Comparative Experimental Studies on Spatial Memory and Learning in Rats and Robots

Alejandra Barrera · Alejandra Cáceres · Alfredo Weitzenfeld · Victor Ramirez-Amaya

Received: 30 September 2009 / Accepted: 1 September 2010 / Published online: 25 September 2010 © Springer Science+Business Media B.V. 2010

Abstract The study of behavioral and neurophysiological mechanisms involved in rat spatial cognition provides a basis for the development of computational models and robotic experimentation of goal-oriented learning tasks. These models and robotics architectures offer neurobiologists and neuroethologists alternative platforms to study, analyze and predict spatial cognition based behaviors. In this paper we present a comparative analysis of spatial cognition in rats and robots by contrasting similar goal-oriented tasks in a cyclical maze, where studies in rat spatial cognition are used to develop computational system-level models of hippocampus and striatum integrating kinesthetic and visual information to produce a cognitive map of the environment and drive robot experimentation. During training, Hebbian learning and reinforcement learning, in the form of Actor-Critic architecture, enable robots to learn the optimal route leading to a goal from a designated fixed location in the maze. During testing, robots exploit maximum expectations of reward stored within the previously acquired cognitive map to reach the goal from different starting positions. A detailed discussion of comparative experiments in rats and robots is presented contrasting learning latency while characterizing behavioral procedures during navigation such as errors associated with the selection of a non-optimal route, body rotations, normalized length of the traveled path, and hesitations. Additionally, we present results from evaluating neural activity in rats through detection of the

A. Barrera $(\boxtimes) \cdot A$. Weitzenfeld

Computer Engineering Department-Robotics and Biorobotics Laboratories, Instituto Tecnológico Autónomo de México (ITAM), Río Hondo #1, Progreso Tizapán, CP 01080, México DF, México

e-mail: abarrera@itam.mx

A. Cáceres · V. Ramirez-Amaya

Neurobiology Institute, Plastic Neural Networks Laboratory, Universidad Nacional Autónoma de México (UNAM), Boulevard Juriquilla #3001, CP 76230, Querétaro, México

V. Ramirez-Amaya e-mail: ramirez@inb.unam.mx immediate early gene Arc to verify the engagement of hippocampus and striatum in information processing while solving the cyclical maze task, such as robots use our corresponding models of those neural structures.

Keywords Hippocampus · Striatum · IEG Arc expression · Spatial learning · Cognitive map · Place recognition · Biorobotics

1 Introduction

Our research group is involved in the study of spatial memory and learning in rats and its experimentation in robots by developing computational models addressing aspects related to cognitive map generation, adaptation, and exploitation during navigation. These cognitive maps code and store spatial information for successful orientation and movement toward specific goals in space [1]. Spatial cognition models developed by our group relate learning and memory processes by (i) interaction of different brain structures to demonstrate skills associated with global and relative positioning in space, (ii) integration of traveled path, (iii) use of kinesthetic and visual cues during orientation, (iv) generation of topological-metric spatial representation of the environment, (iv) adaptation using Hebbian learning [2], (v) representation of animat internal motivational states based on hunger and thirst drives, and (vi) management of rewards implemented by reinforcement learning using an Actor-Critic architecture [3]. This learning architecture is complemented by applying backward reinforcement to successful routes followed by the animat during training thus enabling learning of explored routes. After exploration, the model exploits maximum reward expectations to guide the animat towards the goal from any given departure location. Additionally, the model implements an on-line learning process to adapt the cognitive map to changes in the physical configuration of the environment.

To evaluate the models we have developed both computer simulations and mobile robot experimentations. Tasks previously evaluated include learning of classical spatial reversal behaviors with normal and hippocampus-lesioned rats in a T-maze and in an eight-arm radial maze as carried out by O'Keefe [4]. We describe corresponding results showing a behavioral similarity between robots and rats in [5]. We also experimented with robots designing a place recognition task inspired on a classical goal-oriented navigation experiment devised by Morris [6], where we employed multiple non-cyclical T-mazes surrounded by landmarks. We describe corresponding experiments and results in [7] including modification to landmark configurations after having trained the robot to find the goal. Our current work goes beyond previously published work by developing new spatial tasks with robots and rats using a cyclical maze having external landmarks and providing different possible paths from one location to another one. This paper emphasizes place recognition and goal-oriented navigation behaviors, contrasts learning results in both robots and rats, and evaluates the engagement of rat hippocampus and striatum in information processing while solving the maze by means of studying neural activity in rats through detection of the immediate early gene Arc.

Our work contributes to the study of spatial cognition by developing a model based on recent physiological and behavioral data, and our robotic experiments also contribute to experimental neuroscience by providing a platform to test new hypotheses that might extend our latest understanding of learning and memory in rodents. The rest of this section introduces relevant research background and the main distinctive aspects of our approach against other existing models. We then present the spatial cognition model with emphasis on functional modules representing hippocampus and striatum. A detailed explanation of spatial experiments with rats and robots in the cyclical maze is provided in the following sections together with a comparative analysis of the obtained results.

1.1 Research Background

Research cycles involving animal studies, modeling, and robotic experimentation, have inspired for many years the understanding of the underlying neurophysiology and neuromechanics of biological systems while also inspiring new robotic architectures and applications [8–10]. For example, rat studies can serve as inspiration to the control of autonomous mobile robotic systems [11]. In general, animal studies, in the form of data gathering, feed brain models that generate predictions to be validated by simulation and robotic experimentation, while inspiring new ideas and hypotheses from technology back to models and animal studies, as depicted in Fig. 1.

1.1.1 Hippocampus—Place Cells

A decisive study in the understanding of brain mechanisms involved in the processing of spatial information was undoubtedly the analysis that O'Keefe and Nadel [12] conducted on the neurophysiological and neuropsychological properties of the hippocampus. Their work discovered critical participation of the hippocampus in (i) the development of high-level internal representations of allocentric spatial relations, i.e. representations of the full context and not just of the current position of the animal, and (ii) spatial learning allowing the animal to solve navigation problems that require memory of such representations.



Fig. 1 Framework for the study of animal behavior through cycles of neuroscientific experimentation, theoretical modeling, and physical robotics experimentation. Data and hypotheses obtained from neuroscientific experimentation are the basis for theoretical modeling. Resulting formal models are tested under real robotic environments. The cycle provides a general methodology where new ideas and hypotheses can be proposed in response to knowledge gaps impacting both robotics and biology

The neural substrate of such internal spatial representations was the prior discovery made by O'Keefe and Dostrovsky [13] from individual records of pyramidal cells found in hippocampal substructures CA3 and CA1. Such neurons were called place cells because they show high correlation between their firing rate and the animal location in the environment. Specifically, each of these cells exhibits high rate activation patterns of Gaussian type when the animal is in a continuous and compact area defined as the place field of the cell [13]. The activity derived from place cell population codifies the current location of the animal within a familiar environment. The overall activity is stored in the internal spatial representation referred to as cognitive map by Tolman [14].

Place cell activity appears to be dependent on the location of visual cues in the environment, since rotation of such cues causes a corresponding rotation in place fields [15]. However, place cells maintain their fields when some of the visual cues are removed from the environment [16], and even continue to respond in the dark [17]. Hence, it is assumed that the response of place cells is derived from combining kinesthetic and visual cues information [18].

According to Poucet [19], the cognitive map is generated by the acquisition of topological and metrical information from space. Topological information is related to the knowledge of the spatial relationships between places or objects, whereas metrical information is related to quantitative information about specific angular directions and distances between locations or objects. Poucet suggests that initial cognitive maps are basic topological maps that are refined as the animal explores the environment with the acquisition of metrical information provided by cells in the entorhinal cortex [20].

1.1.2 Motivation and Learning

The motivated behavior is usually oriented towards a goal that in animals may be associated with a drive such as hunger or thirst. However, motivation not only occurs in the presence of instinctive feelings, but can also be stimulated by external incentives such as food smell.

The hypothalamus is regarded as the main area of the brain where information about the internal state of the rat is combined with incentives [21]. Specifically, it is assumed that food pursuit and intake are activities controlled by the lateral hypothalamus [22], which determines the primary or immediate reward these activities produce in animals. In addition to the hypothalamus, the striatum, belonging to the basal ganglia, is also involved in the extraction of information related to rewards from environmental stimuli, and in the use of such information in the generation of goal-oriented behaviors [23, 24].

Reward information is processed in the basal ganglia by dopaminergic neurons responding to primary and secondary rewards. These responses can reflect "errors" in the prediction of rewards, thus constituting teaching signals for reinforcement learning. On the other hand, neurons in the ventral striatum (nucleus accumbens) are activated when animals predict rewards, and adapt expectation activity to new reward situations [23]. Houk et al. [25] suggested the striatum implements an Actor-Critic architecture [3], where an Adaptive Critic predicts expectations of reward values and produces an error signal to adapt reward expectations associated to different rat actions represented in this learning architecture by multiple Actor units.

In goal-oriented behaviors, it is assumed that rats are able to learn spatial tasks by associating rewards with locations in the environment (spatial learning), and rewards with procedures (procedural learning). Rats exhibit route or procedural learning strategies within environments with salient egocentric (local) cues but poor or inexistent salient allocentric (global) cues [26]. When egocentric (local) cues are poor or inexistent, rats manifest a spatial learning strategy integrating egocentric information with the most salient and stable allocentric information in the environment. Neurophysiological studies and other functional neuroanatomical studies have shown that procedural learning relies on the striatal system, whereas spatial learning employs the hippocampal system [27].

1.1.3 Immediate Early Gene Arc Expression

We can verify the engagement of different brain regions in information processing by studying neuronal activity in the brain. This can be achieved with precise anatomical resolution by detecting the expression of genes induced after physiological activity in neurons [28]. Particularly, the detection of the immediate early gene Arc has been used to study the patterns of neural activity in regions such as the hippocampus, striatum and cortex [29–31]. Both, the detection of Arc mRNA and protein, can be used to identify neurons that underwent behaviorally induced neuronal activity [30]. Arc protein expression can be detected 30 min after behavioral stimulation in the same proportion of cells that present electrophysiological activity characteristic of place cells [29, 30].

Arc detection allows mapping neuronal activity potentially in the whole brain, thus it is possible to verify if the behavior of rats in our cyclical maze engages the hippocampus and striatum in information processing such as robots use the functional modules of the spatial cognition model representing those neural structures.

1.2 Related Work

Taking inspiration from spatial cognition in rats, several computer-simulated or robot-tested navigation models have been proposed such as Burgess et al. [32], Brown and Sharp [33], Redish and Touretzky [34], Guazelli et al. [35], Arleo and Gerstner [36], Gaussier et al. [37], Filliat and Meyer [38], Arleo et al. [39], and Milford and Wyeth [40]. Main aspects of our approach include: (i) generation of the spatial map of the whole environment considering the ensemble place cell activity as a code for the current location, and determining topological (where to go next) and metrical (what motor actions perform to get there) relations between different location codes registered during exploration of all places in the environment, (ii) modeling of learning and extinction abilities in rats, defining extinction as the suppression of a behavior based on a reward that is changed in amount, quality or location [12], (iii) modeling of bidirectional interaction between brain functionalities of place representation and learning to exploit expectations of future reward during reinforced spatial tasks, (iv) map adaptation that permits on-line representations of changes in the physical configuration of the environment, and (v) validation of our robotic architecture beyond well-known spatial tasks performed with rats, but also by designing and implementing experiments that produce new behavioral data to be compared with similar experiments in robots. Refer to $\begin{bmatrix} 5 \end{bmatrix}$ for further detail on a comparative analysis between our model and the abovementioned approaches.

2 A Bio-inspired Robotic Model of Spatial Cognition

The spatial cognition model comprises distinct functional modules that capture some properties of rat brain structures involved in learning and memory. These modules are shown in Fig. 2 and described below in terms of an *animat* representing the



Fig. 2 The modules of the spatial cognition model and their interaction. Glossary: *LH* lateral hypothalamus, *RC* retrosplenial cortex, *EC* entorhinal cortex, *VTA* ventral tegmental area, *VS* ventral striatum, *NA* nucleus accumbens, *PIFDL* path integration feature detector layer, *LFDL* landmark feature detector layer, *LL* landmarks layer. Inputs/outputs: *r* primary reinforcement, *sr* secondary reinforcement, \hat{r} effective reinforcement, *DR* dynamic remapping perceptual schema, *WDR* connection weights between DR and PIFDL, *LPS* landmark perceptual schema, *APS* affordances perceptual schema, *PL* landmark information pattern, *LI*, *Ln* information pattern for landmark 1, landmark *n*, *LP* landmarks information pattern, *WI,Wn* connection weights between L1, Ln and LL, *PC* place information pattern, *EX* expectations of maximum reward and their corresponding directions (DX), *DIR* next animat direction, *ROT* animat rotation, *DIS* next animat moving displacement

modeled rat. A more detailed mathematical depiction of each module is presented in Barrera & Weitzenfeld [5].

2.1 Motivation

Motivation module relates to the rat's lateral hypothalamus (LH) controlling food seeking and food intake. The module computes the value of hunger drive (D) and produces the immediate or primary reward (r) the animat gets by the presence of food, corresponding to the navigational goal in the experiments. The reward r depends on the animat internal drive D experimented at any given time.

2.2 Kinesthetic Processing

Kinesthetic processing obtains information from two systems: (i) vestibular organs in the semicircular canals of the inner ear, and (ii) feedback from muscles controlling movement. Information derived from these systems is used by rats to carry out path integration processes. The posterior parietal cortex (PPC), a sensory structure receiving multimodal information such as kinesthetic, visual, and relative to affordances, has been suggested to mediate path integration [41] involving also the retrosplenial cortex (RC) [42].

In our model, we attribute to PPC the representation of the updated animat position in relation to its point of departure (anchor) by integrating past rotations and translations through a dynamic remapping perceptual schema (DR), and to RC, the generation of groups of neurons in the path integration feature detector layer (PIFDL) of the model that respond to specific kinesthetic information patterns (PI) due to the use of Hebbian learning [2]. PI is defined as a matrix of $1 \times n$ activation values registered by all n neurons in PIFDL at any given time t, as follows:

$$PI(t) = DR(t) W_{DR}(t), \tag{1}$$

where DR is a matrix of $1 \times m$ (with *m* depending on the dimension of the modeled environment) input signals to PIFDL, and W_{DR} is a matrix of $m \times n$ connection weights between neurons in DR and PIFDL.

2.3 Landmarks Processing

Landmark processing module computes landmark-related spatial positioning, i.e. distance and relative orientation of each landmark to the animat. According to Redish & Touretzky [34], the entorhinal cortex (EC) is involved in landmarks processing by receiving spatial information about landmarks from PPC. In our model, spatial information about each landmark is encoded in a landmark perceptual schema (*LPS1*, ..., *LPSn*) serving as input to a specific landmark feature detector layer (LFDL1, ..., LFDLn) that produces a landmark information pattern (*L1*, ..., *Ln*). Then, different LFDLs are combined into a single landmarks layer (LL). Hebbian learning updates connection weights between layers LFDLs and LL producing groups of neurons in LL that respond to specific landmark information patterns (*LP*) derived from the integration of all landmarks presented in the animat's environment. In this way, LP is defined as a matrix of $1 \times n$ activation values of all *n* neurons in LL representing an egocentric view from the animat at any given time *t* as follows:

$$LP(t) = L1(t) W1(t) + ... + Ln(t) Wn(t),$$
(2)

where L1, ..., Ln are matrices of $1 \times n$ input signals from layers LFDL1, ..., LFDLn to LL, and W1, ..., Wn are matrices of $n \times n$ connection weights between neurons in LFDL1, ..., LFDLn and LL.

2.4 Affordances Processing

It has been suggested that preceding the rat's motion, nearly half of the cells in PPC exhibit movement-related activity discriminating among basic modes of locomotion: left turns, right turns, and forward motion [43]. The affordances processing module represents PPC cell discrimination among different possible orientations for navigation by use of an affordances perceptual schema (*APS*) encoding possible egocentric (local) turns from -180° to $+180^{\circ}$ in 45° intervals at any given time from any given animat location and orientation.

2.5 Place Representation

Place representation module corresponds to the rat's hippocampus. As shown in Fig. 3a, this module receives input from kinesthetic (*PI*), landmark (*LP*) and affordances (*APS*) modules in addition to its interaction with the learning module. Place representation module comprises a place cell layer (PCL) and a world graph layer (WGL).

2.5.1 Place Cell Layer

Place cell layer (PCL) encompasses hippocampal regions CA3 and CA1, in addition to dentate gyrus (DG). Overlapping place fields in the collection of neurons in PCL are associated with a physical area in the environment that is identified directionally by the ensemble place cell activity pattern (*PC*), and whose extension is determined by affordances changes sensed by the animat during exploration. Specifically, neurons in the path integration feature detector layer (PIFDL) and in the landmarks layer (LL) of the model are connected to neurons in PCL. Synaptic efficacy between layers is maintained by Hebbian learning producing groups of neurons in PCL that respond to specific place cell information patterns (*PC*) derived from kinesthetic and egocentric visual information sensed by the animat while being at certain location and orientation. In this way, *PC* is defined as a matrix of $1 \times n$ activation values registered by the collection of *n* neurons in PCL at any given time *t* as follows:

$$PC(t) = PI(t) W_{PI}(t) + LP(t) W_{LP}(t),$$
(3)

where *PI* and *LP* are matrices of $1 \times n$ input signals from layers PIFDL and LL to PCL, W_{PI} and W_{LP} are matrices of $n \times n$ connection weights between neurons in PIFDL, LL and PCL.

2.5.2 World Graph Layer

Associations between overlapping place fields and physical areas are represented by world graph layer (WGL) through a topological map enabling navigation between



Fig. 3 a The place representation module of the spatial cognition model. Glossary: *PCL* place cell layer, *WGL* world graph layer, *PI* kinesthetic information pattern, *LP* landmarks information pattern, *W_{PI}* connection weights between PIFDL and PCL, W_{LP} connection weights between LL and PCL, W_{Nk} connection weights between any given map node k and its corresponding Actor units, *PC* place information pattern, *APS* affordances perceptual schema, \hat{r} effective reinforcement signal, *EX* expectations of maximum reward and their corresponding directions (DX). **b** Allocentric reference frame representing possible directions from 0° to 315° in 45° intervals to be adopted by the animat. This reference frame is relative to the animat departure location (*TD*) in the exploration process of a given maze

locations in the environment. Specifically, nodes in this map represent associations between kinesthetic and visual information patterns and the place cell population activity, whereas transitions between nodes are associated with metric information derived from animat locomotion such as its moving direction and displacement. Besides this mapping process, WGL also performs place recognition, and we assume that its functionality could be corresponded to the prelimbic cortex that is involved in the control of goal-directed behaviors [44].

Actor units in WGL store place cell activation patterns (*PC*) generated by PCL when the animat is oriented to diverse directions. These directions vary from 0° to 315° in 45° intervals, according to an allocentric (global) reference frame that is relative to the animat departure location in the exploration process as illustrated by Fig. 3b. Hence, every node in the map (a place) connects to a maximum of eight Actor units (eight possible orientations at each place). Every Actor connection is associated with a weight (representing the expectation of reward when orienting to a particular direction from the current location), and an eligibility trace (marking the connection eligible to be reinforced later in time). In this way, Actor units compete to select the next moving direction from the current location or node that allows the

animat to get the greatest reward, thus WGL analyzes Actor weights to obtain the biggest ones (EX) and their corresponding directions (DX).

2.5.2.1 *Place Recognition—Node Creation and Activation* To determine whether or not the animat recognizes a place, WGL searches the current activity pattern *PC* produced by PCL within all Actor units in the map. This search involves the computation of the similarity degree *SD* between *PC* and every stored place cell activity pattern. The model distinguishes among two cases:

- 1. If at least one *SD* exceeds certain threshold, the Actor unit storing the activation pattern with the biggest *SD* is considered the winner.
- 2. If there is no winner, WGL creates a new Actor unit storing pattern *PC* associated to the current animat orientation.

Then, WGL activates or creates a node in the map depending on the following considerations:

- (i) If affordances encoded by *APS* at time t are different from those at time t 1 and a new Actor unit was created, then a new node is created in WGL, connected with that Actor unit, and set as the new active node in the map.
- (ii) If affordances did not change and a new Actor unit was created, then WGL averages the activation pattern stored in the new Actor unit and the pattern stored in the Actor unit of the active map node that is associated to the current animat orientation.
- (iii) If there was an Actor unit winner, an arc from the active node to the node connected to that Actor unit is created if necessary, and this node becomes the new active one.

The flow diagram provided in Fig. 4 summarizes the node creation and activation algorithm, whereas Fig. 5 illustrates the creation process by using a fragment of the map for a cyclical maze, i.e. a maze that provides different possible paths from a given location to another one. Figure 5a shows the trajectory followed by the animat during maze exploration, whereas in Fig. 5b different locations belonging to that trajectory are labeled using letters. Considering first the route from location A and back to A (solid line in Fig. 5a), the animat starts at A oriented towards B corresponding to allocentric direction 90° (see Fig. 3b). As the map is empty, the model creates an Actor unit associated with direction 90° storing pattern PC derived from the current perception of kinesthetic and visual information. Then, the model creates node 1 as illustrated in Fig. 5c, and links it to the new Actor unit 90° (not shown in the figure). The animat chooses to move forward reaching location B, and creates node 2 with its corresponding Actor unit 90° since affordances encoded in the current APS are different from those registered at location A. Affordances remain constant at location C, thus pattern PC generated there is averaged with the pattern stored in the Actor unit 90° of node 2. At location D, affordances change, hence the model creates a new node 3 connected to a new Actor unit 90°. Progressively, the exploration of locations E, F and G in direction 90° leads to the creation of nodes 4 and 5. Then, when the animat turns left at location G orienting itself to 180°, a new Actor unit 180° is created to store the current PC and connected to node 5. Being oriented to 180° , the model creates nodes 6 and 7. At location J, the animat turns left orienting itself



Fig. 4 Flow diagram representing the procedure of creation and activation of nodes within a cognitive map. Glossary: *PC* current place information pattern, *APScurr* affordances perceptual schema encoded at time t, *APSant* affordances perceptual schema encoded at time t-1. Refer to the text for a complete description of the procedure and the corresponding glossary



Fig. 5 The processes of creation and adaptation of a cognitive map. **a** The exploration trajectory followed by the animat in a cyclical maze. The maze is surrounded by four landmarks (L1, L2, L3, L4). **b** Within the maze, different visited locations are labeled with letters from A to O in order of exploration. **c** Map creation while exploring the route indicated as *solid line* in (**a**). **d** Map adaptation while exploring the route indicated as *dotted line* in (**a**). Map nodes are numbered in order of creation, and arcs connecting nodes are labeled with the animat allocentric direction when it moved from one node to the next one

to 270° , thus a new Actor unit 270° is created to store the current *PC* and connected to node 7. Nodes 8 and 9 are created while exploring the corridor in direction 270° . At location M, the animat turns left orienting itself to 0° , thus a new Actor unit 0° is created to store the current *PC* and connected to node 9. Nodes 10 and 11 are created while exploring the corridor in direction 0° . Even though nodes 11 and 3 are representing the same location D in the maze, the animat could not recognize node 3 when visiting D in direction 0° since it was previously explored only in direction 90° and node 3 is only linked to Actor unit 90° . Being oriented to 0° at D, the animat turns right in direction 270° creating a new Actor unit 270° and linking to node 11. Then, node 12 is created to represent locations C and B in direction 270° . When the animat reaches A being oriented to 270° does not recognize node 1 since it includes only the Actor unit 90° , thus the model creates node 13 connecting it to a new Actor unit 270° .

2.5.2.2 Map Adaptation During exploration, an existing topological map may be adapted due to place recognition by the animat. Figure 5d illustrates the adaptation process of the map previously introduced in Fig. 5c. Continuing with the maze exploration process now from locations A to D (dotted line in Fig. 5a), the animat oriented to 270° at A decides to turn left creating a new Actor unit 0° and linking it to node 13. Then the animat chooses to turn left again orienting itself to 90° . At this time, the activity pattern stored in Actor unit 90° of node 1 presents a SD value that exceeds the threshold, hence node 1 is activated. As a result, node 13 is integrated to node 1 since they are representing the same place in the maze. This integration involves connecting Actor units 270° and 0° of node 13 to node 1, linking node 12 to node 1 through a new arc in direction 270°, eliminating both the existing arc from node 12 to node 13 and node 13, and activating node 1 as shown in Fig. 5d. Then, the animat moves forward towards B visiting also C and D and recognizing nodes 2 and 3. Being at D, the animat turns right pointing in direction 0°. At this time, the activity pattern stored in Actor unit 0° of node 11 presents a SD value that exceeds the threshold, thus node 11 is activated. Consequently, existing nodes 3 and 11 are merged into a new one since they are representing the same place in the maze. Therefore, a new node 13 is created integrating Actor units 0° and 270° of node 11 and Actor unit 90° of node 3, as well as all input/output arcs to/from nodes 11 and 3 (i.e., arc 0° from node 10, arc 90° from node 2, arc 270° to node 12, and arc 90° to node 4). Merged nodes 11 and 3 are removed from the map, and the new node 13 becomes the active one (see Fig. 5d).

2.6 Learning

Learning module is related to dopaminergic neurons in the ventral tegmental area and to ventral striatum, processing reward information by using an Actor-Critic architecture. Prior use of this reinforcement learning method to solve goal-search tasks has been documented. The work by Foster, Morris and Dayan [45], for instance, implements an Actor-Critic architecture to enable a simulated rat to solve the reference memory task in a water maze providing the rat with a reward at any given time of the experiment.

In our model, as illustrated in Fig. 6, the Adaptive Critic (AC) includes a Prediction Unit (PU) that estimates the future reward value of any particular place or location at a given time. To do this, every neuron in PCL is connected to PU, and these connections are associated with weights W and eligibility traces E. At each time step t in a trial of an experiment, PU computes the future value P of the activity pattern PC generated by PCL according to (4):

$$P(t) = PC(t) \cdot W_{PCL}(t), \tag{4}$$

where *PC* is a matrix of $1 \times n$ input signals from PCL to PU, and W_{PCL} is a matrix of $1 \times n$ connection weights between neurons in PCL and PU.



AC uses predictions computed at times t and t - 1 to determine the secondary reinforcement, discounting the current prediction at a rate γ to get its present value. The addition of the secondary reinforcement with the primary reinforcement r computed by the motivation module of the model constitutes the effective reinforcement \hat{r} as described by (5):

$$\hat{r}(t) = r(t) + \gamma P(t) - P(t-1).$$
(5)

The effective reinforcement is employed to update the connection weights between PCL and PU in AC (i.e., the reward expectation associated to a place), and also between Actor units and map nodes (i.e., reward expectations associated to different orientations). In the first case we used

$$W_{PCL}(t+1) = W_{PCL}(t) + \beta \,\hat{r}(t) E_{PCL}(t), \tag{6}$$

where β is the learning rate, and E_{PCL} is the matrix of $1 \times n$ eligibility traces corresponding to connections between PCL and PU in AC. In the second case we used

$$W_{Nk}(t+1) = W_{Nk}(t) + \beta \,\hat{r}(t) E_{Nk}(t) \quad \forall \text{ map node } k, \tag{7}$$

🖉 Springer

where W_{Nk} is the vector of connection weights between map node k and a maximum of eight Actor units, and E_{Nk} is the vector of eligibility traces corresponding to a maximum of eight Actor units.

As shown in (6) and (7), both learning rules depend on the eligibility of the connections. At the beginning of any trial in a given experiment, eligibility traces in AC and in Actor units are initialized to 0. At each time step t in a trial, eligibility traces in AC are increased in the connections between PU and the most active neurons within PCL when the action executed by the animat at time t - 1 allowed it to perceive the goal, as shown in (8):

$$E_{PCL}(t) = E_{PCL}(t-1) + \chi PC(t),$$
 (8)

where χ is the increment parameter, and *PC* stores the activity pattern registered by the collection of neurons in PCL. Also at time step *t*, the eligibility trace *e* of the connection between the active node *na* in the map and the Actor unit corresponding to the current animat orientation *dir* is increased by τ as described by (9):

$$e_{na}^{dir}(t) = e_{na}^{dir}(t-1) + \tau.$$
 (9)

Finally, after updating connection weights between PCL and AC, and between Actor units and map nodes at any time step *t* in the trial, all eligibilities decay at certain rates λ and σ respectively, as shown in (10):

$$E_{PCL}(t) = \lambda E_{PCL}(t-1)$$

$$E_{Nk}(t) = \sigma E_{Nk}(t-1) \quad \forall map \ node \ k.$$
(10)

Additionally, to enable goal-directed navigation, the learning module implements a backward reinforcement method where eligibility traces of Actor units are updated in the direction of the arcs connecting the nodes in the path followed by the animat. The backward reinforcement process is carried out after having concluded every successful training trial in the given experiment, and the strategy involved in this method is based on the existence of a factor referred to as goal gradient by Hull [46], according to which the reinforcement effect is the most at the goal location and diminishes progressively as the animal moves backward through the maze. Specifically, in case the animat finds the goal at the end of the path, each eligibility trace is updated in a given amount of reinforcement divided by the amount of steps the animat performed to move from one node to the next one in the path. The reinforcement is initialized to a certain amount at the beginning of any training trial in the experiment, and this amount decreases as the distance from a node to the goal increases.

2.7 Action Selection

Action selection module computes the motor outputs of the model consisting on the next animat direction (DIR), the required rotation to point to that direction (ROT), and the moving displacement (DIS).

Motion is determined by considering (i) all possible affordances or rotations to execute from current location and orientation (APS), (ii) selection of a random rotation between possible affordances (*RPS*, internally computed), (iii) curiosity to execute rotations not yet explored (*CPS*, internally computed), and (iv) expectation

of maximum reward (*EMR*, internally computed by using *EX* and *DX*). These four "signals" are computed by using one or more Gaussian functions whose values are stored in linear arrays or vectors. Specific positions distributed throughout a vector correspond to particular relative rotations between -180° and $+180^{\circ}$ in 45° intervals, and the biggest value of a Gaussian function (the height of the curve's peak) is stored in one of these positions.

The influence of each signal in the final action selection depends on the biggest possible value of the Gaussian functions representing it. Varying the parameter that regulates the height of the curve's peak, the model assigns the following priority to signals: (i) *EMR*, (ii) *APS*, (iii) *CPS*, and (iv) *RPS*.

Vectors derived from the representation of those signals are added, and the specific relative rotation associated with the position of the resultant vector storing the biggest value is used to determine the next animat direction from 0° to 315° in 45° intervals.

In the course of an experiment, while the *EMR* signal is weak, the animat executes a rotation not yet experimented at its current location or a random rotation in case it had tried all possible rotations earlier. Regarding this last situation, it is feasible for the animat to show a "hesitation" behavior, in the sense of performing two or more body rotations at its current location before restarting navigation in the given environment.

3 Animal Experimentation

We used a group of three male rats to carry out the behavioral experiment with a 95×105 cm² maze having six internal corridors as shown in Fig. 7. The maze was placed in a room illuminated using a tenuous light, and surrounded by a white curtain with four different colored geometrical figures representing salient and stable allocentric cues or landmarks.

In order to motivate the animals to learn the spatial task, the consumption of water was restricted during the previous week of the experiment, and absolutely prohibited during the prior 24 h. The tip of a water dispenser was placed at the goal location indicated in Fig. 7 to provide sweet water to rats when reaching this location at the end of any training trial. The solution used as reward consisted of 5 gr. of sucrose dissolved in 50 ml. of water, whereas the water dispenser was filled with 400 μ l of the solution before beginning each trial.

Fig. 7 Top view of the cyclical maze employed in the experiment with rats. Landmarks *L1, L2, L3* and *L4* are illustrated, as well as locations *TD, D1, D2, D3, D4*, and the goal location



During the experiment, rats pass through three phases: habituation, training, and behavioral testing, as described in the following subsections.

3.1 Habituation Procedure

The task begins with a habituation session (pre-training), where the animal is placed in the maze and allowed to explore it freely from location TD (see Fig. 7) being oriented north (90° according to Fig. 3b) until reaching the water dispenser. Water is not provided during the habituation session.

3.2 Training Procedure

During training, the water dispenser provides sweet water as reward to rats. At the beginning of each training trial the rat is placed at location TD (Fig. 7) oriented north (90°). Each trial consists on freely exploring the maze until reaching the goal location and drinking the sweet water. The rat is then removed from the maze and placed in a cage for 60 s after which a new trial is performed. Each rat daily training



Fig. 8 Routes followed by three rats (502, 503, 504) in locating the water (*Goal*) during three training sessions (1, 7 and 17) from a fixed departure location (TD) in the maze. Lines inside each maze illustrate nine different routes. The average latency of arrival to the target is shown below every maze



Fig. 9 a Average latency of arrival to the goal location per training session considering the performance of each rat (502, 503, 504) during the spatial task. **b** Average latency of arrival to the goal location per training session considering the average performance of the three rats and indicating the corresponding variability

session included 9 trials, with each trial ending with the rat reaching the target for a total of 17 sessions. Latencies and routes followed by the animals in locating the water were registered by using the SMART system (San Diego Instruments). Figure 8 illustrates routes followed by different rats during three training sessions, with sessions selected from beginning (Session 1), middle (Session 7), and end (Session 17) of the procedure. It is notorious how the latency of routes decreases as training evolves until animals learn the shortest path to the goal.

Figure 9a shows the average latency of arrival per session for the three rats, whereas Fig. 9b illustrates the average latency of arrival per session considering the average performance of the three rats and indicating the corresponding variability.

3.3 Behavioral Testing Procedure

In order to evaluate rats' ability to reach the target after having concluded the training procedure, rats were separately tested during 12 trials starting three times from four different locations in the maze, D1, D2, D3, and D4 (see Fig. 7), pointing

in four different directions, north (90°), north (90°), east (0°), and south (270°), respectively (see Fig. 3b).

During each testing trial, the rat is placed at the corresponding initial location and orientation, and the trial consists on exploring the maze freely until reaching the goal location and drinking the sweet water. The rat is then removed from the maze. During tests, latencies and routes followed by the animals in locating the water were also registered by using the SMART system as depicted by Fig. 10, where it can be



Fig. 10 Routes followed by rats (502, 503, 504) in locating the water (*Goal*) during tests departing from locations D1 to north (90°), D2 to north (90°), D3 to east (0°), and D4 to south (270°) in the maze. *Lines* inside each maze illustrate three different routes. The average latency of arrival to the goal is shown below each maze

observed that rats did not always show an optimal performance following the shortest route.

3.4 Monitoring Neural Activity by Detecting the Expression of the Immediate Early Gene Arc

After the animals were trained and tested in the cyclical maze task, a sacrifice session was scheduled for five rats, three of which were the animals trained in the task and the other two were cage control animals with the same age.



Fig. 11 Representative full coronal montage images obtained with the Mosaic X software in the Apotome system using a $10 \times /0.45$ objective. Nuclear Sytox counterstaining is shown in *green*, and arc expressing cells are shown in *red*. The images were obtained from cage control animals (**a** and **c**), and from animals previously trained and exposed to the cyclical maze (**b** and **d**). *Top images* (**a** and **b**) correspond to an anteroposterior position around -4.0 mm from bregma where the hippocampus is located, and bottom images (**c** and **d**) correspond to a position around 0.2 mm from bregma where the striatum is located (dorsal, lateral, ventral and medial striatum). The *white bar* represents 1 mm

Trained animals underwent a final behavioral session consisting in placing them at the same release location used during training for a total of 7 trials. This was done in order to complete 5 min of behavioral exploration, which is used to guarantee the maximal amount of behaviorally activated neurons in the hippocampus using other spatial learning tasks [29, 30].

Thirty minutes after the behavioral exposure, animals were sacrificed by decapitation and their brain was removed and processed for Arc immunohistochemistry as described in [30]. It is important to mention that all animal brains were placed in one single block and sliced together in 20 μ m sections, thus each slide contained two caged animals brain sections and three trained animals brain sections.

Immunostained sections (3 or 4) were imaged using a \times 10/0.45 plan apochromat objective and the Mosaic X imaging software on the Apotome system (Zeiss ® Standort Göttingen-Vertrieb Deutschland) to obtain whole coronal section images containing the dorsal hippocampus and the striatum as shown in Fig. 11.

Fig. 12 Zoom images from the Mosaic X montage obtained with a $\times 10/0.45$ objective. Nuclear Sytox counterstaining is shown in green, and arc expressing cells are shown in red. Images a and c correspond to the cage control animal, whereas images b and d correspond to one of the animals previously trained and exposed to the cyclical maze. Images **a** and **b** show activated cells in the hippocampus, whereas c and d show activated cells in the ventral striatum. The white bar represents 250 µm



The imaging parameters were set using the staining signals observed in one of the caged animals and maintained through the rest of the imaging procedure for each slide. The collected images were used to measure the area occupied by the Arc staining signal (in red) and corrected by the region determined by the Sytox green (Invitrogen ® Carlsbad, California) counterstained signal (in green) using the metamorph imaging software (Molecular Devices ® Downingtown, PA). In brief, a region of interest (ROI) was drawn using the area occupied by the counterstained signal that stains the cell nuclei of all cells and allows us to identify the different anatomical regions of the hippocampus and striatum. Once the ROI was drawn, a threshold was set in a cage animal images from the other animals (one more cage and the three exploration animals). This measure resulted in a gross quantification of the animals that executed the task and the corresponding cage control subjects.

Results revealed a clear activation of hippocampus and striatum that can be observed in the images shown in Fig. 12, as well as by the threshold quantification shown in Fig. 13. Therefore, rats engaged both hippocampus and striatum in information processing and used them to solve the task.





4 Robot Experimentation

The rat cognitive model was designed and implemented using the NSL system [47]. The computational model interacts with a real robotic environment through an external visual processing module that takes as input three non-overlapping snapshots $(0^{\circ}, +90^{\circ}, -90)$ taken by the robot at each step using its local camera, and a motor control module that executes rotations and translations on the robot. Refer to [5] for further detail on the robotic implementation of the model.

The AIBO robot experiment was done in a $180 \times 210 \text{ cm}^2$ maze having six internal corridors as shown in Fig. 14. The maze is surrounded by four colored cylinders representing salient and stable allocentric cues or landmarks. During the experiment, similarly to rats, the robot passes through three phases: habituation, training, and testing, as described in the following subsections.

4.1 Habituation Procedure

As with rats, the experiment begins with a habituation procedure where the robot departs from location TD (Fig. 14) oriented north (90° according to Fig. 3b). The robot explores the maze freely until reaching location HE corresponding to the goal. No reinforcement is provided to the robot while exploring the maze, thus it does not carry out any learning process of goal locations, and its actions are determined just by curiosity and randomness. As a result of the habituation procedure, the robot builds a cognitive map of the maze including nodes to represent the explored locations, and Actor units associated with different directions in which they were explored. Figure 15 illustrates spatial maps generated during habituation by four robots.

Fig. 14 Top view of the cyclical maze used in the experiment with AIBO robots. Landmarks *L1, L2, L3* and *L4* are illustrated, as well as locations *TD, D1, D2, D3, D4*, and the goal location





Fig. 15 Spatial maps generated by four robots during the habituation procedure of the spatial task. Nodes are numbered in order of creation, and arcs between nodes are labeled with the robot allocentric direction (see Fig. 3b) when it moved from one node to the next one

4.2 Training Procedure

After habituation, reinforcement is introduced at location HE of the maze, the robot is placed at TD oriented north (90°) , and the training procedure begins from the cognitive map previously generated.

The robot perceives the same visual and kinesthetic information sensed during habituation since the environmental configuration remains constant. The robot recognizes during training previously explored locations that were navigated in similar directions. On the other hand, when the robot visits unexplored locations or orientations, it adds new nodes and connections to the map. Additionally, when experimenting with different orientations at any given location, the robot tries to merge possible redundant nodes by creating new ones and reconstructing links as necessary (as described above in 'Section 2.5.2.2'). Figure 16 illustrates the spatial maps presented in Fig. 15 including all modifications applied by robots during training.

In a typical training trial, the robot explores the maze from TD until finding the goal location. Exploration is interrupted if the robot spends more than 600 s without reaching the goal. Depending on how exhaustively the robot explores the



Fig. 16 Spatial maps adapted by four robots as a result of the training process in the spatial task. *White nodes* were created during the habituation procedure, whereas *black nodes* were added by robots during training. Nodes are numbered in order of creation, and arcs between nodes are labeled with the robot allocentric direction (see Fig. 3b) when it moved from one node to the next one. All missed numbers within the sequence of nodes in each map correspond to nodes that were merged with others when robots identified them as duplicates representing one same place in the environment

maze during habituation, its motor actions during training are more determined by the random factor and less determined by the curiosity factor since it recognizes all visited maze locations. While the robot's reward expectations are not big enough to exceed the randomness, it follows indirect routes to the goal (i.e., routes that involve visiting places more than once), or direct but not optimal routes.

Training phase consisted of one session including 20 trials per robot. As with rats, latencies and routes followed by robots to reach the goal were recorded. Figure 17 illustrates routes registered during three training trials selected from beginning (Trial 1), middle (Trial 8), and end (Trial 20) of the procedure. It can be observed how the learning latency decreases until reaching the goal location by following the optimal path.



Fig. 17 Routes followed by four robots in locating the goal during three training trials from a fixed departure location (TD) being oriented north (90°) in the maze. Illustrated trials were taken from beginning (*Trial 1*), middle (*Trial 8*), and end (*Trial 20*) of the learning process. The line inside each maze represents the route followed by the robot. The *small arrow* placed at location TD indicates the robot orientation, and all other *small arrows* over the line represent locations where the robot hesitated while taking its next motor action. The average latency of arrival to the target in those three training trials is shown below each maze

Figure 18a shows the average latency of arrival to the goal for each of the four robots, whereas Fig. 18b depicts the average latency of arrival for the four robots indicating the corresponding variability.





Fig. 18 a Average latency of arrival to the goal per training trial considering the performance of each of the four robots during the spatial task. **b** Average latency of arrival per training trial considering the average performance of the four robots and indicating the corresponding variability

4.3 Testing Procedure

In order to evaluate robots' ability to reach the target after finishing the training procedure, we tested robot trajectories during 12 trials with robots liberated three times from the same initial locations and orientations as used to test rats. Figure 14 illustrates locations D1, D2, D3 and D4, from where robots depart with orientation north (90°), north (90°), east (0°) and south (270°), respectively.

During each testing trial, the robot is placed at the initial location and orientation, and the trial consists on exploring the maze freely until reaching the goal location. After reaching the goal, the robot is removed from the maze.

During tests, latencies and routes followed by robots to reach the target were recorded as shown by Fig. 19. Each robot exploits the cognitive map built during habituation and adapted during training to reach the target successfully. During some



Fig. 19 Routes followed by four robots while locating the goal during tests departing from locations D1 to north (90°), D2 to north (90°), D3 to east (0°), and D4 to south (270°) within the maze. Lines inside each maze illustrate routes recorded from three trials. When the robot followed the same route in those three trails, the corresponding maze shows only one line. *Small arrows* over the routes represent hesitations executed by the robot in deciding its next motor action. The average latency of arrival to the goal location in three trials per departure point is shown below each maze

of the testing trials, robots still modified their spatial maps by adding new nodes to represent unexplored locations and/or merging existent nodes. Figure 20 illustrates nodes recognized by robots within their spatial maps during tests, and new nodes derived from additions or integrations.

5 Comparative Results

As a result of training, the latency of arrival to the target location decreases progressively until stabilizing in both rats and robots as shown in Figs. 9 and 18.



Fig. 20 Spatial maps adapted by four robots as a result of tests carried out in the cyclical maze task. *Black nodes* indicate existing nodes recognized by robots, or new nodes created by them as a result of visiting unexplored places or integrating duplicate nodes that correspond to same places. Information related to maximum reward expectations stored in Actor units of recognized nodes was exploited by robots to reach the goal location successfully regardless the point of departure within the maze. *White nodes* are existing nodes not used by robots during tests. Nodes are numbered in order of creation, and arcs between nodes are labeled with the robot allocentric direction (see Fig. 3b) when it moved from one node to the next one

Hence, both subjects were able to learn the shortest route leading to the goal from the initial fixed location. To confirm that the learning criterion was reached, we obtained for both rats and robots the path length registered during each training trial, normalized to the shortest path length possible in the maze from the starting point to the goal location. Figure 21a presents the evolution of the rat average learning through the training process in terms of the normalized route length. It is feasible to observe a similar curve as the one related to the average latency of arrival to the goal (Fig. 9b). Likewise, Fig. 21b depicts the evolution of the robot average learning during training in terms of the normalized route length, where we can observe a very similar curve as that for the latency of arrival to the goal (Fig. 18b). For both, rats and robots, the normalized path length shows a decreasing tendency until it stops varying as a result of having learnt the optimal route from the fixed initial location to the goal. Specifically, in training session 1, rats traveled approximately 2.5 times the optimal distance to the goal, whereas in training session 16, they traveled approximately 1



Fig. 21 Average performance during training in the task measured through the distance traveled normalized to the minimal distance to the goal; i.e., the average number of times that **a** rats and **b** robots traveled the minimal distance possible from the initial fixed location to the goal. Both graphs show the variability of the results

time the optimal distance to the goal. Likewise, robots traveled approximately 4.5 times the optimal distance to the goal in training trial 1, whereas in training trial 20, they traveled 1 time the optimal distance as expected.

In Fig. 22 we include a comparison between number of hesitations, i.e., movement stops followed by more than one head or body turn before restarting navigation, monitored during training for rats and robots. It can be observed that both subjects reported a decreasing number of hesitations. Rats showed less than one hesitation during the last training session, whereas the robots' average performance indicates that they stop hesitating in their navigation decisions once they have learnt the shortest path to the goal, i.e., around trial 13.

It is possible to characterize comparatively the performance of rats and robots during tests by analyzing behavioral procedures such as errors associated with the selection of a non-optimal route, body rotations, normalized length of the traveled path, and hesitations during navigation. Figure 23 illustrates the comparison of these behavioral procedures in a graphical manner.



Fig. 22 Average number of hesitations (i.e., movement stops followed by more than one body turn before restarting navigation) committed by **a** rats and **b** robots during training from the same fixed location in the maze

Errors are calculated as the percentage of executed testing trials where the subject did not follow the shortest route from the given departure location to the goal. As depicted by Fig. 23a, robots, in contrast to rats, followed optimal routes to the goal during all trials from locations D1 and D2. When starting from location D3, 50% of the trials performed by robots correspond to non-optimal paths to the target, whereas 67% of the trials performed by the rats correspond to non-optimal paths, and when departing from D4, 67% of the robot trials and 89% of the rat trials were incorrect.

Regarding tests with robots departing from D3 and D4, they reported some trials performing direct routes (i.e., non-optimal paths that do not involve visiting places of the maze more than once) or indirect routes (i.e., non-optimal paths that do involve visiting places of the maze more than once), since those locations were unexplored or barely explored during training as shown in the spatial maps of Fig. 16. In those cases, robots navigated randomly until recognizing a place previously learned, typically belonging to the shortest path to the goal, and followed the rest of that route.

On the other hand, as it was not informative to compare latencies of arrival to the goal between rats and robots since robots spent around 8 s at each step in



0.20 0.00

D1

D2 D3 Departure location during tests

D4

taking pictures of the environment, deciding their next motor action and executing it, we monitored instead the number of body rotations performed independently by both subjects, and the path length registered during each testing trial normalized to the shortest path length possible from the corresponding departure location to the target (i.e., the number of times the optimal distance). As shown in Fig. 23b, the average number of body rotations in rats when departing from any given test location exceeds slightly the number of robot body rotations, partially indicating that rats took longer to reach the target in contrast to robots. In fact, according to results depicted in Fig. 23c, rats traveled distances from 0.5 to 1.7 times larger than the minimal one possible from any given test departure location to the target; whereas robots navigated the minimal distance departing from location D2, distances slightly exceeding the minimal one possible from locations D1 and D3, and a 0.7 times larger distance starting from D4.

Finally, even though rats and robots had few hesitations while deciding their motor actions during tests, rats hesitated more than robots did during navigation to the goal location from any point of departure according to the results presented in Fig. 23d. Particularly, in contrast to rats, robots did not hesitate when departing from location D2 since this place belongs to the optimal route learnt during training.

6 Conclusions and Discussion

According to [48], robots can be "biological models" in several different senses: (i) they can be modeled *on* animals by using the biology as a source of ideas when attempting to build a robot with some target capability; (ii) they can be models *for* animals by employing robotic theory/technology as a source of explanatory mechanisms in biology; or (iii) they can be models *of* animals by using robots as a simulation technology to test biological hypotheses.

The work presented in this paper exemplifies this last kind of "biorobot". Indeed, our aim is to provide the robotic model of spatial cognition in rats to neurobiologists/neuroethologists as a technological platform to test with robots biological experiments whose results can predict rodents' spatial behavior. In this way, experimentalists could test neuroscientific hypothesis in the robotic model and obtain results in less than 2 h instead of spending several weeks experimenting directly with animals.

Webb [48] describes a number of dimensions on which robotic models of biological systems can differ, such as (i) performance match; i.e. to what extent the model behavior matches the target behavior; (ii) abstraction, i.e. the amount of detail included in the model; and (iii) structural accuracy, i.e. how well the model represents the actual mechanisms underlying the behavior.

Both, the demonstration that rats, as robots, used the hippocampal and striatal modules during the execution of the cyclical maze task discussed throughout the paper, and the results derived from the corresponding comparative behavioral analysis, allow us to validate our robotic model in terms of performance match with rats. Indeed, rats and robots reached successfully the learning criterion during the training process of the spatial task. When rats were tested to find the goal location in the relatively complex cyclical maze departing from different locations, they did not always show an optimal performance following the shortest route, and this behavior was also exhibited by robots during tests.

Allocentric information derived from global cues or landmarks was used by robots firstly to build the cognitive map and then to recognize within it maze locations previously explored. This is the function of the hippocampus module of our model, and biological studies had shown that the rat hippocampus performs such kind of information processing task [18, 19], which is congruent with our observation of hippocampal activity during the execution of the cyclical maze task. The striatum module, on the other hand, allows robots to learn routes that lead to goals by reinforcing motor actions executed at every place belonging to those routes, which are represented by chains of nodes in the spatial map. In this regard, we observed that during the execution of the task, rats increased the neuronal activity in the striatum suggesting that they may also perform a similar computation.

In locating the goal from any given departure location, robots navigated randomly until recognizing a place belonging to a route previously learnt, and exploited properly the information stored in Actor units to reach the goal location by executing rotations associated with maximum reward expectations. In this way, an efficient performance during tests depends on how exhaustively robots explore the environment during training visiting all possible direct routes from the fixed departure location to the goal.

Even though we have demonstrated the behavioral performance match between robots and rats during the execution of the cyclical maze task, we should acknowledge the following obvious differences between both subjects: (i) robots learned faster than rats; (ii) robots reported less incorrect trials during tests; (iii) robots executed less number of body rotations during any testing trial; (iv) robots traveled less number of times the minimal distance to the goal during any testing trial; and (v) robots hesitated less than rats during any testing trial. Exploring what contributes to this enhanced performance in robots may yield insights as to how learning can be improved in rats.

As mentioned in the beginning of this paper, our research group on biology and robotics has been studying the brain areas involved in spatial memory and learning in rodents, especially the hippocampus and the striatum, and we have modeled the underlying mechanisms of these structures through different functional modules that incorporate neuroscientific hypotheses. Even though our robotic model has reached a good level of abstraction and structural accuracy, it remains to demonstrate that the hippocampal and striatal activation in the rat brain is related with the same information processing modeled in the robot. Lesion or inactivation studies in different behavioral conditions may show some light in this regard. Additionally, the model needs to consider the following aspects:

- The function of head-direction cells [49] providing information on the rats' rotation magnitude and movement direction. Extending our affordances processing module by modeling the function of these cells and supplying their activation as input to the place representation module, will allow the robot to navigate in different spatial contexts (i.e., mazes with corridors as well as open field environments).
- The manner in which receptive fields of hippocampal place cells adapt their distribution by increasing their density in the presence of relevant elements

in a given navigational task (e.g., walls, corners, or doors), obstacles, and the perception of goals.

• The functional differences between hippocampal substructures CA1, CA3 and DG relative to their capabilities of pattern completion and pattern separation [50–52] in conditions where environmental changes promote new navigational behaviors.

We plan to extend the model in those directions to provide the robot with enhanced adaptive capabilities, which will be demonstrate through the comparative implementation of several reinforced navigational tasks with rats and robots to evaluate (i) adaptation to internal changes in the maze such as closing of existing corridors or opening of new ones, as well as (ii) navigational routes and the latency of arrival to the target location in an open field arena where an obstacle is introduced after the training process.

Certainly, our "biorobot" model, as discussed by Webb [48], offers an alternative experimental platform to address specific biological questions in an attempt to provide a finer understanding about spatial memory and learning in rodents.

References

- 1. Roberts, W.A.: Principles of Animal Cognition, pp. 201–230. McGraw Hill, USA (1998)
- Hebb, D.O.: The Organization of Behavior: A Neuropsychological Theory. Wiley-Interscience, New York (1949)
- Barto, A.G.: Adaptive critics and the basal ganglia. In: Houk, J.C., Davis, J.L., Beiser, D. (eds.) Models of Information Processing in the Basal Ganglia, pp. 215–232. MIT Press, Cambridge (1995)
- O'Keefe, J.: Spatial memory within and without the hippocampal system. In: Seifert, W. (ed.) Neurobiology of the Hippocampus, pp. 375–403. Academic Press, New York (1983)
- Barrera, A., Weitzenfeld, A.: Biologically-inspired robot spatial cognition based on rat neurophysiological studies. Auton. Robots, Springer 25(1–2), 147–169 (2008)
- 6. Morris, R.G.M.: Spatial localization does not require the presence of local cues. Learn. Motiv. **12**, 239–260 (1981)
- Barrera, A., Weitzenfeld, A.: Rat-inspired model of robot target learning and place recognition. In: Proceedings of the 15th Mediterranean Conference on Control and Automation—MED. Athens, Greece (2007)
- 8. Arkin, R.C.: Behavioral Based Robotics. MIT Press (1998)
- 9. Webb, B.: What does robotics offer animal behaviour? Anim. Behav. 60, 545–558 (2000)
- Weitzenfeld, A.: A prey catching and predator avoidance neural-schema architecture for single and multiple robots. J. Intell. Robot. Syst., Springer 51(2), 203–233 (2008)
- Weitzenfeld, A.: From schemas to neural networks: a multi-level modeling approach to biologically-inspired autonomous robotic systems. J. Robot. Auton. Syst. 56(2), 177–197 (2008)
- 12. O'Keefe, J., Nadel, L.: The Hippocampus as a Cognitive Map. Oxford University Press (1978)
- O'Keefe, J., Dostrovsky, J.: The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. Brain Res. 34(1), 171–175 (1971)
- 14. Tolman, E.: Cognitive maps in rats and men. Psychol. Rev. 55, 189-208 (1948)
- McNaughton, B.L., Knierim, J.J., Wilson, M.A.: Vector encoding and the vestibular foundations of spatial cognition. In: Gazzaniga, M. (ed.) The Cognitive Neurosciences, pp. 585–595. MIT Press, Boston (1994)
- O'Keefe, J., Conway, D.H.: Hippocampal place units in the freely moving rat: why they fire where they fire. Exp. Brain Res. 31, 573–590 (1978)
- 17. Quirk, G.J., Muller, R.U., Kubie, J.L.: The firing of hippocampal place cells in the dark depends on the rat's recent experience. J. Neurosci. **10**(6), 2008–2017 (1990)
- Jeffery, K.J., O'Keefe, J.M.: Learned interaction of visual and idiothetic cues in the control of place field orientation. Exp. Brain Res. 127, 151–161 (1999)

- 19. Poucet, B.: Spatial cognitive maps in animals: new hypotheses on their structure and neural mechanisms. Psychol. Rev. **100**(2), 163–182 (1993)
- Moser, E.I., Kropff, E., Moser, M.-B.: Place cells, grid cells, and the brain's spatial representation system. Annu. Rev. Neurosci. 31, 69–89 (2008)
- Risold, P., Thompson, R., Swanson, L.: The structural organization of connections between hypothalamus and cerebral cortex. Brains Res. Rev. 24(2–3), 197–254 (1997)
- 22. Kelley, A.: Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. Neurosci. Biobehav. Rev. 27(8), 765–776 (2004)
- Schultz, W., Tremblay, L., Hollerman, J.: Reward prediction in primate basal ganglia and frontal cortex. Neuropharmacology 37(4–5), 421–429 (1998)
- Schultz, W., Tremblay, L., Hollerman, J.: Reward processing in primate orbitofrontal cortex and basal ganglia. Cereb. Cortex 10(3), 272–283 (2000)
- 25. Houk, J.C., Adams, J.L., Barto, A.G.: A model of how the basal ganglia generate and use neural signals that predict reinforcement. In: Houk, J.C., Davis, J.L., Beiser, D.G. (eds.) Models of Information Processing in the Basal Ganglia, pp. 249–270. MIT Press, Cambridge (1995)
- Collett, T.S., Graham, P.: Animal navigation: path integration, visual landmarks and cognitive maps. Curr. Biol. 14(12), R475–7 (2004)
- Hartley, T., Burgess, N.: Complementary memory systems: competition, cooperation and compensation. Trends Neurosci. 28(4), 169–170 (2005)
- Lanahan, A., Worley, P.: Immediate-early genes and synaptic function. Neurobiol. Learn. Mem. 70(1–2), 37–43 (1998)
- Guzowski, J.F., McNaughton, B.L., Barnes, C.A., Worley, P.F.: Environment-specific expression of the immediate-early gene Arc in hippocampal neuronal ensembles. Nat. Neurosci. 2(12), 1120–1124 (1999)
- Ramirez-Amaya, V., Vazdarjanova, A., Mikhael, D., Rosi, S., Worley, P.F., Barnes, C.A.: Spatial exploration-induced arc mRNA and protein expression: evidence for selective, network-specific reactivation. J. Neurosci. 25, 1761–1768 (2005)
- Vazdarjanova, A., Ramirez-Amaya, V., Insel, N., Plummer, T.K., Rosi, S., Chowdhury, S., Mikhael, D., Worley, P.F., Guzowski, J.F., Barnes, C.A.: Spatial exploration induces ARC, a plasticity-related immediate-early gene, only in calcium/calmodulin-dependent protein kinase II-positive principal excitatory and inhibitory neurons of the rat forebrain. J. Comp. Neurol. 498(3), 317–329 (2006)
- Burgess, N., Recce, M., O'Keefe, J.: A model of hippocampal function. Neural Netw. 7(6–7), 1065–1081 (1994)
- Brown, M.A., Sharp, P.E.: Simulation of spatial learning in the Morris Water Maze by a neural network model of the hippocampal formation and nucleus accumbens. Hippocampus 5, 171–188 (1995)
- Redish, A., Touretzky, D.: Cognitive maps beyond the hippocampus. Hippocampus 7(1), 15–35 (1997)
- Guazzelli, A., Corbacho, F.J., Bota, M., Arbib, M.A.: Affordances, motivation, and the world graph theory. Adapt. Behav. 6(3–4), 435–471 (1998)
- Arleo, A., Gerstner, W.: Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. Biol. Cybern. 83, 287–299 (2000)
- Gaussier, P., Revel, A., Banquet, J.P., Babeau, V.: From view cells and place cells to cognitive map learning: processing stages of the hippocampal system. Biol. Cybern. 86, 15–28 (2002)
- 38. Filliat, D., Meyer, J.-A.: Global localization and topological map learning for robot navigation. In: Hallam et al. (eds.) From Animals to Animats 7 Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior, pp 131–140. The MIT Press (2002)
- Arleo, A., Smeraldi, F., Gerstner, W.: Cognitive navigation based on nonuniform Gabor space sampling, unsupervised growing networks, and reinforcement learning. IEEE Trans. Neural Netw. 15(3), 639–652 (2004)
- Milford, M., Wyeth, G.: Spatial mapping and map exploitation: a bio-inspired engineering perspective. In: Winter, S., Duckham, M., Kulik, L., Kuipers, B. (eds.) Spatial Information Theory, pp. 203–221. Springer-Verlag, Heidelberg (2007)
- Parron, C., Save, E.: Evidence for entorhinal and parietal cortices involvement in path integration in the rat. Exp. Brain Res. 159(3), 349–359 (2004)
- Cho, J., Sharp, P.: Head direction, place, and movement correlates for cells in the rat retrosplenial cortex. Behav. Neurosci. 115(1), 3–25 (2001)

- McNaughton, B., Mizumori, S., Barnes, C., Leonard, B., Marquis, M., Green, E.: Cortical representation of motion during unrestrained spatial navigation in the rat. Cereb. Cortex 4, 27–39 (1994)
- Granon, S., Poucet, B.: Involvement of the rat prefrontal cortex in cognitive functions: a central role for the prelimbic area. Psychobiology 28(2), 229–237 (2000)
- Foster, D.J., Morris, R.G.M., Dayan, P.: A model of hippocampally dependent navigation, using the temporal difference learning rule. Hippocampus 10, 1–16 (2000)
- 46. Hull, C.L.: The goal gradient hypothesis and maze learning. Psychol. Rev. 39, 25–43 (1932)
- Weitzenfeld, A., Arbib, M., Alexander, A.: The neural simulation language. MIT Press, Cambridge (2002)
- Webb, B.: Can robots make good models of biological behaviour? Behav. Brain Sci. 24, 1033– 1050 (2001)
- 49. Ranck, J.B., Jr.: Head-direction cells in the deep layers of dorsal presubiculum in freely moving rats. Soc. Neurosci. Abstr. **10**, 599 (1984)
- Leutgeb, S., Leutgeb, J.K.: Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map. Learn. Mem. 14(11), 745–757 (2007)
- Ramirez-Amaya, V., Marrone, D.F., Gage, F.H., Worley, P.F., Barnes, C.A.: Integration of new neurons into functional neural networks. J. Neurosci. 26(47), 12237–12241 (2006)
- 52. Guzowski, J.F., Knierim, J.J., Moser, E.I.: Ensemble dynamics of hippocampal regions CA3 and CA1. Neuron **44**(4), 581–584 (2004)