A Computational Model for a Multi-Goal Spatial Navigation Task inspired by Rodent Studies

Martin Llofriu
Computer Science and Eng Dept
University of South Florida
Tampa, FL, USA
mllofriualon@mail.usf.edu

Pablo Scleidorovich
Computer Science and Eng Dept
University of South Florida
Tampa, FL, USA
pablos@mail.usf.edu

Gonzalo Tejera
Facultad de Ingeniería
Universidad de la República
Montevideo, Uruguay
gtejera@fing.edu.uy

Marco Contreras
Facultad de Ciencias
Universidad Mayor
Santiago, Chile
mcontrerasabar@gmail.com

Tatiana Pelc
Psychology Dept
University of Arizona
Tucson, AZ, USA
tatiana.pelc@gmail.com

Jean-Marc Fellous
Psychology Dept
University of Arizona
Tucson, AZ, USA
fellous@email.arizona.edu

Alfredo Weitzenfeld
Computer Science and Eng Dept
University of South Florida
Tampa, FL, USA
aweitzenfeld@usf.edu

Abstract—**We present a biologically-inspired computational model of the rodent hippocampus based on recent studies of the hippocampus showing that its longitudinal axis is involved in complex spatial navigation. While both poles of the hippocampus, i.e. septal (dorsal) and temporal (ventral), encode spatial information; the septal area has traditionally been attributed more to navigation and action selection; whereas the temporal pole has been more involved with learning and motivation. In this work we hypothesize that the septal-temporal organization of the hippocampus axis also provides a multi-scale spatial representation that may be exploited during complex rodent navigation. To test this hypothesis, we developed a multi-scale model of the hippocampus evaluated it with a simulated rat on a multi-goal task, initially in a simplified environment, and then on a more complex environment where multiple obstacles are introduced. In addition to the hippocampus providing a spatial representation of the environment, the model includes an actor-critic framework for the motivated learning of the different tasks.**

Keywords—spatial cognition, computational neuroscience, neural networks, learning, navigation

I. Introduction

Spatial navigation in rodents has been studied for quite some time suggesting the existence of a cognitive map in the rat's hippocampus [1-2]. The biological basis that supports the cognitive map has received a lot of attention. However, how this information is functionally used for navigational purposes is not fully clear. This paper extends our understanding of spatial navigation in rodents by developing new computational models based on some of the latest rodent studies of the hippocampus.

Many spatially tuned cells are found in the hippocampal formation and related structures in rodents and other mammals. In particular, place cells firing in the hippocampus are highly correlated with the position of the animal in an allocentric frame of reference [3]. In the enthorinal cortex, grid cells fire when the animal is at the vertices of a grid laid out over the environment [4]. Additionally, head direction cells signal the orientation of the animal's head, also in an allocentric frame of reference [5].

Classical studies have shown multi-scale activation field gradients along the dorso-ventral (septo-temporal) axis of both place cells and grid cells, with smaller place fields towards the septal portions and larger fields towards the temporal portions [6-7].

While there are multiple examples of single scale computational models of spatial navigation inspired by rodent studies of the hippocampus (e.g. [8-16]), limited work has been devoted to exploring the navigational purpose of multi-scale spatial representations in the hippocampus. In our previous work [17-18], we analyzed a simple circular open maze to show the theoretical advantages of larger scales of representation during the learning of a simple single goal oriented task.

The goal of our new computational model is to evaluate the role of different place field sizes in relation to the spatial complexity of the environment. In particular, we extend our previous computational model and navigation task from single goal to multi-goal navigation based on our most recent experimental studies also involving the introduction of obstacles in the environment [19-20].

II. Task and Methods

A rat is trained to learn a goal-oriented navigation task and then perform a recall session on in modified environment where obstacles are introduced [20-21]. Eight feeders or goals are laid over a circular open field maze, where a LED light is placed above each feeder as a cue used during leaning. Fig. 1a shows the layout of the maze during training, and Fig. 1b shows the introduction of obstacles during recall.

During each experiment, a subset of three feeders, known as the *set*, are selected to give rewards, whereas the other ones do not have reward (sugar water). The set of 3 feeders is fixed throughout the experiment which consists of three different phases:

NSF IIS Robust Intelligence research collaboration grant #1117303

1. A ***non-delayed cue*** phase where each feeder from the set is randomly chosen and its associated light is flashed until the rat feeds from it. This was repeated *100* times.
2. A ***delayed cue*** phase. The rat is allowed to go through the feeders freely and without flashing cues. If too much time passes without the rat feeding from one of the 3-feeder target, one of the correct feeder lights is flashed until the rat reaches it. This phase is executed until the rat consecutively reaches 15 feeders from the set, with no more than 2 cues and without making any mistake.
3. A ***delayed cue with obstacles*** phase that only differs from the previous phase in that a set of obstacles is placed in the environment. Obstacles consist on 12.5 cm wide barriers. Some of them are put against the maze wall, whereas the rest are placed towards the middle of the maze. Some of the barriers near the wall are placed together to form a bigger (25 cm) barrier.

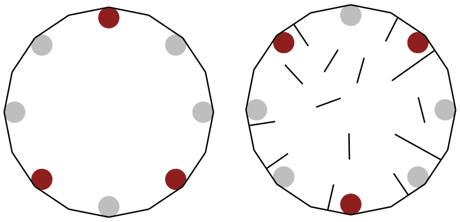

Fig. 1. (left) The maze layout. Circles represent the feeders, with the learning set represented in red. Black lines show the walls in the environment. (right) A sample disposition of obstacles for the recall phase. The interior black lines represent the obstacles.

III. Model

The computational model is shown in Fig. 2. The model is based on a reinforcement learning architecture that uses information provided by different brain regions in rodents: hippocampus (HPC), subiculum (SUB) and prefrontal cortex (PFC). The output of these regions is fed to a learning module comprised by: ventral tegmental area (VTA), dorso-medial striatum and ventral striatum (Nucleus Accumbens - NA).

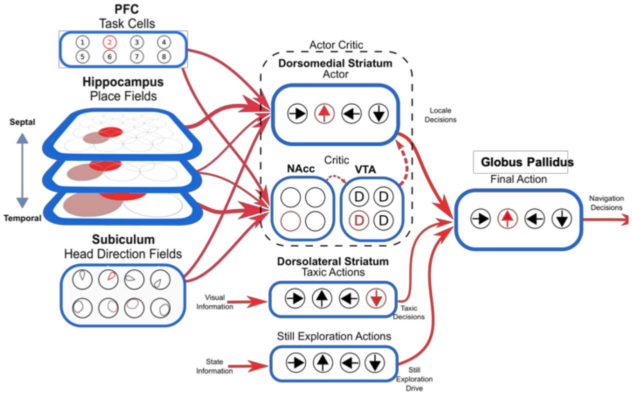

Fig. 2. The figure shows the multi-scale computational model architecture for spatial navigation.

Hippocampal place cells are modeled by having different size activation fields along the longitudinal axis. The different scales project output to a value estimating network, where information is input to the nucleus accumbens (Nacc) and relayed to the dopaminergic ventral tegmental area (VTA), and to action selection structures, composed by the dorsomedial striatum. The striatum also receives input from the PFC indicating the current state of the task and from subicular head direction cells. Dopaminergic error signals are projected to the dorsomedial striatum, where they are used to learn the associations between situations (stimulus) and actions (response). Additionally, visual information drives a taxic behavior module (dorsolateral striatum), and a still exploration module. All action selection information converges to a common structure for final action selection (Globus Pallidus), made in a winner take all fashion. Red arrows indicate connectivity, the thicker the arrow the stronger the connectivity. The level of red indicates current activation for all units (circles).

A. Place Cells

Place cells are modelled using a Gaussian kernel function which, set to 0 outside the given radius, as illustrated by Fig. 3.

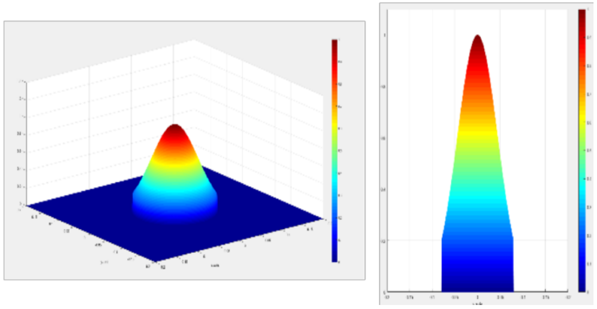

Fig. 3. Illustration of a Gaussian kernel function modeling the place cell activation field.

Equation (1) shows the kernel function.

$$K(d) = \begin{cases} 0 & \text{if} \quad d > 1 \\ e^{d^2 \cdot log\,(\alpha)} & \text{otherwise} \end{cases} \tag{1}$$

where:

- $K(d)$ is the place field kernel function.
- α is a parameter smaller than 1 representing the activation of a cell at its border.

To calculate the activation of a place cell, we first compute the distance of the rat to the place cell's center, and then we apply the kernel to the calculated distance normalized by the place cell's radius of activation as shown in (2) and (3).

$$d_i = ||\vec{x} - \vec{x}_i|| \tag{2}$$

$$PC_i(d_i) = K(\frac{d_i}{r_i}) \tag{3}$$

where:

- $\vec{x}$ is the rat position
- $\vec{x}_i$ is the center of place cell i.
- d_i is the distance from the rat's position to the center of place cell i.
- r_i is the radius of activation of place cell i

- $PC_i(d_i)$ is the activation of place cell i

Fig. 4 illustrates the place-field Gaussian kernel function for a place cell activation field. Notice that when the distance to the center of the place cell is equal to the radius of activation $(d_i = r_i)$, then the activation of the cell is equal to α, i.e. $(PC(d_i) = \alpha)$. Furthermore, when the activation is larger than the radius $(d_i > r_i)$, the activation becomes null, i.e. $(PC(d_i) = 0)$.

The radius used for dorsal and ventral place cells is 0.068m and 0.14m, respectively. This is based on [32], and scaled by the ratio of the corresponding environment sizes.

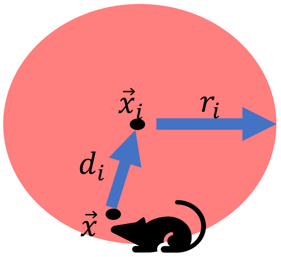

Fig. 4. Illustation of the place-field kernel function based on the rat position $\vec{x}$, its distance d_i from the center $\vec{x}_i$ of place cell i, where r_i is the radius of activation for place cell i.

B. "Obstacle" Place Cells (OPC)

An important aspect of the new model is the addition of obstacles to the environment. It has been observed that obstacles impact place cell firing patterns in different ways, including place cells being "silenced" when obstacles are found within the cell's field [22]. Placing obstacles also results in the activation of "obstacle" cells (also referred to as "wall", "boundary", or "barrier" cells), that fire only when an obstacle is present [23-24]. Additionally, these types of cell have been shown to affect place cell firing depending on which side of the obstacle they fire.

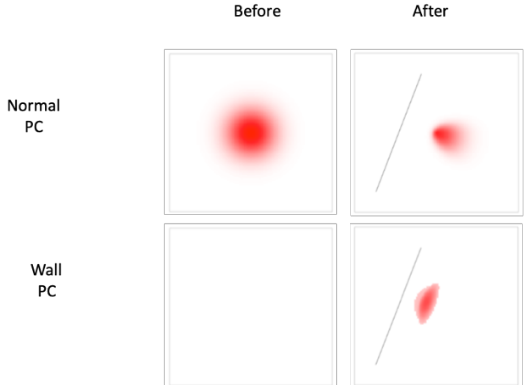

Fig. 5. Interaction between place fields and obstacles. Top row shows a normal place cell (NPC) field before and after introducing an obstacle. Bottom row shows an obstacle place cell (OPC) field before and after introducing an obstacle.

Based on these findings, we take into consideration "obstacle" place cells (OPC), when having obstacles in the environment. In our model, "normal" place cells (NPC) are considered cells whose activation is negatively modulated when nearby obstacles are introduced. On the other hand, OPCs are considered cells that activate only under the presence of nearby obstacles. The interactions of these two types of cells with an obstacle are illustrated in Fig. 5.

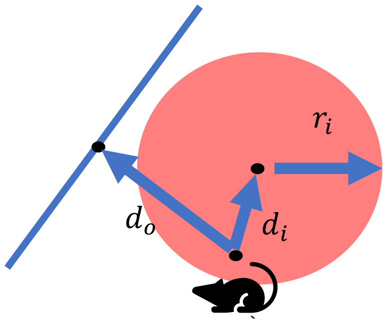

Fig. 6. Illustration of the kernel function modified by the presence of an obsacle or wall, where d_0 is the shortest (orthogoal) distance from the current rat position to the obstacle.

To model OPCs, the original equation for PCs is modified, as shown in (4), by multiplying it with a modulator function, described further on. The function takes as input the distance from the rat to the nearest obstacle or wall, as well as the distance between the rat and the place cell's center (both normalized by the PC radius). Fig. 6 illustrates the new function.

$$OPC_i(d_i, d_o) = m_i\left(\frac{d_o - d_i}{r_i}, \frac{d_o}{r_i}\right) \cdot PC_i(d_i) \qquad (4)$$

where

- OPC_i is the function to calculate the firing rate of the obstacle interactive place cell i.
- d_o is the distance from the rat to the closest obstacle.
- m_i is a function that modulates the activation of cell i according to the distance to the closest obstacle.

The modulator function serves two objectives. First, it prevents place cells from firing when an obstacle is located between the rat and the place cell's center. Second, it provides the behavior for NPCs and OPCs. To accomplish the first purpose, the function returns 0 when the distance to the PC's center is bigger than the distance to the closest obstacle. To accomplish the second, the equation of the modulator differs for NPCs and OPCs. For NPCs, it returns a sigmoid function that decreases the closer the rat is to an obstacle, while for OPCs, the modulator returns the product of two sigmoidal functions S (one of them inverted) so that the cell only activates if close to an obstacle. Equations (5) and (6) describe the modulator function.

$$m_i(d_{io}, d'_o) = \begin{cases} 0 & \text{if } d_{io} \leq 0 \\ S^1(d_{io}) & \text{if cell } i \text{ is NPC} \\ S^2(d_{io}) \cdot \left(1 - S^3(d'_o)\right) & \text{if cell } i \text{ is OPC} \end{cases} \qquad (5)$$

$$S^k(d) = \frac{1}{1+e^{-a_k \cdot d + b_k}} \quad (6)$$

where

- m_i is the modulator function for cell i
- S^k are linearly scaled sigmoid functions with different parameters.
- $a_1, a_2, a_3, b_1, b_2, b_3$ are constant parameters that linearly scale the sigmoid functions.
- d_{io} and d'_o are the input parameters provided in Eq 4.

C. Head Direction Cells (HD)

Similarly to place cells, we model head direction cell (HD) firing using the same Gaussian kernel shown in (1). To calculate the activation of a HD cell, first we compute the angular difference between the cell's preferred direction and the rat orientation as shown in (7). Then, this value is normalized by the cell's angular activation radius and used as input for the kernel function as shown in (8). Fig. 7 illustrates the concept.

$$\Delta_i = \begin{cases} 2\pi - |\theta - \theta_i| & \text{if } |\theta - \theta_i| > \pi \\ |\theta - \theta_i| & \text{otherwise} \end{cases} \quad (7)$$

$$HD_i(\Delta_i) = K\left(\frac{\Delta_i}{ar_i}\right) \quad (8)$$

where

- θ_i is the preferred direction for HD cell i.
- θ is the rat's orientation.
- Note that both θ and θ_i are assumed to be in the range $[-\pi, \pi]$.
- Δ_i is the angular difference between the rat's orientation and the cell's preferred direction.
- ar_i is the angular activation radius for HD cell i
- HD_i is the activation function for HD cell i.

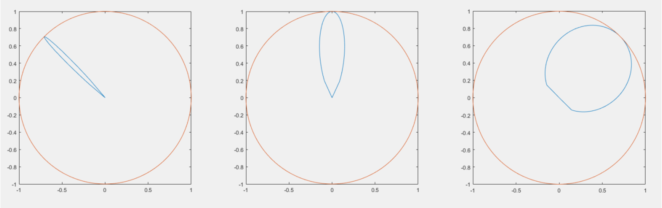

Fig. 7. Three head direction cells centered at 135°, 90° and 45° respectively. The polar plots show the activation of each cell for all posible rat orientations.

D. Task Cells (TC)

Task Cells (TC), as referred to in this work, signal the currently pursued sub-goal (e.g. feeder). This information is needed given the multi-goal nature of the task. Namely, since there is more than one goal, the navigational decisions to be performed in a certain place depends on the currently pursued goal. This aspect relates to the multiple map hypothesis [25] proposing that some place cell activity depends not only on the location but also on the current sub-task being carried out. We model the tuning of place cells to the task using information coming from the Pre-Frontal Cortex (PFC) and spatial information from the striatum. This conforms with a multiple map hypothesis, but multiple maps would be first found in the striatum, upon the convergence of the place and state information. Equation (9) describes the modeling of task cells.

$$TC_i(g) = \begin{cases} 1 & \text{if } g = g_i \\ 0 & \text{if } g \neq g_i \end{cases} \quad (9)$$

where

- TC_i is the activation of the task cell i.
- g is the current goal of the rat.
- g_i is the goal (feeder) associated to task cell i.

E. Striatal Cells (SC)

Striatal Cells (SC), both dorsal and ventral, receive inputs from place cells (HPC), head direction cells (SUB) and task cells in the Pre-Frontal Cortex (PFC). Each cell in the striatum is tuned to respond to one cell of each input source. The cell is tuned to the pursue of a particular goal, head direction and place. The resulting activation of each striatal cell is computed as the product of the corresponding place, head direction and task cells' activities as shown in (10).

$$SC_i = OPC_{j_i} \cdot HD_{k_i} \cdot TC_{l_i} \quad (10)$$

where

- SC_i is striatal cell i.
- OPC_{j_i} is the obstacle place cell associated to striatal cell i.
- HD_{k_i} is the head direction cell associated to striatal cell i.
- TC_{l_i} is the task cell associated to striatal cell i.

There are 400k striatal neurons in the model, each receiving input from an individual place cell, an individual head direction cell, and an individual task cell. From those cells, 200k receive input exclusively from dorsal hippocampus place cells, contributing only to action selection (dorsal striatum); 120k receive input from both dorsal and ventral hippocampus place cells and are split into action selection and value estimation (mid striatum); and 80k receive input exclusively from ventral hippocampus place cells, contributing to value estimation only (ventral striatum – NAcc).

F. Locale Action Learning

Our model includes a locale learning module that learns a function from where the rat performs the appropriate egocentric actions, i.e. turn left/right, go forward, or eat. Since our locale decisions are modulated by the task choice (the pursued goal), we extend the concept of location to include non-spatial aspects. A location is described by place $\vec{x}$, heading θ, and goal g, as shown in (11).

$$l = (\vec{x}, \theta, \mathrm{g}) \quad (11)$$

Dopamine release has been related to reinforcement learning and to a potential "error signal" [26]. However, this is normally done in the context of a stimulus-response conditioning task. In our model, we apply this idea but take the striatal unit population code as our input, instead of a simple stimulus. Then, the value of each combination of place, head direction, and task state is slowly modified to reflect the expected reward that the rat is going to obtain after departing from that location.

We use the Actor Critic [27] architecture because it keeps a separate representation for the value estimation module and the action selection module. This accommodates our distributed VTA-NA system for value estimation and dopamine release modulation, and the dorso-medial striatum for locale action selection.

Actor Critic methods also present a subtle advantage for our task, as they learn from negative outcomes faster than other off-policy reinforcement learning algorithms, such as Q-Learning [28]. This is important because the task involves not only learning how to arrive at the proper feeders, but also how to avoid wasting time navigating towards incorrect ones. Since our algorithm has to learn a population code, a modification of the traditional Actor Critic algorithm was implemented, and could be interpreted as RL over soft-states [29].

Equation (12) shows how the state value function is computed for a given location.

$$V^t(l) = \frac{1}{\sum_i SC_i(l)} \cdot \sum_i SC_i(l) * V_i^t \tag{12}$$

where

- V^t is the value function at time t.
- l is a given location as defined in Eq 11.
- SC_i is the activation function for striatal cell i.
- V_i^t is the value associated to striatal cell i at time t

Equation (13) shows the calculation of the error signal.

$$e_t = r_{t+1} + \gamma \cdot V^t(l_{t+1}) - V^t(l_t) \tag{13}$$

where

- e_t is the error signal at time t.
- l_t and l_{t+1} are the locations at times t and $t+1$ respectively.
- r_{t+1} is the reward received at time $t+1$.
- γ is the discount factor.

Equations (14) and (15) describe how to update the state and action values associated to each cell, respectively.

$$V_i^{t+1} = V_i^t + \alpha \cdot SC_i(l_t) \cdot e_t \tag{14}$$

$$Q_{ij}^{t+1} = Q_{ij}^t + \alpha \cdot SC_i(l_t) \cdot e_t \tag{15}$$

where

- V_i^t and V_i^{t+1} are the state values associated to striatal cell i at times t and $t+1$, respectively.
- Q_{ij}^t and Q_{ij}^{t+1} are the action values associated to striatal cell i, action j at time t and $t+1$, respectively.
- l_t is the location at time t.
- e_t is the error signal at time t.
- SC_i is the activation function for striatal cell i.
- α is a constant learning rate.

Equation (16) describes how these action values are used in the action selection process by computing a set of votes for each action j.

$$rl\_votes_t(j) = \frac{\sum_i SC_i(l_t) * Q_{ij}^t}{\sum_i SC_i(l_t)} \tag{16}$$

where

- $rl\_votes_t(j)$ are the votes at time t for action j computed from the action values.
- $SC_i(l_t)$ is the value of striatal cell i at location l_t .
- Q_{ij}^t is the action value associated to striatal cell i action j at time t

Additionally, the Actor Critic algorithm was enhanced with eligibility traces to improve the learning rate. Eligibility traces maintain a notion of the past activity of each cell. Then, upon unexpected changes in value estimation, not only the last active cells are updated, but all cells that were active in the recent past. Equations (17) and (18) show the update rule for the eligibility traces, while (19) and (20) show the modified update rule using the traces instead of the activation.

$$\lambda_i^{t+1} = \max(SC_i(l_t), \beta \cdot \lambda_i^t) \tag{17}$$

$$\lambda_{ij}^{t+1} = \begin{cases} \beta \cdot \lambda_{ij}^t \; if \; a_t \neq a_j \\ max(SC_i(l_t), \beta \cdot \lambda_{ij}^t) \end{cases} \tag{18}$$

$$V_i^{t+1} = V_i^t + \alpha \cdot \lambda_i^t \cdot e_t \tag{19}$$

$$Q_{ij}^{t+1} = Q_{ij}^t + \alpha \cdot \lambda_{ij}^t \cdot e_t \tag{20}$$

where

- λ_i^t is the eligibility trace associated to striatal cell i at time t.
- λ_{ij}^t is the eligibility trace associated to striatal cell i, action j at time t.
- β is a constant that regulates the exponential rate of decay of the traces.

G. HPC Layers and Connectivity

In the model, place cells are organized into layers, where each layer contains cells with different activation place fields along the longitudinal axis. Three layers are included in the model, corresponding to septal, middle and temporal HPC, each representing a different activation field size.

A distinction between dorsal and ventral striatum has been suggested in the framework of reinforcement learning and the actor critic implementation, where dorsal is associated with stimulus-response learning (actor) and ventral to value learning (critic) [30].

As previously shown in Fig. 2, septal cells are connected only to the action selection module and temporal cells are connected to the value estimation module, corresponding to the dorsal and ventral striatum respectively. Each cell in the intermediate layer is connected with each module with probability *0.5*.

H. Taxic Modules

Reinforcement learning algorithms usually devote a lot of time to random exploration of the state-action space. In a navigational task, this would correspond to a rat that moves randomly with no directionality at all, until it learns a reasonable policy. Rats, in general, show great directionality in their movements while performing initial exploration. In the model, we have incorporated three taxic modules to guide the rat to visual stimuli:

- The first taxic module guides the rat towards flashing feeders. This module reflects prior knowledge that the rat has about flashing feeders.
- The second taxic module guides the rat to non-flashing feeders.
- The third module guides the rat to obstacle endpoints, providing a way to navigate through a maze of obstacles when no feeders are visible.

All three modules vote on each possible movement action depending on the expected reward of getting to the given visual stimulus. In addition, the votes are inversely proportional to the number of steps it would take the rat to reach the feeder or obstacle. Equation (21) summarizes the above:

$$taxic\_votes_t^k\ (j) = vr_k + sr_k * sn_t^k \qquad (21)$$

where:

- $taxic\_votes_t^k(j)$ are the votes for action j for taxic module k at time t
- vr_k is a system parameter representing the expected reward associated with the visual stimulus for taxic module k.
- sr_k is a small negative reward given after each step to account for the motion effort. The value differs for the different taxic modules.
- sn_t^k represents the number of steps needed to reach the stimulus associated to taxic module k, computed from the angular and linear distance.

In addition to voting on each action, the feeder-related taxic modules contribute to value estimation of a given place or situation by modifying the error signal, as described in and (22) and (23).

$$V_T^t = \max_{k,j}\ taxic\_\ votes_t^k(j) \qquad (22)$$

$$e_t = e_t^{rl}\ + (\gamma_T \cdot V_T^{t+1} - V_T^t) \qquad (23)$$

where

- e_t^{rl} is the reinforcement learning error.
- γ_T is a discount factor for the taxic value.
- V_T^t is a state value estimation computed from the taxic modules.

If the rat is seeing a flashing feeder, the value of that location will be increased by the expected value of going to the flashing feeder. Value is also estimated using a constant expectancy, while also considering the number of steps it would take to reach the interest point. This allows the rat to learn the positive outcome of an action that takes it to a “promising place” where a feeder is first observed, before receiving the actual reward. In addition, this allows for detecting the negative outcome of trying to eat from a feeder without success. It is the contrast between the high value estimated by the taxic module and the zero outcome what produces a high error signal (or decay in dopamine release) upon failure.

I. Exploration

In contrast to what is usually done with RL algorithms, there is no continuous exploration drive built into the model; the rat learns the proper actions by navigating using the taxic strategies and observing their outcomes.

However, some situations arise in which the system as a whole cannot propose any action. This may happen due to the lack of visual stimulus, or due to a negative value estimation of all action outcomes by the actor critic or by a combination of both (the system only chooses positive valued actions).

In these cases, after the rat has been still for a certain number of simulation steps, an exploratory module takes over for a fixed and small number of steps, where the rat executes random actions.

J. Action Selection

Action selection is performed in a collaborative fashion through a voting mechanism, which instantiates the action selection mechanism of the Globus Pallidus. Then, all votes are tallied and the action with the most votes wins, in a winner take all fashion.

The step size for forward actions was 0.05 meters in the model, and the turn angle was π/16 for rotations. Note that all actions are egocentric.

IV. Results

Figs. 8 and 9 show the paths traveled by the simulated rat during learning and recall, respectively. The recall session includes obstacles in the environment.

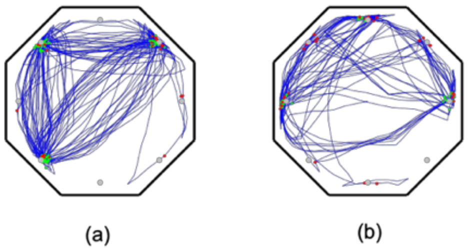

Fig. 8. Paths followed by the simulated rat during training for different rat groups. The green dots signal a successful eat attempt, whereas the red ones signal an unsuccessful one. Panels show: (a) a typical non-delayed cue training session; (b) a typical delayed cue training session.

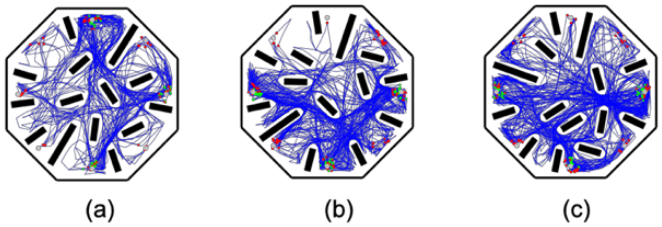

Fig. 9. Paths followed by the simulated rat during recall for different rat groups. The green dots signal a successful eat attempt, whereas the red ones signal an unsuccessful one. Panels show: (a) a delayed cue with obstacle session for the control group; (b) a delayed cue with obstacle session for the septal group; and (c) a delayed cue with obstacle session for the temporal group. Delayed cue paths were chosen from the individual with the performance closest to its group median.

Fig. 10 shows a boxplot [31] with the completion times in seconds for the delayed cued phase with small obstacles. The "Temporal" and "Septal" groups represent partial deactivation of the temporal and septal groups, respectively.

A Kruskal-Wallis test was performed over the data and significant differences were found ($p < 0.0001$). A Dunn test post-hoc pairwise comparison was made. Significant difference was found between groups Control and Septal ($p < 0.05$) and Control and Temporal ($p < 0.001$). No significant difference was found between Septal and Temporal completion times.

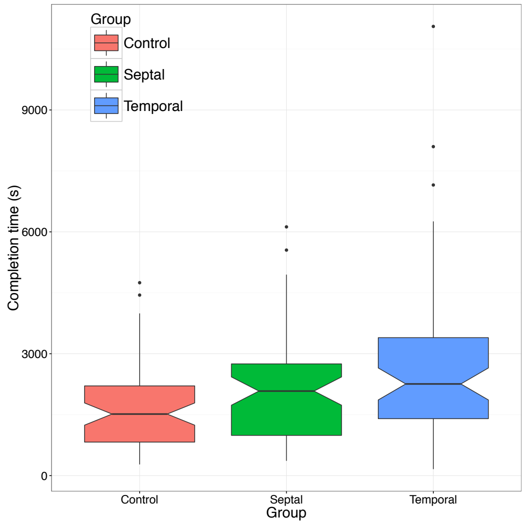

Fig. 10. Completion times for the delayed cue with small obstacles for all three groups over 64 individuals. The septal portion of the HPC was inactivated in the Septal group and the temporal portion in the Temporal. The plot shows boxplots using the 1.5 IQ outlier criteria.

We note how deactivation of either the septal or temporal groups results in a lower performance than the control group (no inactivation). Additionally, inactivation of the temporal group results in a lower performance than the inactivation of the septal group.

Fig. 11 shows the evolution of completion times for different values of the eligibility traces decay system parameter λ. Each panel shows the resulting completion times for a given value of the parameter. As eligibility traces decay faster (lower values), performance decreases for all groups. In addition, as the traces decay faster, the difference between groups becomes more apparent. The ventral group is notably more impaired under fast decaying traces conditions.

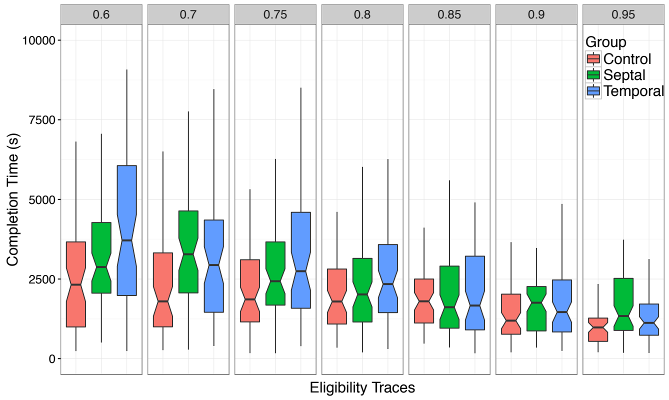

Fig. 11. Completion times in seconds for the delayed cue phase with obstacles while varying the eligibility traces decay parameter (top x axis).

V. Discussion and Conclusions

We presented a model of rat spatial navigation using reinforcement learning to find and memorize a subset of correct feeders. The model is able to reproduce data from experiments with rodents involving multiple goals and obstacles. Results are consistent with rat experimental data based on similar protocols where both septal and temporal inactivation impair the animal's performance, with the latter group showing the most impairments [19-20]. The model also shows how a differentiation of functions along the longitudinal axis of the hippocampus could explain the differences in performance observed, where the septal portion of the hippocampus is attributed the function of action selection and the temporal portion is involved in mapping locations to value.

Additionally, it can be observed that obstacle based place field inhibition provides two navigational advantages. In the first place, the disappearance of previous place cells and the appearance of new obstacle cells disrupt the learned policy. This prevents the rat from trying to execute a policy that is no longer consistent with the environment, because of an obstacle. Secondly, by allowing cells to fire only on one side of the obstacle, previously learned values do not propagate to regions that are no longer close to the place cell center. The introduction of the obstacle changes the distance the rat must travel from one point to the other, and it is useful that the value mapping changes as well.

In the future we plan to evaluate the model with varying obstacles configurations, environment sizes, and navigation learning tasks. We also plan to evaluate these tasks in physical robotic platforms to analyze the effect of real-time aspects of the environment.

ACKNOWLEDGMENT

This work is funded by NSF IIS Robust Intelligence research collaboration grant #1117303 at USF and U. Arizona entitled "Investigations of the Role of Dorsal versus Ventral Place and Grid Cells during Multi-Scale Spatial Navigation in Rats and Robots".

REFERENCES

[1] Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626.

[2] O'Keefe, J., & Nadel, L. (1978). *The Hippocampus as a Cognitive Map*. Oxford : New York: Oxford University Press.

[3] O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1).

[4] Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806. https://doi.org/10.1038/nature03721

[5] Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. Description and quantitative analysis. *The Journal of Neuroscience*, *10*(2), 420–435.

[6] Jung, M. W., Wiener, S. I., & McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience*, *14*(12), 7347–7356.

[7] Brun, V. H., Solstad, T., Kjelstrup, K. B., Fyhn, M., Witter, M. P., Moser, E. I., & Moser, M.-B. (2008). Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus*, *18*(12), 1200–1212. https://doi.org/10.1002/hipo.20504

[8] Barrera, A., & Weitzenfeld, A. (2008). Biologically-inspired robot spatial cognition based on rat neurophysiological studies. *Autonomous Robots*, *25*(1–2), 147–169. https://doi.org/10.1007/s10514-007-9074-3

[9] Caluwaerts, K., Staffa, M., N'Guyen, S., Grand, C., Dollé, L., Favre-Félix, A., Khamassi, M. (2012). A biologically inspired meta-control navigation system for the Psikharpax rat robot. *Bioinspiration & Biomimetics*, *7*(2), 25009. https://doi.org/10.1088/1748-3182/7/2/025009

[10] Filliat, D., & Meyer, J. (2002). Global Localization and Topological Map-Learning for Robot Navigation.

[11] Gaussier, P., Revel, A., Banquet, J. P., & Babeau, V. (2002). From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. *Biological Cybernetics*, *86*(1), 15–28.

[12] Guazzelli, A., Bota, M., Corbacho, F. J., & Arbib, M. A. (1998). Affordances. Motivations, and the World Graph Theory. *Adaptive Behavior*, *6*(3–4), 435–471. https://doi.org/10.1177/105971239800600305

[13] Milford, M., & Wyeth, G. (2010). Persistent navigation and mapping using a biologically inspired SLAM system. *The International Journal of Robotics Research*, *29*(9), 1131–1153.

[14] Arleo, A., & Gerstner, W. (2000). Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biological Cybernetics*, *83*(3), 287–299.

[15] Dollé, L., Sheynikhovich, D., Girard, B., Chavarriaga, R., & Guillot, A. (2010). Path planning versus cue responding: a bio-inspired model of switching between navigation strategies. *Biological Cybernetics*, *103*(4), 299–317. https://doi.org/10.1007/s00422-010-0400-z

[16] Erdem, U. M., & Hasselmo, M. (2012). A goal-directed spatial navigation model using forward trajectory planning based on grid cells. *European Journal of Neuroscience*, *35*(6), 916–931. https://doi.org/10.1111/j.1460-9568.2012.08015.x

[17] Lyttle, D., Gereke, B., Lin, K., & Fellous, J.-M. (2013). Spatial scale and place field stability in a grid-to-place cell model of the dorsoventral axis of the hippocampus. *Hippocampus*, *23*(8), 729–744. https://doi.org/10.1002/hipo.22132

[18] Llofriu, M., Tejera, G., Contreras, M., Pelc, T., Fellous, J.-M., & Weitzenfeld, A. (2015). Goal-Oriented Robot Navigation Learning using a Multi-Scale Space Representation. *Neural Networks*.

[19] Contreras, M., Pelc, T., Llofriu, M., Weitzenfeld, A. and Fellous, J.M., 2018, The ventral hippocampus is involved in multi-goal obstacle-rich spatial navigation, Hippocampus, Wiley, Aug, https://doi.org/10.1002/hipo.22993.

[20] Harland B, Contreras M and Fellous JM. (2018) A Role for the Longitudinal Axis of the Hippocampus in Multiscale Representations of Large and Complex Spatial Environments and Mnemonic Hierarchies**.** In 'Hippocampus: Plasticity and Functions', Ales Stuchlik (editor), ISBN: 978-1-78923-357-5.

[21] Jones, B., Bukoski, E., Nadel, L., & Fellous, J.-M. (2012). Remaking memories: reconsolidation updates positively motivated spatial memory in rats. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *19*(3), 91–98. https://doi.org/10.1101/lm.023408.111

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

[23] Rivard, B., Li, Y., Lenck-Santini, P.-P., Poucet, B., & Muller, R. U. (2004). Representation of Objects in Space by Two Classes of Hippocampal Pyramidal Cells. *The Journal of General Physiology*, *124*(1), 9–25. https://doi.org/10.1085/jgp.200409015

[24] Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., & Burgess, N. (2009). Boundary Vector Cells in the subiculum of the hippocampal formation. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(31), 9771–9777. https://doi.org/10.1523/JNEUROSCI.1319-09.2009

[25] Redish, A. D. (1999). Beyond the Cognitive Map: From Place Cells to Episodic Memory. MIT Press.

[26] Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27.

[27] Sutton, McAllester, Singh, Mansour (1999). Policy gradient methods for reinforcement learning with function approximation: actor-critic algorithms with value function approximation. NIPS'99 Proceedings of the 12th International Conference on Neural Information Processing Systems, pp 1057-1067.

[28] Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, Mass: MIT Press.

[29] Singh, S. P., Jaakkola, T., & Jordan, M. I. (1995). Reinforcement Learning with Soft State Aggregation. In *Advances in Neural Information Processing Systems 7* (pp. 361–368). MIT Press.

[30] Balleine, B. W., & O'Doherty, J. P. (2009). Human and Rodent Homologies in Action Control: Corticostriatal Determinants of Goal-Directed and Habitual Action. *Neuropsychopharmacology*, *35*(1), 48–69. https://doi.org/10.1038/npp.2009.131

[31] Krzywinski, M., & Altman, N. (2014). Points of Significance: Visualizing samples with box plots. *Nature Methods*, *11*(2), 119–120. https://doi.org/10.1038/nmeth.2813

[32] Jung, M. W., Wiener, S. I., & McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience, 14*(12), 7347-7356.