
Flexible navigation with neuromodulated cognitive maps

Anonymous Author(s)

Affiliation

Address

email

1

Abstract

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

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

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

32

1 Introduction

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

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

46 Another important component are neuromodulators. Their actions include modulation of neuronal
47 dynamics, for instance adjusting the synaptic strength, tuning place fields [11, 12, 13]. Further,
48 neuromodulators such as dopamine transmit reward signals reshaping place cell tuning [14, 15, 16, 17],
49 support novelty detection [12], and encode prediction errors [18, 13], particularly via LEC inputs
50 [19, 20]—mechanisms closely related to reinforcement learning principles.

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

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

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

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

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

82 **2 Methods**

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

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

92 **Place Cell Formation** The primary spatial representation is formed by a set of grid cell modules,
 93 each encoding a periodic tiling of 2D space, which directly maps to a toroidal manifold \mathbf{T}^2 (Fig.
 94 1b). Departing from traditional grid cell modeling approaches [52, 53], we generate population
 95 activity directly via Gaussian tuning over the torus, continuously updated using the agent's velocity
 96 vector—an approach used in prior work [8].

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

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

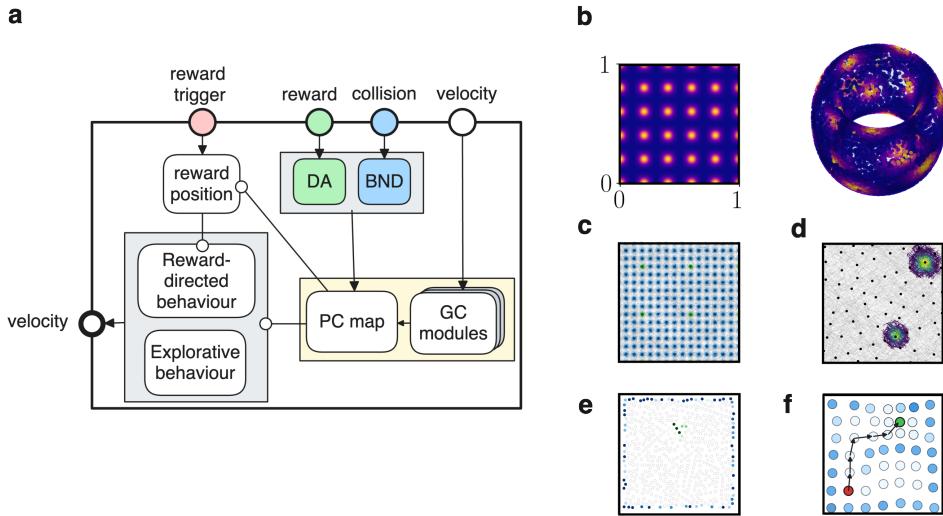


Figure 1: **Model layout and spatial representations** - **a:** the full architecture of the model, consisting of three main sensory input, targeting the two modulators and the cognitive map module, and the executive components, represented by a policy module, two behavioral programs and a reward receiver. **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.

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

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

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

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

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

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

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

$$\Delta \mathbf{W}_i^{GC,PC} = c^k v^k \varphi_{\sigma^k} (\mathbf{u}^{GC} - \mathbf{W}_i^{GC,PC}) \quad (2)$$

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

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

$$\beta_i = c_a^k m_i \bar{\beta} + (1 - m_i) \bar{\beta} \quad (3)$$

130 where c_a^k is a scaling gain parameter, for which a value of 1 signifies that no modulation takes place.

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

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

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

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

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

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

158 Optimization of the model hyper-parameters was carried out using the evolutionary Covariance-Matrix
 159 Adaptation strategy (CMA-ES) [62] with a population size of 90. The choice of the hyper-parameters
 160 to evolved took into account the type of dynamcis for which it was difficult to select pre-defined
 161 values; the number of evolved variables was 10. The use of such method derived from the impracticality
 162 of other optimization algorithms, also given the non-differentiability of several dynamic. Viable
 163 alternatives were based on reinforcement learning, Bayesian and grid search, but evolution was more
 164 appealing for exploration and visualization of the parameter space exploration. For evaluating the

165 individuals it was used a multi-objective fitness function: maximization of the reward count, and
 166 minimization of the collision count from the time the reward position was discovered.

167 3 Results

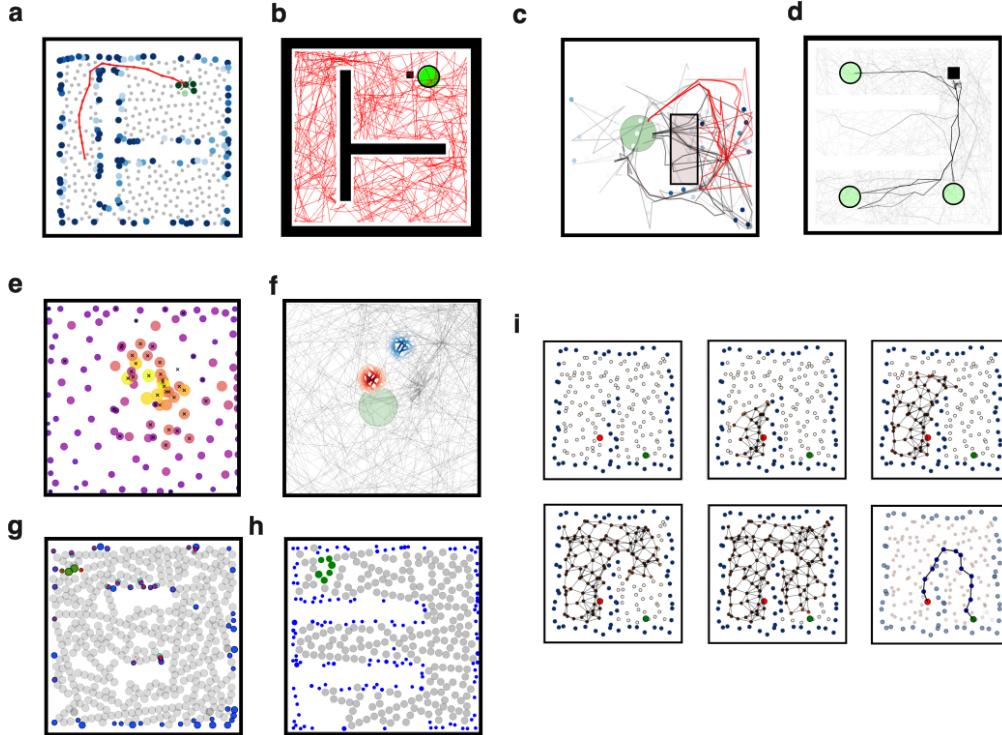


Figure 2: **Cognitive maps and performance results** - **a:** a cognitive map over a space, together with the plan (red line) to reach a target location from a starting position.. **b:** the same environment but with the reward (green circle), trajectory (red line), agent position (black square). - **c:** 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.

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

173 The left panel of Fig. 2a-b displays place cells associated with collisions and reward events, signaling
 174 boundaries (in blue) and reward (in green) locations. The overlap of these two representations and the
 175 non-modulated place cells (in grey) is what we refer to as a cognitive map, since these are the main
 176 sources of spatial and contextual information used during planned navigation, whose path is depicted
 177 as a gray line. The right panel instead portrays the actual environment with walls (black), reward
 178 location (green), and multiple trajectories (red). During exploration, the main areas were visited
 179 until the reward position was located and the goal-directed navigation dominated, as highlighted

180 by the density of the path lines. Considering the position of the walls and corners, the layout of
 181 this environment does not always make the target locations visible, as it is a non-convex area and
 182 therefore can be classified as wayfinding [63]. The challenge of not being able to use straight lines
 183 is overcome by the graph approach using local data and the consideration of boundary place cells,
 184 allowing the agent to plan accordingly. In addition, considering the Gaussian receptive fields and
 185 the approximately homogenous distribution of place field centers supported by lateral inhibition, the
 186 calculated path accounting for node-length also implicitly minimizes effective path-length, although
 187 not necessarily exactly. Figure 2g visualizes part of the path-finding process.

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

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

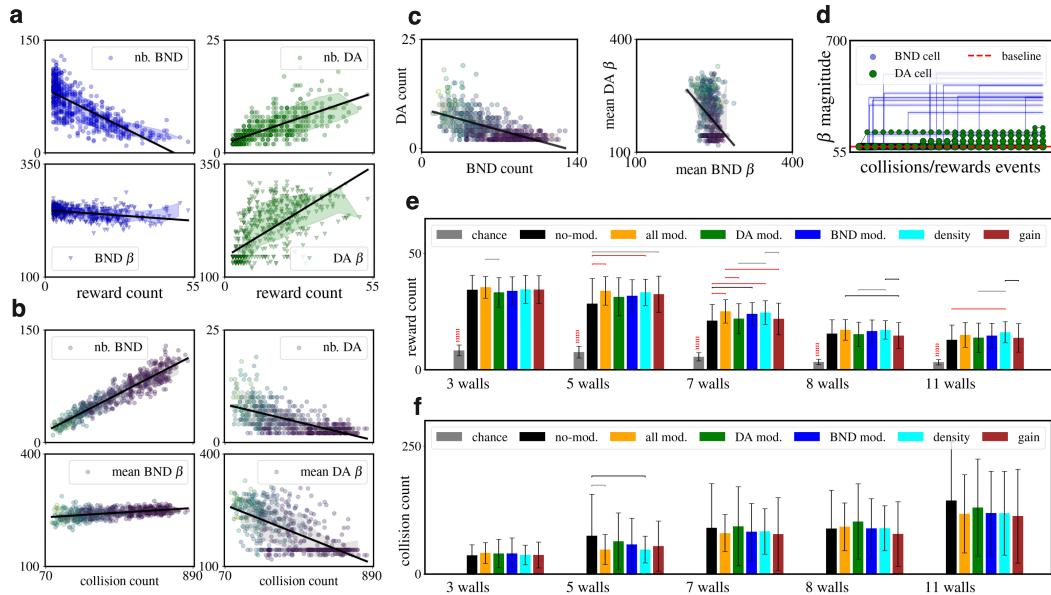


Figure 3: **Cognitive maps and performance results** - **a:** effect of reward count on reward and boundary modulated cells (green and blue respectively), both in total count (top row) and average gain modulation magnitude (bottom row); simulation of 160 independent runs. - **b:** similar plot but with 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 modulatory 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

200 **Adaptive goal representation through sensory error** Then, we tested the adaptability to environ-
 201 mental changes. In this scenario, the reward object was moved after being fetched a fixed number
 202 of times. Here, the difficulty was to unlearning previous locations and discovering new ones, in a
 203 protocol similar to [39]. In Fig. 2d is reported the set of trajectories over many trials with the reward

204 displaced in three possible locations. The agent was capable of planning behavior, as earlier, but
205 also exploring and finding the new rewards, as shown by the density of lines. Whenever a goal path
206 resulted in a failed prediction, the DA-based sensory error weakened the association between the
207 place cells and the reward signal, leading to an extinction of its representation at that location.

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

210 **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 c_a^{BND} , c_a^{DA} that yielded the best results were both larger than 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 environments, 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 distribution of these re-location vector is biased towards the rewarding area, a similar observation to Fig. 2e in which the cells involved by modulation of field positions are marked with a black cross. In Fig. 3d 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 beta. This feature can be considered another adaptation property.

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

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

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

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

239 Figure 2d showcases the distribution of place cells with the circle size and color representing the
240 node degree, which aligns discretely with the position of reward and the density. Furthermore, in Fig.
241 2e the place field of one cell is shown, before and after several reward occurrences and consequent
242 center relocation.

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

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

254 As for modulation, there is a strong tendency of increasing the magnitude of the gain for both reward
255 c_a^{DA} and collision c_a^{BND} events, with the effect of further reducing the size the active place fields. This

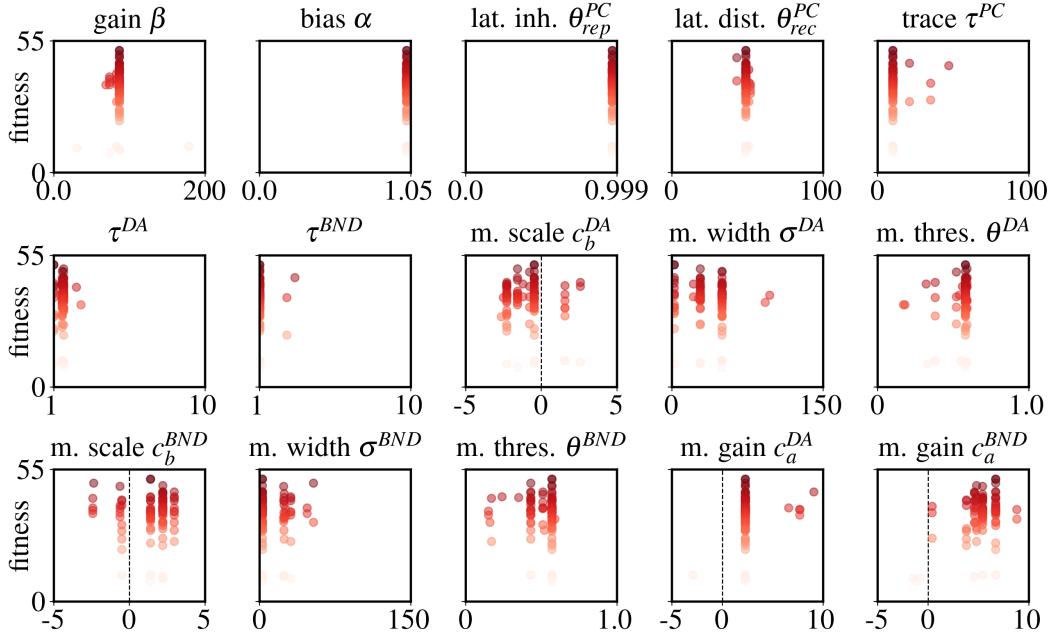


Figure 4: Distribution of evolved hyper-parameters - Results relative to the last generation, from a run with population size of 90 individuals. The hyper-parameters are: neural activation gain β , neural activation bias α , lateral inhibition threshold θ_{inh}^{PC} , lateral distance threshold θ_{rec}^{PC} , 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} .

256 result is in the direction of experimental observation showing shrinking of place fields near objects
257 and walls [64, 65].

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

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

265 4 Discussion

266 Exploration and planning in novel and past environments are essential behaviors of animals, directly
267 affecting their success in spatial understanding and goal reaching.

268 An important element behind these abilities is the formation of a map of their surroundings, enriched
269 with information gained from new experiences, known as a cognitive map. In this work, we presented
270 a rate network model inspired by the CA1 hippocampal region [?], which differs from previous
271 approaches as it operates online, uses neuromodulation-based plasticity and does not rely on external
272 coordinates.

273 We used simplified grid cells together with synaptic plasticity as a mechanism to develop information-
274 rich representations based on place cells updated through experience, grouped with common perspec-
275 tives on cognitive maps [?]. In the spirit of minimizing the geometric assumptions in the neural
276 space, we treated the generated place network as a topological graph, with sensory information added

277 locally through the action of neuromodulators. This idea aligned with the concept of a *labeled graph*
278 [? ?], however, it is also true that no metric violations were possible in these settings.

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

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

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

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

311 Lastly, the relevance of active modulation in the neuronal properties of place cells was confirmed
312 through simulated ablation experiments.

313 These tests reported a significant impact of altering the density of place cells on the total count of
314 collected rewards. In general, these results are consistent with the experimental observations of
315 alteration of place cells following reward events [? ?], in particular in terms of increased clustering
316 of cells [? ?], reminiscent of changes in firing rate after contextual changes [? ?].

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

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

338
339

340 **Acknowledgements & Statements**

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

348 **References**

- 349 [1] Reginald Golledge, Dan Jacobson, Rob Kitchin, and Mark Blades. Cognitive Maps, Spatial
350 Abilities, and Human Wayfinding. *GEOGRAPHICAL REVIEW OF JAPAN SERIES B*, 73:93–
351 104, December 2000.
- 352 [2] Russell A. Epstein and Lindsay K. Vass. Neural systems for landmark-based wayfinding
353 in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
354 369(1635):20120533, February 2014.
- 355 [3] Francesca Sargolini, Marianne Fyhn, Torkel Hafting, Bruce L. McNaughton, Menno P. Witter,
356 May-Britt Moser, and Edvard I. Moser. Conjunctive Representation of Position, Direction, and
357 Velocity in Entorhinal Cortex. *Science*, 312(5774):758–762, May 2006.
- 358 [4] Emilio Kropff, James E. Carmichael, May-Britt Moser, and Edvard I. Moser. Speed cells in the
359 medial entorhinal cortex. *Nature*, 523(7561):419–424, July 2015.
- 360 [5] Trygve Solstad, Edvard I. Moser, and Gaute T. Einevoll. From grid cells to place cells: A
361 mathematical model. *Hippocampus*, 16(12):1026–1031, 2006.
- 362 [6] Daniel Bush, Caswell Barry, and Neil Burgess. What do grid cells contribute to place cell firing?
363 *Trends in Neurosciences*, 37(3):136–145, March 2014.
- 364 [7] Torsten Neher, Amir Hossein Azizi, and Sen Cheng. From grid cells to place cells with realistic
365 field sizes. *PLOS ONE*, 12(7):e0181618, July 2017.
- 366 [8] Tianyi Li, Angelo Arleo, and Denis Sheynikhovich. *Modeling Place Cells and Grid Cells in
367 Multi-Compartment Environments: Hippocampal-Entorhinal Loop as a Multisensory Integration
368 Circuit*. April 2019.
- 369 [9] Olesia M. Bilash, Spyridon Chavlis, Cara D. Johnson, Panayiota Poirazi, and Jayeeta Basu.
370 Lateral entorhinal cortex inputs modulate hippocampal dendritic excitability by recruiting a
371 local disinhibitory microcircuit. *Cell Reports*, 42(1):111962, January 2023.
- 372 [10] Flavio Donato, Anja Xu Schwartzlose, and Renan Augusto Viana Mendes. How Do You Build
373 a Cognitive Map? The Development of Circuits and Computations for the Representation of
374 Space in the Brain. *Annual Review of Neuroscience*, 46(Volume 46, 2023):281–299, July 2023.
- 375 [11] John E. Lisman and Anthony A. Grace. The Hippocampal-VTA Loop: Controlling the Entry of
376 Information into Long-Term Memory. *Neuron*, 46(5):703–713, June 2005.

- 377 [12] Adrian J. Duszkiewicz, Colin G. McNamara, Tomonori Takeuchi, and Lisa Genzel. Novelty
378 and Dopaminergic Modulation of Memory Persistence: A Tale of Two Systems. *Trends in*
379 *Neurosciences*, 42(2):102–114, February 2019.
- 380 [13] Wolfram Schultz, Peter Dayan, and P. Read Montague. A Neural Substrate of Prediction and
381 Reward. *Science*, 275(5306):1593–1599, March 1997.
- 382 [14] Kimberly A. Kempadoo, Eugene V. Mosharov, Se Joon Choi, David Sulzer, and Eric R.
383 Kandel. Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial
384 learning and memory. *Proceedings of the National Academy of Sciences*, 113(51):14835–14840,
385 December 2016.
- 386 [15] Aude Retailleau and Thomas Boraud. The Michelin red guide of the brain: Role of dopamine
387 in goal-oriented navigation. *Frontiers in Systems Neuroscience*, 8, March 2014.
- 388 [16] Katie C. Bittner, Aaron D. Milstein, Christine Grienberger, Sandro Romani, and Jeffrey C.
389 Magee. Behavioral time scale synaptic plasticity underlies CA1 place fields. *Science*,
390 357(6355):1033–1036, September 2017.
- 391 [17] Alexandra Mansell Kaufman, Tristan Geiller, and Attila Losonczy. A Role for the Locus
392 Coeruleus in Hippocampal CA1 Place Cell Reorganization during Spatial Reward Learning.
393 *Neuron*, 105(6):1018–1026.e4, March 2020.
- 394 [18] Denis Sheynikhovich, Satoru Otani, Jing Bai, and Angelo Arleo. Long-term memory, synaptic
395 plasticity and dopamine in rodent medial prefrontal cortex: Role in executive functions. *Frontiers*
396 in *Behavioral Neuroscience*, 16, January 2023.
- 397 [19] Kei M. Igarashi, Hiroshi T. Ito, Edvard I. Moser, and May-Britt Moser. Functional diversity
398 along the transverse axis of hippocampal area CA1. *FEBS Letters*, 588(15):2470–2476, August
399 2014.
- 400 [20] Hiroshi T. Ito and Erin M. Schuman. Functional division of hippocampal area CA1 via
401 modulatory gating of entorhinal cortical inputs. *Hippocampus*, 22(2):372–387, 2012.
- 402 [21] Michael Peer, Iva K. Brunec, Nora S. Newcombe, and Russell A. Epstein. Structuring Knowl-
403 edge with Cognitive Maps and Cognitive Graphs. *Trends in cognitive sciences*, 25(1):37–54,
404 January 2021.
- 405 [22] Elizabeth R. Chrastil and William H. Warren. From Cognitive Maps to Cognitive Graphs. *PLoS*
406 ONE, 9(11):e112544, November 2014.
- 407 [23] Steffen Werner, Bernd Krieg-Brückner, and Theo Herrmann. Modelling Navigational Knowl-
408 edge by Route Graphs. In Christian Freksa, Christopher Habel, Wilfried Brauer, and Karl F.
409 Wender, editors, *Spatial Cognition II: Integrating Abstract Theories, Empirical Studies, Formal*
410 *Methods, and Practical Applications*, pages 295–316. Springer, Berlin, Heidelberg, 2000.
- 411 [24] C. R. Gallistel and Audrey E. Cramer. Computations on Metric Maps in Mammals: Getting
412 Oriented and Choosing a Multi-Destination Route. *Journal of Experimental Biology*, 199(1):211–
413 217, January 1996.
- 414 [25] Michael Peer, Catherine Nadar, and Russell A. Epstein. The format of the cognitive map
415 depends on the structure of the environment. *Journal of Experimental Psychology: General*,
416 153(1):224–240, January 2024.
- 417 [26] William H. Warren. Non-Euclidean navigation. *Journal of Experimental Biology*,
418 222(Suppl_1):jeb187971, February 2019.
- 419 [27] Mark Wagner. Comparing the psychophysical and geometric characteristics of spatial perception
420 and cognitive maps. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*,
421 15(1):6–21, 2008.
- 422 [28] Rainer Rothkegel, Karl F. Wender, and Sabine Schumacher. Judging Spatial Relations from
423 Memory. In Christian Freksa, Christopher Habel, and Karl F. Wender, editors, *Spatial Cognition:*
424 *An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, pages 79–
425 105. Springer, Berlin, Heidelberg, 1998.

- 426 [29] Tobias Meilinger. The Network of Reference Frames Theory: A Synthesis of Graphs and
 427 Cognitive Maps. In Christian Freksa, Nora S. Newcombe, Peter Gärdenfors, and Stefan Wölfl,
 428 editors, *Spatial Cognition VI. Learning, Reasoning, and Talking about Space*, pages 344–360,
 429 Berlin, Heidelberg, 2008. Springer.
- 430 [30] Jane X. Wang, Zeb Kurth-Nelson, Dhruva Tirumala, Hubert Soyer, Joel Z. Leibo, Remi Munos,
 431 Charles Blundell, Dharshan Kumaran, and Matt Botvinick. Learning to reinforcement learn,
 432 January 2017.
- 433 [31] Victor R. Schinazi, Daniele Nardi, Nora S. Newcombe, Thomas F. Shipley, and Russell A.
 434 Epstein. Hippocampal size predicts rapid learning of a cognitive map in humans. *Hippocampus*,
 435 23(6):515–528, 2013.
- 436 [32] Bruno Poucet. Spatial cognitive maps in animals: New hypotheses on their structure and neural
 437 mechanisms. *Psychological Review*, 100(2):163–182, 1993.
- 438 [33] Paul Stoewer, Achim Schilling, Andreas Maier, and Patrick Krauss. Neural network based
 439 formation of cognitive maps of semantic spaces and the putative emergence of abstract concepts.
 440 *Scientific Reports*, 13(1):3644, March 2023.
- 441 [34] William de Cothi, Nils Nyberg, Eva-Maria Griesbauer, Carole Ghanamé, Fiona Zisch, Julie M.
 442 Lefort, Lydia Fletcher, Coco Newton, Sophie Renaudineau, Daniel Bendor, Roddy Grieves,
 443 Éléonore Duvelle, Caswell Barry, and Hugo J. Spiers. Predictive maps in rats and humans for
 444 spatial navigation. *Current Biology*, 32(17):3676–3689.e5, September 2022.
- 445 [35] James C. R. Whittington, Timothy H. Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil
 446 Burgess, and Timothy E. J. Behrens. The Tolman-Eichenbaum Machine: Unifying Space and
 447 Relational Memory through Generalization in the Hippocampal Formation. *Cell*, 183(5):1249–
 448 1263.e23, November 2020.
- 449 [36] Andrea Banino, Caswell Barry, Benigno Uria, Charles Blundell, Timothy Lillicrap, Piotr
 450 Mirowski, Alexander Pritzel, Martin J. Chadwick, Thomas Degris, Joseph Modayil, Greg
 451 Wayne, Hubert Soyer, Fabio Viola, Brian Zhang, Ross Goroshin, Neil Rabinowitz, Razvan
 452 Pascanu, Charlie Beattie, Stig Petersen, Amir Sadik, Stephen Gaffney, Helen King, Koray
 453 Kavukcuoglu, Demis Hassabis, Raia Hadsell, and Dharshan Kumaran. Vector-based navigation
 454 using grid-like representations in artificial agents. *Nature*, 557(7705):429–433, May 2018.
- 455 [37] Ben Sorscher, Gabriel C. Mel, Samuel A. Ocko, Lisa M. Giocomo, and Surya Ganguli. A unified
 456 theory for the computational and mechanistic origins of grid cells. *Neuron*, 111(1):121–137.e13,
 457 January 2023.
- 458 [38] Christopher J. Cueva and Xue-Xin Wei. Emergence of grid-like representations by training
 459 recurrent neural networks to perform spatial localization, March 2018.
- 460 [39] Zuzanna Brzozko, Susanna B. Mierau, and Ole Paulsen. Neuromodulation of Spike-Timing-
 461 Dependent Plasticity: Past, Present, and Future. *Neuron*, 103(4):563–581, August 2019.
- 462 [40] Marielena Sosa, Mark H. Plitt, and Lisa M. Giocomo. Hippocampal sequences span experience
 463 relative to rewards. *bioRxiv*, page 2023.12.27.573490, February 2024.
- 464 [41] Abdullahi Ali, Nasir Ahmad, Elgar de Groot, Marcel A. J. van Gerven, and Tim C. Kietzmann.
 465 Predictive coding is a consequence of energy efficiency in recurrent neural networks, November
 466 2021.
- 467 [42] Jacopo Bono, Sara Zannone, Victor Pedrosa, and Claudia Clopath. Learning predictive cognitive
 468 maps with spiking neurons during behavior and replays. *eLife*, 12:e80671, March 2023.
- 469 [43] Wolfram Schultz. Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*,
 470 18(1):23–32, March 2016.
- 471 [44] Jeffrey B. Inglis, Vivian V. Valentin, and F. Gregory Ashby. Modulation of Dopamine for
 472 Adaptive Learning: A Neurocomputational Model. *Computational brain & behavior*, 4(1):34–
 473 52, March 2021.

- 474 [45] Philippe N. Tobler, Christopher D. Fiorillo, and Wolfram Schultz. Adaptive Coding of Reward
475 Value by Dopamine Neurons. *Science*, 307(5715):1642–1645, March 2005.
- 476 [46] Roshan Cools. Chemistry of the Adaptive Mind: Lessons from Dopamine. *Neuron*, 104(1):113–
477 131, October 2019.
- 478 [47] Aaron D Milstein, Yiding Li, Katie C Bittner, Christine Grienberger, Ivan Soltesz, Jeffrey C
479 Magee, and Sandro Romani. Bidirectional synaptic plasticity rapidly modifies hippocampal
480 representations. *eLife*, 10:e73046, December 2021.
- 481 [48] André A. Fenton. Remapping revisited: How the hippocampus represents different spaces.
482 *Nature Reviews Neuroscience*, 25(6):428–448, June 2024.
- 483 [49] Luxin Zhou and Yong Gu. Cortical Mechanisms of Multisensory Linear Self-motion Perception.
484 *Neuroscience Bulletin*, 39(1):125–137, July 2022.
- 485 [50] Steven J. Jerjian, Devin R. Harsch, and Christopher R. Fetsch. Self-motion perception and
486 sequential decision-making: Where are we heading? *Philosophical Transactions of the Royal
487 Society B: Biological Sciences*, 378(1886):20220333, August 2023.
- 488 [51] Ian Q. Whishaw and Brian L. Brooks. Calibrating space: Exploration is important for allothetic
489 and idiothetic navigation. *Hippocampus*, 9(6):659–667, 1999.
- 490 [52] Yuri Dabaghian. Grid Cells, Border Cells and Discrete Complex Analysis.
- 491 [53] Vemund Sigmundson Schøyen, Kosio Beshkov, Markus Borud Pettersen, Erik Hermansen,
492 Konstantin Holzhausen, Anders Malthe-Sørensen, Marianne Fyhn, and Mikkel Elle Lep-
493 perød. Hexagons all the way down: Grid cells as a conformal isometric map of space. *PLOS
494 Computational Biology*, 21(2):e1012804, February 2025.
- 495 [54] Richard S Sutton and Andrew G Barto. The Reinforcement Learning Problem.
- 496 [55] P. R. Montague, P. Dayan, and T. J. Sejnowski. A framework for mesencephalic dopamine
497 systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16(5):1936–1947,
498 March 1996.
- 499 [56] Seetha Krishnan, Chad Heer, Chery Cherian, and Mark E. J. Sheffield. Reward expectation
500 extinction restructures and degrades CA1 spatial maps through loss of a dopaminergic reward
501 proximity signal. *Nature Communications*, 13(1):6662, November 2022.
- 502 [57] Colin G. McNamara, Álvaro Tejero-Cantero, Stéphanie Trouche, Natalia Campo-Urriza, and
503 David Dupret. Dopaminergic neurons promote hippocampal reactivation and spatial memory
504 persistence. *Nature Neuroscience*, 17(12):1658–1660, December 2014.
- 505 [58] Frédéric Michon, Esther Krul, Jyh-Jang Sun, and Fabian Kloosterman. Single-trial dynamics
506 of hippocampal spatial representations are modulated by reward value. *Current biology: CB*,
507 31(20):4423–4435.e5, October 2021.
- 508 [59] Philip Shamash and Tiago Branco. Mice identify subgoal locations through an action-driven
509 mapping process, December 2021.
- 510 [60] Ben Sorscher, Gabriel Mel, Surya Ganguli, and Samuel Ocko. A unified theory for the origin of
511 grid cells through the lens of pattern formation. In *Advances in Neural Information Processing
512 Systems*, volume 32. Curran Associates, Inc., 2019.
- 513 [61] Zuzanna Brzosko, Sara Zannone, Wolfram Schultz, Claudia Clopath, and Ole Paulsen. Se-
514 quential neuromodulation of Hebbian plasticity offers mechanism for effective reward-based
515 navigation. *eLife*, 6:e27756, 2017.
- 516 [62] Christian Igel, Nikolaus Hansen, and Stefan Roth. Covariance Matrix Adaptation for Multi-
517 objective Optimization. *Evolutionary Computation*, 15(1):1–28, March 2007.
- 518 [63] Tobias Meilinger, Marianne Strickrodt, and Heinrich H. Bühlhoff. Qualitative differences in
519 memory for vista and environmental spaces are caused by opaque borders, not movement or
520 successive presentation. *Cognition*, 155:77–95, October 2016.

- 521 [64] S.N. Burke, A.P. Maurer, S. Nematollahi, A.R. Uprety, J.L. Wallace, and C.A. Barnes. The Influence
 522 of Objects on Place Field Expression and Size in Distal Hippocampal CA1. *Hippocampus*,
 523 21(7):783–801, July 2011.
- 524 [65] Sander Tanni, William De Cothi, and Caswell Barry. State transitions in the statistically stable
 525 place cell population correspond to rate of perceptual change. *Current Biology*, 32(16):3505–
 526 3514.e7, August 2022.
- 527 [66] Michael I. Anderson and Kathryn J. Jeffery. Heterogeneous Modulation of Place Cell Firing by
 528 Changes in Context. *Journal of Neuroscience*, 23(26):8827–8835, October 2003.
- 529 [67] Inah Lee, Amy L. Griffin, Eric A. Zilli, Howard Eichenbaum, and Michael E. Hasselmo. Gradual
 530 Translocation of Spatial Correlates of Neuronal Firing in the Hippocampus toward Prospective
 531 Reward Locations. *Neuron*, 51(5):639–650, September 2006.
- 532 [68] Marianne Fyhn, Sturla Molden, Stig Hollup, May-Britt Moser, and Edvard I. Moser. Hippocam-
 533 pal Neurons Responding to First-Time Dislocation of a Target Object. *Neuron*, 35(3):555–566,
 534 August 2002.
- 535 [69] David Dupret, Joseph O’Neill, Barty Pleydell-Bouverie, and Jozsef Csicsvari. The reorgani-
 536 zation and reactivation of hippocampal maps predict spatial memory performance. *Nature
 537 Neuroscience*, 13(8):995–1002, August 2010.

538 5 Appendix

539 5.1 Grid cell module

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

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

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

551 A grid cell module l of size N^{gc} is identified by a set of positions defined over a square centered
 552 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
 553 space has boundary conditions for each dimension, such that, for instance, when $x_t + s_l^{gc} \cdot v_x > 2$
 554 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
 555 velocity in the local space of the module l with respect to the real global agent velocity $\mathbf{v} = \{v_x, v_y\}$.
 556 When the module is initialized, the starting positions of its cells are uniformly distributed over the
 557 square forming a lattice. When the agent is reset in a new position at the beginning of new trial, a
 558 displacement vector is applied to the last cells positions such that the mapping between the module
 559 local space and the global environment is preserved.

560 The firing rate vector of each cell is determined with respect to a 2D Gaussian tuning curve centered
 561 at the origin at $(0, 0)$, and it is calculated as

562 $r_i = \exp\left(-\frac{x_i^2+y_i^2}{\sigma_l^{gc}}\right)$, where σ_l^{gc} is the width of the tuning curve for module l . An illustration of the
 563 receptive field over a 2D environment and a toroidal space is reported in Figure 5a-b.

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

566
 567 In our model, each grid cell had a tuning width of 0.04. They were defined as 8 modules of size 36,
 568 and the relative speed scales were $\{1, 0.8, 0.7, 0.5, 0.4, 0.3, 0.2, 0.1, 0.07\}$.

569 **5.2 Place cells**

570 **Tuning formation** The tuning of a new place cell is simply defined as the current GC population
 571 vector \mathbf{u}_i^{GC} , and its index is that of the first silent cell, which is added to the forward weight matrix
 572 $\mathbf{W}_i^{GC,PC} \leftarrow \mathbf{u}_i^{GC}$.

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

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

582 **Activity** The current firing rate of the PC population is determined by the cosine similarity between
 583 the GC input and the forward weight matrix, then passed through a generalized sigmoid $\phi(z) =$
 584 $[1 + \exp(-\beta(z - \alpha))]^{-1}$. The parameter α represents the activation threshold, or horizontal offset,
 585 while β the gain, or steepness.

$$u_i^{PC} = \phi \left(\cos \left(\mathbf{u}^{GC}, \mathbf{W}_i^{GC,PC} \right) \right) \quad (4)$$

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

$$m_i = -m_i/\tau^{PC} + u_i \quad (5)$$

587 It is used as a proxy for a memory trace.

588
 589 In the model, a PC population is defined by its average place field size, determining the granularity
 590 of the representation of the place. In plot 5b it is illustrated an example of place cells layer tuning
 591 obtained from a continuous trajectory over a square environment.

592 **5.3 Modulation**

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

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

597 $\tau_{DA} = 2, \tau^{BND} = 1$

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

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

601 where the weight contribution of the Hebbian update, and η^k is the learning rate: $\eta^{DA} = 0.9, \eta^{BND} =$
 602 0.9. Additionally, connections values are kept non-negative.

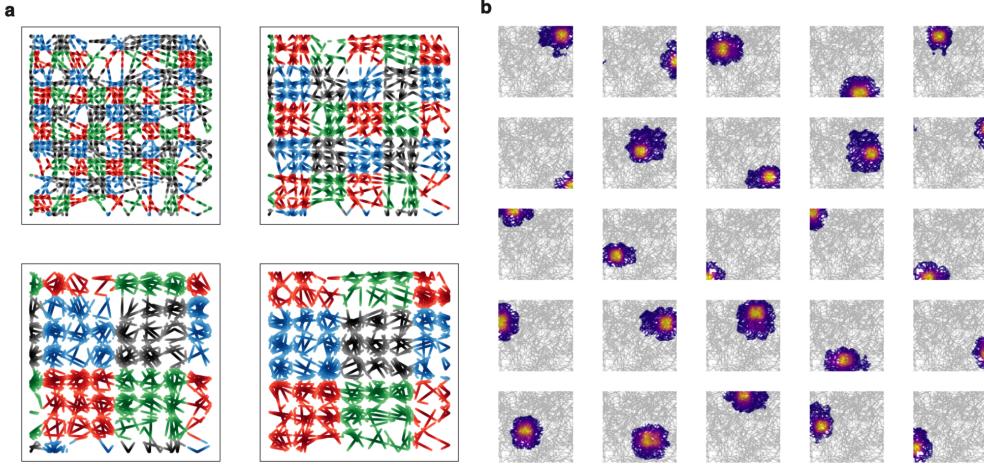


Figure 5: Place fields obtained from grid cells activity - a: grid cell modules with different granularity represented over a continuous trajectory in an open space. For visualization purposes, each module is represented as composed of four sub-modules of 9 grid cells each, whose periodic tuning generates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

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

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

$$\beta_i = c_a^k m_i \bar{\beta} + (1 - m_i) \bar{\beta} \quad (8)$$

606 where c_a^k is a scaling gain parameter, and if it is 1 then no modulation takes place.

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

$$\mathbf{q}_i = c_b^k v^k \exp \left(-\frac{\|\mathbf{W}_i^{GC,PC} - \mathbf{u}^{GC}\|}{\sigma^k} \right) \quad (9)$$

610 where c_b^k is a scaling relocation parameter, while σ^k the width of the Gaussian distance. This
611 displaced is used to move in GC activity space and get the new GC population vector to use as
612 tuning pattern.

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

617 5.4 Decision making

618 Behaviour selection logic

619 The possible behaviours are *exploration* and *exploitation*, and an action is defined as a 2D velocity
620 vector. For exploration, an action can be generated either as random navigation, using a polar vector
621 of fixed magnitude (the speed) and angle sampled from a uniform distribution, or as a step within
622 a goal-directed navigation plan to reach a random destination, which corresponds to a randomly

623 sampled existing place cells. In the goal-directed navigation the magnitude of the velocity vector
 624 is less or equal than a fixed speed value, depending on the distance from the next target position in
 625 the plan. Instead, for the exploitation behaviour, the action is a step within a goal-directed navigation
 626 towards the reward location. The behaviour selection process depends on the experience of collision,
 627 the presence of a plan, and the success in the navigation planning. A diagram of this logic is reported
 628 in Figure 6.

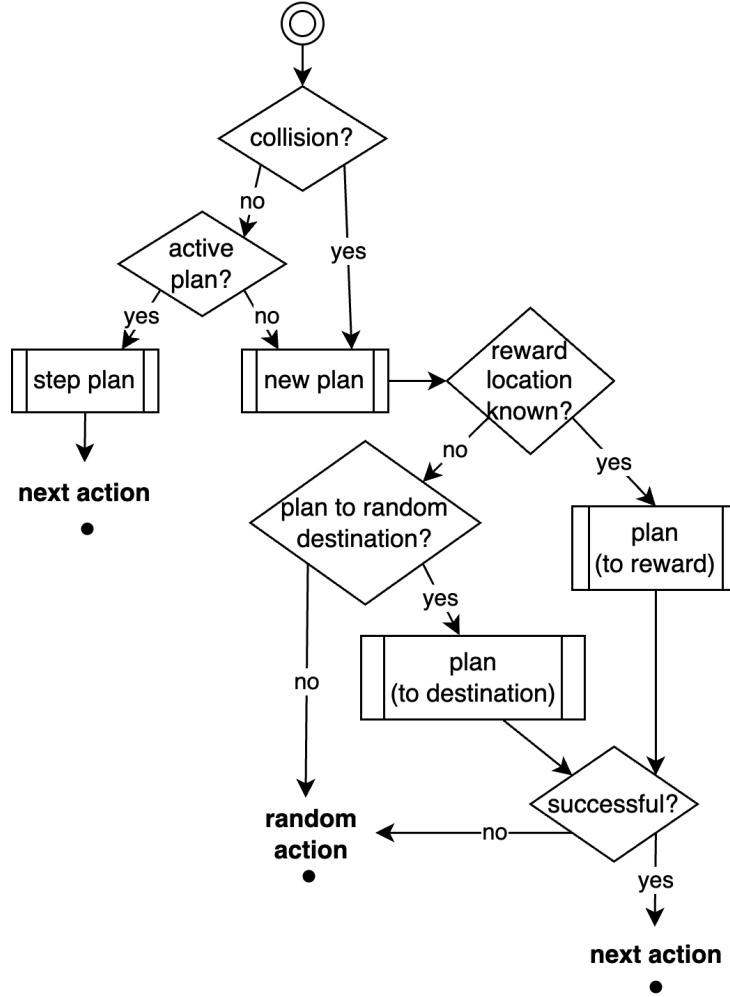


Figure 6: Diagram of the behaviour selection process

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

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

633 **Path-planning algorithm**
 634

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

```
1: Initialize activity vector  $a \leftarrow \mathbf{0} \in \mathbb{R}^n$ ; set  $a_s \leftarrow 1$ 
2: Initialize history list  $H \leftarrow []$                                  $\triangleright$  — Forward propagation phase —
3: for  $t = 1$  to  $\text{MAX\_PATH\_DEPTH}$  do
4:    $a \leftarrow C \cdot a$ 
5:    $a \leftarrow a \circ w$                                                $\triangleright$  Element-wise multiplication with node weights
6:    $a \leftarrow \sigma(4(a - 0.6))$                                       $\triangleright$  Apply sigmoidal activation:  $\sigma(x) = \frac{1}{1+e^{-x}}$ 
7:    $a_i \leftarrow 0$  if  $a_i < 0.1$  (thresholding)
8:   Append  $a$  to  $H$ 
9:   if  $a_e > 0$  then
10:    break
11:   end if
12: end for
13: if maximum depth reached then
14:   return []                                                        $\triangleright$  No path found
15: end if
16: if  $|H| < 3$  then
17:   return  $[s, e]$                                                $\triangleright$  Path is trivially short
18: end if
19: Initialize path index stack  $G \leftarrow [[e]]$ 
20: for  $t = 1$  to  $\text{MAX\_PATH\_DEPTH}$  do
21:   Let  $m \leftarrow C_G[t-1]$                                           $\triangleright$  Get neighbors of current group
22:   if  $m_s > 0$  then
23:     break
24:   end if
25:    $a \leftarrow H[-(t+1)] \circ m$ 
26:   Append  $\{i : a_i > 0\}$  to  $G$ 
27: end for                                                  $\triangleright$  — Path reconstruction —
28:  $P \leftarrow [e]$ 
29: for  $t = 1$  to  $|G| - 1$  do
30:   Initialize  $a \leftarrow \mathbf{0} \in \mathbb{R}^n$ 
31:   Set  $a_i \leftarrow 1$  for all  $i \in G[t]$ 
32:    $a \leftarrow a \circ C_P[-1]$ 
33:   if  $\sum a = 0$  then
34:     return []                                                        $\triangleright$  No valid neighbor
35:   else
36:     Choose  $j \in \{i : a_i > 0\}$  uniformly at random
37:     Append  $j$  to  $P$ 
38:   end if
39: end for
40: Append  $s$  to  $P$ 
41: return  $\text{reverse}(P)$ 
```

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

638 **5.5 Environments**

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

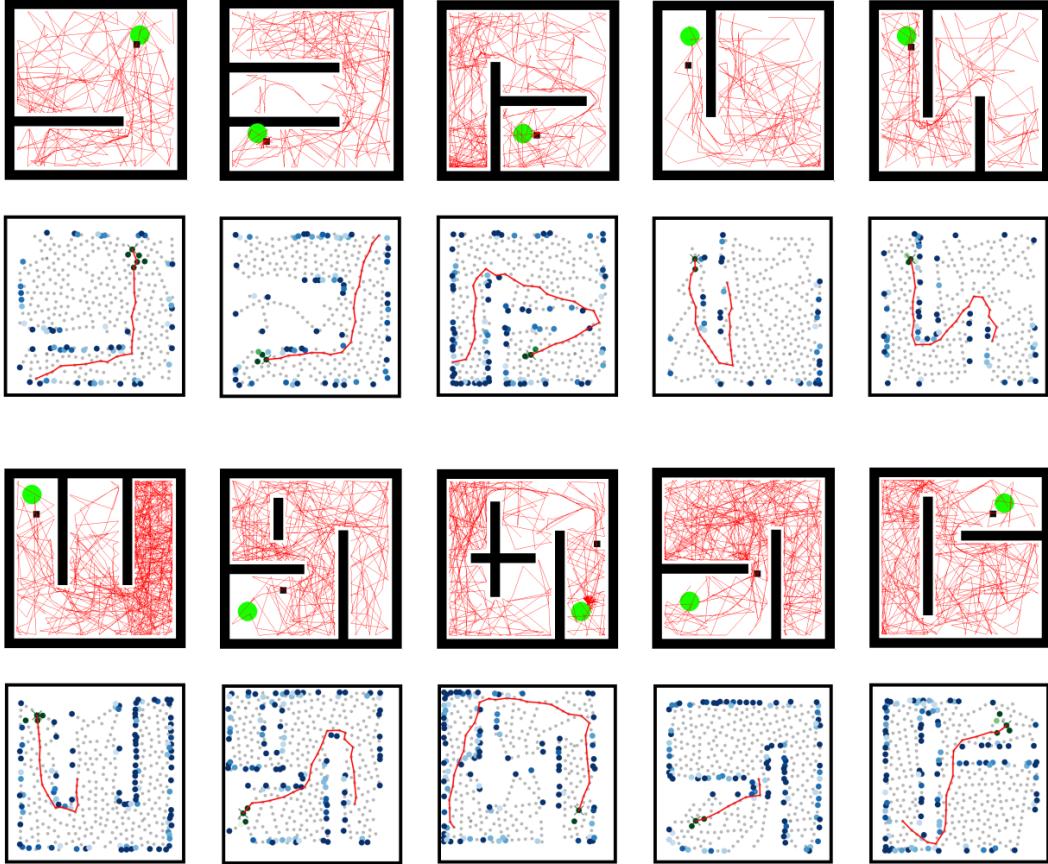


Figure 7: Sample of generated environments

643 The reward object is defined as a circle with size the 5% of the total environment area. When the
 644 agent position is within its boundary, it is provided a binary signal $R \sim \mathcal{B}(p_r)$ drawn from a Bernoulli
 645 with probability $p_r = 0.6$. The duration of the reward fetching is set to 2 time steps.
 646 The agent object is defined as a square with size the 3.3% of the total environment area.
 647 The testing protocol was inspired by the behavior of animals that venture into new territories in search
 648 of food. It was divided into two parts:

- 649 • **exploration phase:** the agent was placed in a random location within the environment for
 650 10'000 time steps. In this phase the reward is not present. Further, in order to force greater
 651 exploration of the environment, every 3'000 steps it was teleported to another random
 652 location. This external intervention was meant to mitigate the randomness in the exploratory
 653 behavioural strategy of the agent.
- 654 • **reward phase:** a reward is insert in a random location, and it is available to be discovered.
 655 When it is encountered, the agent is teleported to a random location within the environement,
 656 and after a fixed about 100 time step it is enabled its reward-seeking behaviour, in the form
 657 of goal-directed navigation. The total duration of this phase is set to 20'000 time steps.

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

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

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

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

678 The evolved hyper-parameters are: neural gain β , lateral inhibition threshold θ_{inh}^{PC} , lateral distance
679 threshold θ_{rec}^{PC} , activity trace time constant τ^{PC} , reward modulation scale c_b^{DA} , reward modulation
680 spread x_b^{DA} , boundary modulation scale c_b^{BND} , boundary modulation spread c_b^{BND} , reward gain
681 modulation c_a^{DA} , and boundary gain modulation c_a^{BND} .