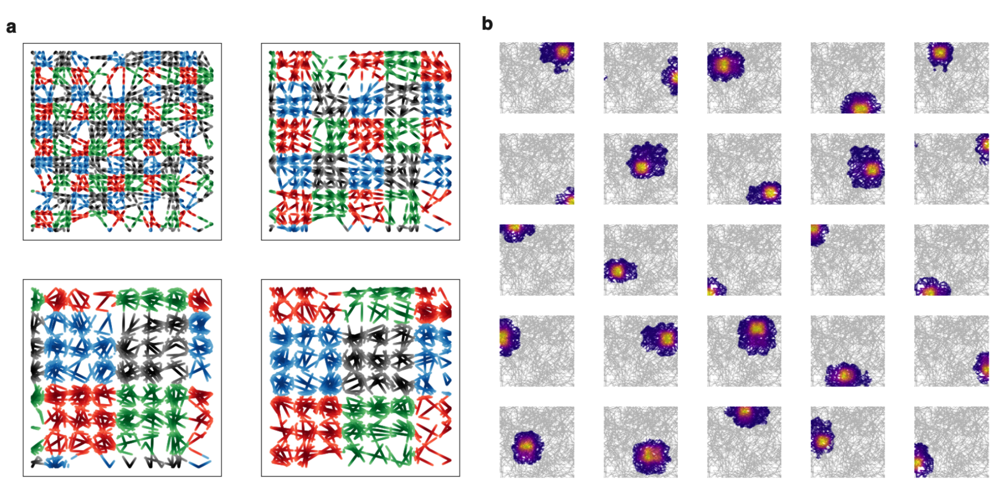

Figure 5: **Place fields obtained from grid cells activity - a**: *grid cell modules with different granularity represented over a continuous trajectory in an open space. For visualization purposes, each module is represented as composed of four sub-modules of 9 grid cells each, whose periodic tuning generates activity that repeats in space*. **b**: *place cells whose spatial tuning has been obtained from the concatenation of the grid cells population vector.*

5.3 Modulation

Neuromodulation is implemented as a reward-sensitive signal, represented as DA (mimicking the function of dopamine), and a collision-sensitive signal, represented as BND (for boundary). Its dynamics are defined in terms of a leaky variable v whose state is perturbed by an external input x, whose qualitative meaning differs for each neuromodulator k.

$$\begin{aligned} \mathrm{v}^{\mathrm{k}} &= -\mathrm{v}^{\mathrm{k}}/\tau^{\mathrm{k}} + \mathrm{x}^{\mathrm{k}} \\ \mathrm{v}^{\mathrm{k}} &= \max(\mathrm{v}^{\mathrm{k}}, 0) \end{aligned} \tag{6}$$

$\tau_{\mathrm{DA}} = 2$, $\tau^{\mathrm{BND}} = 1$

Learning rule The connection weights $\mathbf{W}^{\mathrm{k}}$ are updated according to a plasticity rule composed of an Hebbian term, involving the leaky variable, the place cells that are above a certain threshold θ^{k}, and the current connection weights value:

$$\Delta\mathbf{W}^{\mathrm{k}} = \eta^{\mathrm{k}}\mathbf{u}^{\mathrm{PC}}\left(\mathrm{v}^{\mathrm{k}} - \mathbf{W}^{\mathrm{k}}\right) \tag{7}$$

where the weight contribution of the Hebbian update, and η^{k} is the learning rate: $\eta^{\mathrm{DA}} = 0.9$, $\eta^{\mathrm{BND}} = 0.9$. Additionally, connections values are kept non-negative.

Active neuronal modulation Neuromodulation acts on the neuronal profile of the place cells by affecting the value of the activation gain and relocate the center of their tuning.

Gain modulation is implemented using the activity traces and a constant reference gain value $\bar{\beta}$:

$$\beta_{\mathrm{i}} = \mathrm{c}_{\mathrm{a}}^{\mathrm{k}}\mathrm{m}_{\mathrm{i}}\bar{\beta} + (1 - \mathrm{m}_{\mathrm{i}})\bar{\beta} \tag{8}$$

where $\mathrm{c}_{\mathrm{a}}^{\mathrm{k}}$ is a scaling gain parameter, and if it is 1 then no modulation takes place.

Concerning center relocation, it is applied to recently active neurons with non-zero trace m_{i}. For a place cell i with position $\mathbf{W}_{\mathrm{i}}^{\mathrm{GC,PC}}$ (in the grid cell space), it is calculated a displacement vector q_{i} with respect to the current position $\mathbf{u}^{\mathrm{GC}}$:

$$\mathrm{q}_{\mathrm{i}} = \mathrm{c}_{\mathrm{b}}^{\mathrm{k}}\mathrm{v}^{\mathrm{k}}\exp\left(-\frac{\|\mathbf{W}_{\mathrm{i}}^{\mathrm{GC,PC}} - \mathbf{u}^{\mathrm{GC}}\|}{\sigma^{\mathrm{k}}}\right) \tag{9}$$

where $\mathrm{c}_{\mathrm{b}}^{\mathrm{k}}$ is a scaling relocation parameter, while σ^{k} the width of the Gaussian distance. This displacemented is used to move in GC activity space and get the new GC population vector to use as tuning pattern.

For this modulatory action, they were considered solely cells whose activity trace was greated than modulator-specific threshold θ^{k}, so to prevent involving silent cells. Also in this case, it is ensured that the new place field center is at a minimum distance $\theta_{\mathrm{min}}^{\mathrm{PC}}$ from the others; here Euclidean distance is used.

5.4 Decision making

Behaviour selection logic
The possible behaviours are *exploration* and *exploitation*, and an action is defined as a 2D velocity vector. For exploration, an action can be generated either as random navigation, using a polar vector of fixed magnitude (the speed) and angle sampled from a uniform distribution, or as a step within a goal-directed navigation plan to reach a random destination, which corresponds to a randomly sampled existing place cells. In the goal-directed navigation the magnitude of the velocity vector is less or equal than a fixed speed value, depending on the distance from the next target position in the plan Instead, for the exploitation behaviour, the action is a step within a goal-directed navigation towards the reward location. The behaviour selection process depends on the experience of collision,

the presence of a plan, and the success in the navigation planning. A diagram of this logic is reported in Figure 6.

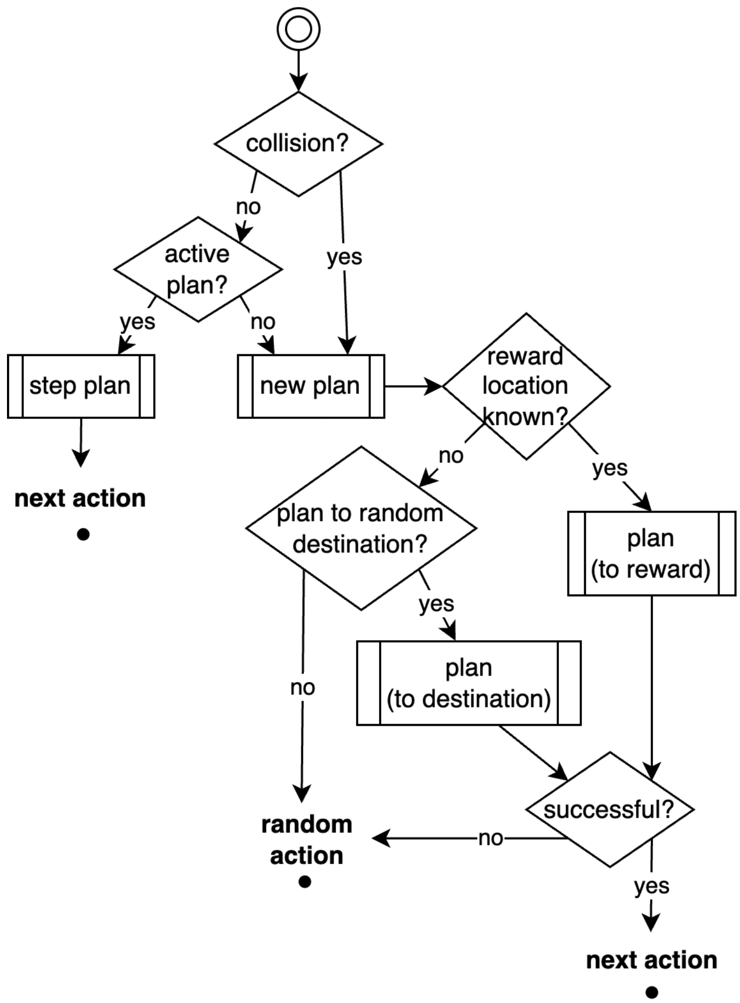

Figure 6: **Diagram of the behaviour selection process**

The positions of the agent and of the target location for planning are identified by the place cells population vector. In particular, the reward position (x_r, y_r is determined by the weighted average of the centers x_i, y_i of the place cells with respect to their DA-modulated connections weights. Further, only the top 5 place cells are considered.

$$
\begin{aligned}
x_r &= \sum_i^5 \mathbf{W}_i^{DA} x_i \\
y_r &= \sum_i^5 \mathbf{W}_i^{DA} y_i
\end{aligned}
\tag{10}
$$

Path-planning algorithm

Algorithm 1 ACTIVITY-BASED PATHFINDING

Require: Connectivity matrix $C \in \mathbb{R}^{n \times n}$, node weights $w \in \mathbb{R}^n$, start node s, end node e
Ensure: A list of nodes forming a (short) path from s to e, or empty if none found

Initialize activity vector $a \leftarrow \mathbf{0} \in \mathbb{R}^n$; set $a_s \leftarrow 1$
Initialize history list $H \leftarrow [\,]$ ▷ *— Forward propagation phase —*
for t = 1 to `MAX_PATH_DEPTH` **do**
    $a \leftarrow C \cdot a$
    $a \leftarrow a \circ w$ ▷ Element-wise multiplication with node weights
    $a \leftarrow \sigma(4(a - 0.6))$ ▷ Apply sigmoidal activation: $\sigma(x) = \frac{1}{1+e^{-x}}$
    $a_i \leftarrow 0$ if $a_i < 0.1$ (thresholding)
    Append a to H
    if $a_e > 0$ **then**
        break
    end if
end for
if maximum depth reached **then**
    return [ ] ▷ No path found
end if
if $|H| < 3$ **then**
    return $[s, e]$ ▷ Path is trivially short
end if ▷ *— Backward traceback phase —*
Initialize path index stack $G \leftarrow [[e]]$
for t = 1 to `MAX_PATH_DEPTH` **do**
    Let $m \leftarrow C_{G[t-1]}$ ▷ Get neighbors of current group
    if $m_s > 0$ **then**
        break
    end if
    $a \leftarrow H[-(t+1)] \circ m$
    Append $\{i : a_i > 0\}$ to G
end for ▷ *— Path reconstruction —*
$P \leftarrow [e]$
for $t = 1$ to $|G| - 1$ **do**
    Initialize $a \leftarrow \mathbf{0} \in \mathbb{R}^n$
    Set $a_i \leftarrow 1$ for all $i \in G[t]$
    $a \leftarrow a \circ C_{P[-1]}$
    if $\sum a = 0$ **then**
        return [ ] ▷ No valid neighbor
    else
        Choose $j \in \{i : a_i > 0\}$ uniformly at random
        Append j to P
    end if
end for
Append s to P
return reverse(P)

The planning of a new route is implemented as a path-finding algorithm based on the place cell graph, provided as connectivity matrix C. Its particularity is the use of a weighting $\widetilde{\mathbf{W}}$ of the nodes according to the neuromodulation map. A description is reported in algorithm 5.4.

5.5 Environments

The game in which test the model has been developed with the python library Pygame, used under license GNU LGPL version 2.1 and available at `https://github.com/pygame/pygame`. The environment layout consisting in a customizable arrangement of vertical and horizontal hard walls with variable length and fixed width. Below in Figure 7 some samples are shown.

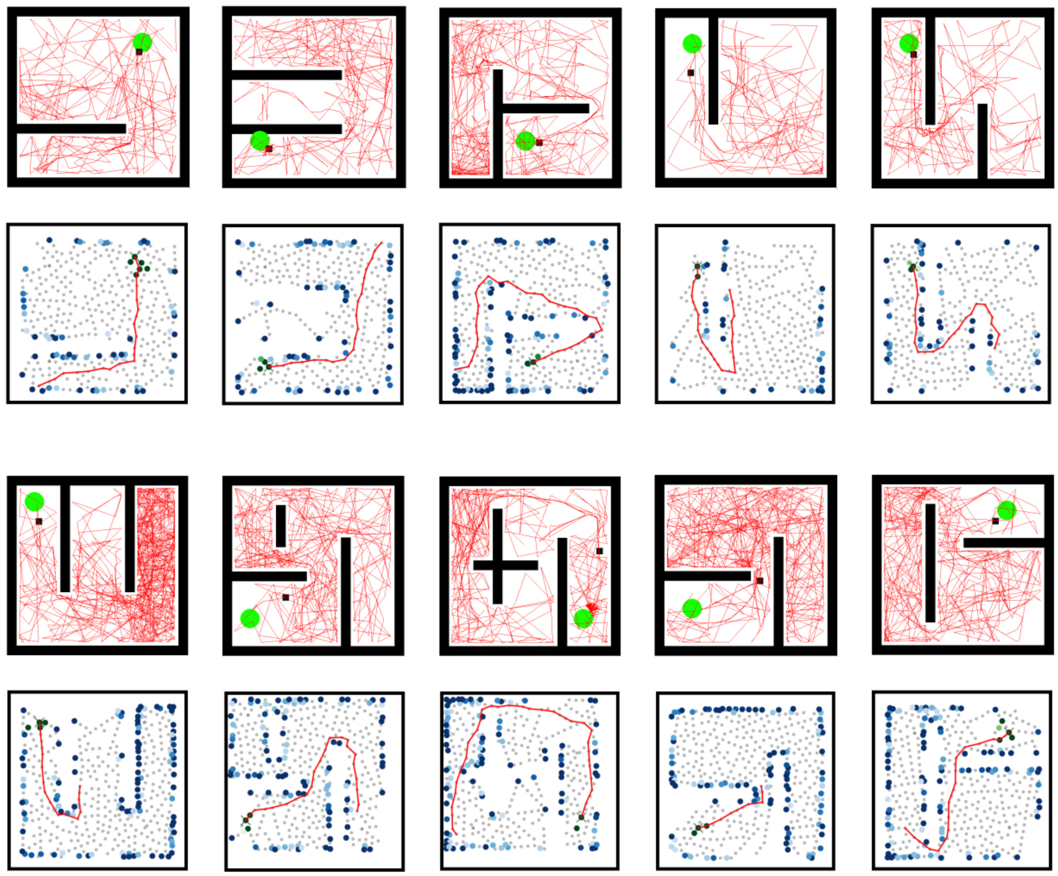

Figure 7: **Sample of generated environments**

The reward object is defined as a circle with size the 5% of the total environment area. When the agent position is within its boundary, it is provided a binary signal $R \sim \mathcal{B}(p_r)$ drawn from a Bernoulli with probability $p_r = 0.6$. The duration of the reward fetching is set to 2 time steps.

The agent object is defined as a square with size the 3.3% of the total environment area.

The testing protocol was inspired by the behavior of animals that venture into new territories in search of food. It was divided into two parts:

- **exploration phase**: the agent was placed in a random location within the environment for 10'000 time steps. In this phase the reward is not present. Further, in order to force greater exploration of the environment, every 3'000 steps it was teleported to another random location. This external intervention was meant to mitigate the randomness in the exploratory behavioural strategy of the agent.
- **reward phase**: a reward is insert in a random location, and it is available to be discovered. When it is encountered, the agent is teleported to a random location within the environement, and after a fixed about $\tilde{1}00$ time step it is enabled its reward-seeking behaviour, in the form of goal-directed navigation. The total duration of this phase is set to 20'000 time steps.

An episode is defined as a continuous trajectory during the reward phase, namely a set of time steps starting from when the agent is place in a position until either it finds the reward or the simulation ends.

Detour experiment The protocol is modified such that after a fixed number of episodes the layout of the environment is changed, *e.g.* a wall is inserted. This experiment is meant to test the ability to reach the reward location by using the same cognitive map, and possibly update it with the new sensory information, such as the detection of the new boundaries.

Changing reward experiment During the reward phase, the reward location is changed after a fixed number of fetches.

Optimization The model hyper-parameters such as the constants for the neural dynamics and the behaviour selection have been optimized through a evolutionary algorithm. Initally, an initial population of individuals with different random genomes (string of hyper-parameters values) is sampled and evaluated. Then, the population of a new generation is constructed from the first by combining and mutating the genomes of the top ranked individuals from the previous generation. More in details, for the sampling of the new generation we used the Covariance Matrix Adaptation algorithm, in which the shape of distribution of genome values is iteratively adapted according to the recent performances. The fitness used for evaluating the individuals was of a tuple consisting of the number of collected rewards and the number of collisions. In particular, the latter was calculated starting from the time when the reward position was first discovered, in this way at least the stochasticity of the exploratory behaviour was excluded.

The evolved hyper-parameters are: neural gain β, lateral inhibition threshold $\theta^{\mathrm{PC}}_{\mathrm{inh}}$, lateral distance threshold $\theta^{\mathrm{PC}}_{\mathrm{rec}}$, activity trace time constant τ^{PC}, reward modulation scale $\mathrm{c}^{\mathrm{DA}}_{\mathrm{b}}$, reward modulation spread $\mathrm{x}^{\mathrm{DA}}_{\mathrm{b}}$, boundary modulation scale $\mathrm{c}^{\mathrm{BND}}_{\mathrm{b}}$, boundary modulation spread $\mathrm{c}^{\mathrm{BND}}_{\mathrm{b}}$, reward gain modulation $\mathrm{c}^{\mathrm{DA}}_{\mathrm{a}}$, and boundary gain modulation $\mathrm{c}^{\mathrm{BND}}_{\mathrm{a}}$.