Flexible navigation with neuromodulated cognitive maps

Krubeal Danieli

1 Introduction

During navigation, animals dynamically create rich representations of the environment, forming personalized cognitive maps. The hippocampal area CA1 features spatial cells that adapt based on behavior and internal states. Computational models have usually obtained spatial tuning by training a deep recurrent network for solving navigation tasks such as path integration [1, 2, 3], lasting multiple numerous epochs and using backpropagation. However, these training methods do not closely align with real-time local learning paradigms used by animals.
In this study, it is introduced a rate model that generates place cells in one-shot as the agent navigates the environment by simply assigning the current spatial observation to a selected neuron while ensuring a sparse representation (*i.e.* spaced place fields).

An important ingredient for the learning dynamics of our model is neuromodulation. Neuromodulation is an important ingredient for biological neuronal dynamics, with different molecules covering a wide range of functions. Previous models [4, 5] inspired by experimental results crafted a simple spiking plasticity rule for reward-directed navigation where acetilcholine mediates explorative behaviour and dopamine reinforces memory of reward locations. Other approached using deep artificial networks have applied neuromodulation in conjuction with other training practices, such as dropout probability [6]. In this work, modulators are described as synaptic resources that are consumed by plasticity events, and their dynamics are modelled as leaky integrators. Further, acetilcholine is used to mediate the generation of new place fields, while dopamine mediates the slow remapping of the place centers in conjunction with a reward signal. The concentration of acetilcholine is affected by the presence of active neurons or by the occurrence of a weight update. Dopamine, on the other hand, is influenced by the presence of a reward.

This model successfully creates a representation of visited areas and recurrent connections are defined among similarly tuned cells. Importantly, plasticity hyper-parameters such as the equilibrium concentration and decay time-constant of modulators influence the density of place cells, impacting the encoding of behaviorally relevant information [7].

This network is then used to solve a goal-directed navigation task, where the agent is trained to reach a target location. The agent is equipped with a policy that modulates the exploration behaviour and the decision-making process.

2 Methods

The model is constructed around the concept of a cognitive map, which an agent builds by freely navigating a closed environment and reaching a discovered goal location. The full schema of its components its illustrated in plot 1-**a** below. The architecture relies on the core assumption that the agent has receives minimal external information, consisting solely of a reward and collision input as two binary values. These two signals are used to enrich the cognitive map with experience-dependent data, which is then used to guide the agent's behavior.

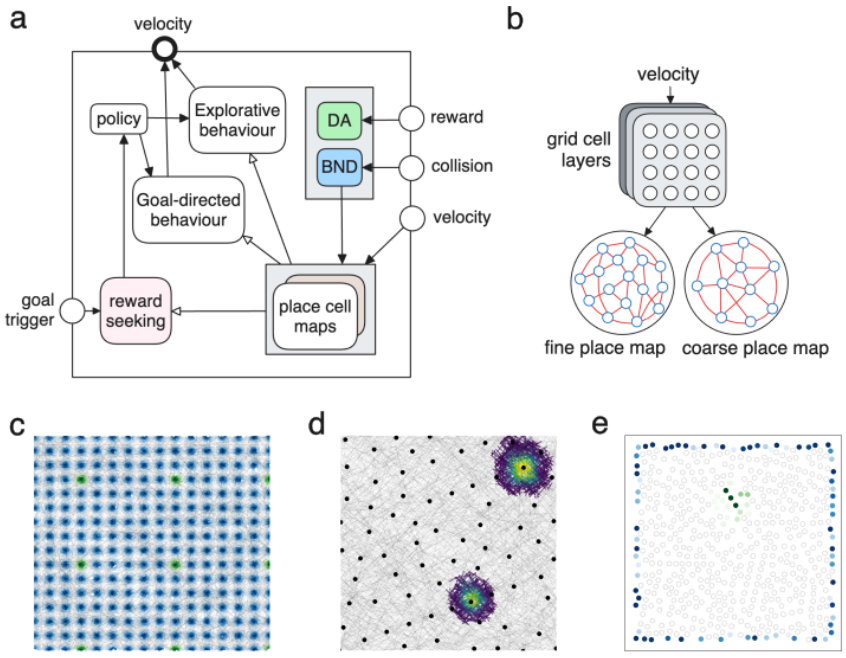

Figure 1: Model layout and spatial representations - **a**: *the full architecture of the model, consisting of three main sensory input, targeting the two modulators and the cognitive map module, and the executive components, represented by a policy module, two behavioural programs and a reward receiver.* **b**: *the cognitive map component, organized with a stack of grid cell modules receiving the velocity input and projecting to two layer of place cells with different place field granularity.* **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.*

The formation of the spatial representation is instead based on idiothetic information, which is the agent's perception of self-motion. In particular, here we assume this cue to be the factual velocity vector, namely the actual displacement of the agent in the environment. In the brain, this signal is thought to result from the integration of inertial and relative motion cues.

2.1 Place cell map

The formation of place cells is obtained from the activity of a set of grid cells organized into modules, or layers. This simple feed-forward architecture is depicted in plot 1-**b**. A grid cell module i has been defined as a set of N^{gc} neurons with gaussian tuning curve evenly distributed over the surface of a two dimensional torus $\mathbf{T}^2$. When the agent moves in the environment, a two dimensional Euclidean space $\mathbf{R}^2$, with a velocity $\mathbf{v} = \{x, y\}$, its position on the torus is updated by the same vector but scaled by a speed scalar s_i^{gc} local to the grid cell module i, which determines its periodicity in space. The initial position on the torus is randomly chosen at the beginning of each episode, since what matters is the sequence of displacements without any reference to a meaningful origin.

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 size 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. In plot 1-**c** is shown the activity of a grid cell module over a trajectory, with the periodicity underlined by the cell in green.

The activity of all grid cell modules, indicated as $\mathbf{u}^{GC}$, is then projected down to two independent layers of initially un-tuned cells, whose feed-forward weights $\mathbf{W}^{GC,PC1}$; $\mathbf{W}^{GC,PC2}$ are initialized at zero. As the agent moves and the grid cells activity changes, if no neurons within a place cell layer are active, then one is randomly chosen and its weights are set to the current (at time t) grid cells' population vector $\mathbf{W}_i^{GC,PC} \leftarrow \mathbf{u}_t^{GC}$. For the plasticity process to be completed, it is also checked the possible overlap with other cells in the same layer, effectively accounting for lateral inhibition. This mechanism is implemented by computing the cosine similarity with the weight vector of the other tuned cells and comparing it with a threshold θ_{rep}^{PC}, with the possibility of aborting the plasticity process if the similarity is too high.

The activity of a tuned place cell i is given, again, by the cosine similarity between the current grid cells' population vector and the weight vector of the cell:

$$\mathbf{u}_i^{PC} = \phi\left(\cos\left(\mathbf{u}^{GC}, \mathbf{W}_i^{GC,PC}\right)\right) \tag{1}$$

where ϕ is a generalized sigmoid function $\phi(z) = [1 + \exp(-\beta(z - \alpha))]^{-1}$ with gain β and threshold α. The two layers of place cells differ in the size of their place fields. This feature is affected by the sensitivity of a cell tuning with

respect to the grid cell activation, determined by the parameters of the sigmoid, and the strength of the lateral inhibition, determined by the similarity threshold. Within a layer, the connections between cells are calculated by the same cosine similariy, but compared against a different threshold $\theta_{\text{rec}}^{\text{PC}}$. One layer is set to be more fine-grained, with an overall higher density of place cells over the space, while the other is more coarse-grained, with overall large place field sizes.

In figure 1-**d** are shown the centers of one of the fine-grained layer and the activity of two cells, with their place field highlighted as an heatmap.

2.2 Neuromodulators

The fine-grained layer of place cells constitutes the main cognitive map of the agent, since it captures the environment with greater detail. The neuromodulators are operationalized as analog sensors of meaningful environmental events, here reward and collision, and map directly to the place cells through plastic connections $\mathrm{W}^{\mathrm{k,PC}}$. For each neuromodulator k, it is defined a leaky variable v^{k} that accumulates the corresponding signal I over time, and decays exponentially to zero in the absence of inputs with time constant τ^{k}:

$$\dot{\mathrm{v}}^{\mathrm{k}} = -\mathrm{v}^{\mathrm{k}}/\tau^{\mathrm{k}} + \mathrm{I} \tag{2}$$

This variable is then paired with the activity of each place cell i for updating the synaptic weights in a Hebbian fashion:

$$\Delta \mathbf{W}_{\mathrm{i}}^{\mathrm{k}} = \eta^{\mathrm{k}} \mathrm{v}^{\mathrm{k}} \mathbf{u}_{\mathrm{i}}^{\mathrm{PC}} \tag{3}$$

where η^{k} is the neuromodulator-specific learning rate.

On the one hand, the reward modulation, signed as $\mathbf{W}^{\mathrm{DA}}$, is sensitive to the instantaneous presence of reward, defined as a boolean value. Over time, its coupling with the population vector $\mathbf{u}_{\mathrm{t}}^{\mathrm{PC}}$ delineates a region of the environment where the reward has been experienced. On the other hand, the collision modulation, referred to as $\mathbf{W}^{\mathrm{BND}}$, signals the occurrence of a collision with a boundary, which is again given as a boolean. After enough events, the profile of the resulting weight matrix with the place cells provides an approximation of the shape of the environment given by its boundaries. From the perspective of the agent, this intuition of the topology of its surroundings is crucial for effectively planning routes to target locations.

At each moment during navigation, the weight matrices $\mathbf{W}^{\mathrm{DA,PC}}$, $\mathbf{W}^{\mathrm{BND,PC}}$ act as scalar fields over the neural space of the place cells, and their simultaneous contributions delineate what in this work is referred to as a cognitive map. In plot 1-**e** is shown the activity of the two neuromodulators over the fine-grained place cells map, showcasing the bounds of the environment and the reward location.

2.3 Policy and behavior

In this work, the first interest was to test the usefulness of our simple cognitive map built from minimal assumptions for the tasks of exploration and goal-

directed navigation. To this end, we defined a simple hard-coded policy that toggles between these two behaviours according to the presence of a goal signal, externally provided, and the presence of an actual goal representation, taken care of by a special component called *reward seeking*, depicted in the pink box of plot 1-**a**. Exploration is accomplished by a random walk, with a variable number of steps in the same direction to avoid stagnation, and occasional plans to visit random positions within the known map, again for limiting stagnation. Goal-directed navigation, either for reaching a random position for exploration or the actual reward location, is achieved by calculating the shortest path between the closest place cells centers of the current and target positions. The place cells are hence treated as nodes of a graph, and their connections constitute its edges. We use a Dijkstra algorithm applied to the coarse-grained layer, which contains less and more spread out cells and it is thus cheaper to compute, to derive a coarse-grained plan. However, in the case the agent gets stuck or the distance to the target is shorter than the cells' distance, the planning switches to the other layer for devising a fine-grained plan, which it is followed until either the target or the next node in the coarse-grained are reached.

The advantage of this dual-layer planning lies in its flexibility, as it lightens the computational load of planning by exploing the sparser and lighter map and only invokes the detailed one when necessary. In the process of behaviour, learning does not occur explicitly, but it is instead accounted for in the online formation of the cognitive map.

2.4 Map re-configuration

Our second aim was to investigate the possibility of online alteration of the place cells density, in terms of its potential performance improvement. This action is motivated by the consolidated phenomenon of hippocampal rate remapping, for which the place cells change their firing pattern according to contextual shifts (**ref**). Additionally, there is growing experimental evidence that place fields can be moved in space following behaviourally relevant events, such as the occurrence of reward, according to a plasticity rule known as behavioural time-scale plasticity (BTSP) (**ref**).

In our model, we associated this process to both collision and reward signals, whose location is set to be the center towards which the cells within a certain radius r_{BND}, r_{DA} are pulled. The centers of the cells involved are shifted with a force proportional to the Gaussian distance from the center of the signal, and the strength is weighted by a parameter λ_{BND}, λ_{DA}.

Our working hypothesis is that the online re-configuration of the cognitive map can lead to a representation of the environment more tailored with the agent experience. This changes introduced by this mechanism should then be reflected in the adaptability and navigation abilities.

References

[1] 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.

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

[3] 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.

[4] Zuzanna Brzosko, Wolfram Schultz, and Ole Paulsen. Retroactive modulation of spike timing-dependent plasticity by dopamine. *eLife*, 4:e09685, October 2015.

[5] 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.

[6] Jie Mei, Rouzbeh Meshkinnejad, and Yalda Mohsenzadeh. Effects of neuromodulation-inspired mechanisms on the performance of deep neural networks in a spatial learning task. *iScience*, 26(2):106026, February 2023.

[7] 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.