27

Robot Topological Mapping and Goal-Oriented Navigation Based on Rat Spatial Cognition

Alejandra Barrera
Mexico's Autonomous Technological Institute (ITAM)
Mexico

1. Introduction

Spatial mapping has been subject of extensive research in the robotics field. Approaches to map building have been proposed, such as topological (Franz et al., 1998), metric (Movarec & Elfes, 1985), and hybrid maps combining these two approaches (Guivant et al., 2004; Kuipers et al., 2004; Bosse et al., 2004; Zivkovic et al., 2005). Diverse issues have arisen as critical to implement practical and robust solutions to the SLAM (Simultaneous Localization And Mapping) problem in real, large and static environments, such as data association, which relates to whether or not two features observed at different points in time correspond to one and the same object or place in the physical world (Hähnel et al., 2003), and perceptual ambiguity that occurs when trying to distinguish between places in the environment that may provide equivalent visual patterns (Frese, 2006).

In general terms, relevant aspects of concern are related to the evaluation of SLAM algorithms according to their computational complexity, their approach to the data association problem, and the sort of environment representation built. However, few efforts have been documented relative to the use of spatial representations to navigate directly to designated goal locations (Stentz & Hebert, 1995; Thrun et al., 2000). Even more, we have not found reports of attempts dealing with the "unlearning" process of previously learnt target locations. Therefore, one of the purposes of our research consists on addressing the imbalance between spatial mapping and map exploitation detected in the SLAM literature, as well as the lack of target unlearning research, by taking inspiration from these abilities in animals. Indeed, research cycles involving animal studies, computational 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. Specifically, the study of behavioral and neurophysiological mechanisms involved in rat spatial cognition provides a basis for the development of computational models of robot mapping and robot experimentation during goal-oriented navigation tasks. These models and robotic architectures offer neurobiologists and neuroethologists alternative platforms to study, analyze and predict spatial cognition based behaviors.

This chapter presents a computational system-level model of rat spatial cognition addressing aspects relative to cognitive map generation, adaptation, and exploitation during

navigation. This model relates rat 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 unknown environment, (v) pattern association using Hebbian learning, (vi) representation of the rat internal motivational state based on the hunger drive, (vii) management of rewards implemented by reinforcement learning using an Actor-Critic architecture allowing both learning and unlearning of goal locations, (viii) propagation of reward information through the cognitive map enabling both learning and unlearning of routes to a goal in the unknown environment, (ix) exploitation of maximum reward expectations after environment exploration to enable navigation towards the goal from any given departure location, and (x) on-line adaptation of the cognitive map to changes in the physical configuration of the environment.

The chapter starts by introducing relevant research background and related work. After providing a detailed description of the spatial cognition model, the manuscript presents results from its evaluation through mobile robot experimentation during tasks such as (i) learning and unlearning of reversal behaviors in a T-maze and in an 8-arm radial maze, (ii) place recognition and goal-oriented navigation in multiple T-mazes having external landmarks where the robot is trained to learn the correct route to a goal from a fixed location in the maze, and tested to reach the goal from different starting positions, and (iii) on-line adaptation of the cognitive map in those multiple T-mazes where landmark configuration is modified after having trained the robot to find the goal. Finally, the chapter concludes by addressing contributions of the presented work from biological and robotics perspectives, and discussing next directions to extend computational modeling of spatial cognition in rodents and corresponding experimentation in robots.

2. Research Background

The position in space is part of the fundamental information that animals and people learn. In addition to the memory of the position of objects or places, the memory to move through space is equally important to reach a point B from a point A at which the individual is located. We use the term spatial cognition to refer to the process of coding, storing and exploiting spatial information for successful orientation and movement toward specific goals in space (Roberts, 1998).

2.1 Cognitive Mapping

In understanding the brain mechanisms involved in the processing of spatial information, O'Keefe and Nadel (1978) 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 animal position, and (ii) spatial learning allowing the animal to solve navigation problems that require memory of such representations. The neural substrate of such internal spatial representations was the prior discovery made by O'Keefe and Dostrovsky (1971) from individual records of place cells found in hippocampal substructures CA3 and CA1. Each of these cells exhibits high rate activation when the animal is in a continuous and compact area defined as the place field of the cell. The activity derived from the place cell population codifies the current location of

the animal within a familiar environment, and it is stored in the internal spatial representation referred to as cognitive map by Tolman (1948).

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 (McNaughton et al., 1994a). However, place cells maintain their fields when some of the visual cues are removed from the environment (O'Keefe & Conway, 1978), and even continue to respond in the dark (Quirk et al., 1990). Hence, it is assumed that the response of place cells is derived from combining kinesthetic and visual cues information (Jeffery & O'Keefe, 1999).

According to Poucet (1993), 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 (Moser et al., 2008).

2.2 Motivation and Learning

The motivated behavior is usually oriented towards a goal that in animals may be associated with a drive such as hunger, and can also be stimulated by external incentives such as food smell. The hypothalamus is regarded as the main brain area where information about the internal state of the rat is combined with incentives (Risold et al., 1997). Specifically, it is assumed that food pursuit and intake are activities controlled by the lateral hypothalamus (Kelley, 2004), which determines the primary reward these activities produce in animals.

In addition to the hypothalamus, the striatum in 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 (Schultz et al., 1998).

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 (Schultz et. al, 1998). Houk et al. (1995) suggested the striatum implements an Actor-Critic architecture (Barto, 1995), 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) (O'Keefe & Nadel, 1978). Neurophysiological and neuroanatomical studies have shown that procedural learning relies on the striatal system, whereas spatial learning employs the hippocampal system (Hartley & Burgess, 2005).

2.3 Experimental Basis for Spatial Cognition Processes in Rats

Interested in demonstrating the use of both learning strategies, spatial and procedural, in normal rats and rats with hippocampal lesions, O'Keefe (1983) implemented two discrimination tasks.

The first task took place in an 8-arm radial maze without external signals (see Fig. 1(a)), where five arms were removed to form a T-shaped maze (see Fig. 1(b)) and a reward (food) was placed at the end of the left arm of the T. Every five trials during training, the maze was modified so that other three arms constitute the T, preventing the animal to base its choice of the correct arm on signals outside the maze, but on the spatial or procedural strategies. As a result of training, rats with injured hippocampus learned the task faster than normal rats.
To determine the type of learning strategy used by animals, O'Keefe continued the experiment after reaching the learning criterion now alternating between testing trials in the radial maze and in the T maze. During tests, all rats chose arms in the radial maze located on the same side of the reward arm in the T, thus showing the use of a procedural strategy since rats learned to rotate left.

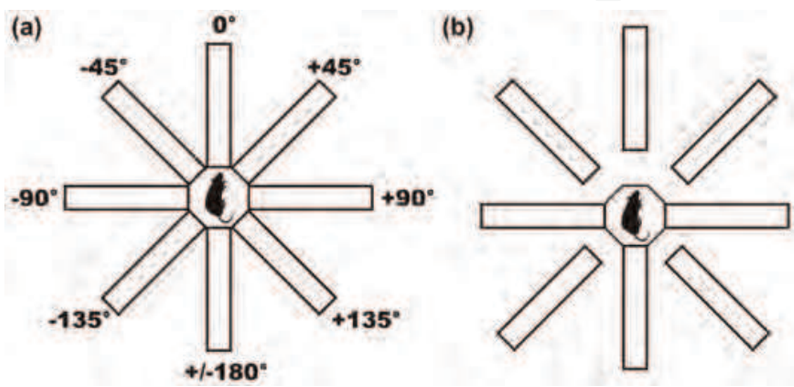

Fig. 1. Diagrams of mazes employed by O'Keefe during discrimination tasks. (a) The 8-arm radial maze showing the orientation of arms according to an egocentric (local) reference frame, which specifies the rat behavior in terms of rotations relative to its body midline. (b) A T-maze resultant from separating five arms of the original 8-arm radial maze.

The second task consisted on a reversal discrimination training all animals to turn left in the T-maze, and testing their orientation decisions when changing the reward location at the end of the right arm of the T. During tests, O'Keefe alternated between two trials in the T and one trial in the radial maze. Results obtained from injured animals showed an abrupt change in the rotation performed at the T-junction from the left to the right arm, but in the radial maze, the rat orientation change was continuous and incremental from the training quadrant (-90° -45°), through straight ahead (0°) until the testing quadrant (+45°, +90°) (see Fig. 1(a)). On the other hand, the performance of healthy rats in the T-maze proceeded in the same manner as that of injured animals, but their orientation in the radial maze did not change on a continuous and incremental manner, but randomly.
The outcome of the reversal task allowed O'Keefe to ratify the following:
• Injured rats used a procedural learning strategy during training, associating the reward with the response of turning left. During tests, extinction of the reward on the left T-maze arm gradually modified the animals' orientation until turning right.
• Healthy rats employed a spatial learning strategy during training before reaching the learning criterion. However, after this event, they based their learning on a procedural strategy that abruptly ended with the reward extinction due to the intervention of the hippocampus to shoot again the environment exploration and the corresponding use of the spatial learning strategy.

According to O'Keefe and Nadel (1978), the cognitive map is responsible for estimating the rat's position within the environment and its relative direction to the goal location, enabling the animal to reach this location from any other one. One classical spatial task showing this goal-directed navigation behavior is the water maze introduced by Morris (1981).

Under the original Morris experiment (Morris, 1981), the maze consisted of a circular tank filled with an opaque mixture of milk and water including a platform in a fixed location. In each training trial, normal rats and rats with hippocampal lesions were independently placed at one of various departure locations near the tank's circumference, and required to swim until they located the platform, upon which they could stand and escape from the cold water. After training, rats were tested in two situations: (i) with the platform visible and (ii) with the platform submerged inside the tank and visual cues placed around the arena. In the first case, all rats were able to find the platform immediately, whereas in the second, only normal rats found it from any known or novel starting location at the periphery of the tank.

The experiment also tested situations where the platform was removed from the tank or placed in a random location from trial to trial. In the first case, healthy rats looked for the platform persistently in its previous location, whereas in the second one, healthy rats could not learn to navigate directly to the platform.

Successful navigation in the water maze task by hippocampus healthy animals involved the use of a spatial learning strategy that allowed them to relate the goal location with external landmarks, store this relation in a cognitive map, and use the map to find a correct route.

2.4 Related Work

Taking inspiration from spatial cognition in rats, several computer-simulated or robot-tested navigation models have been proposed. From this point, we use the term "animat" to refer the modeled rat.

The models by Burgess et al. (1994) and Brown and Sharp (1995) include location and head direction signals, as well as reward-related changes in the efficacy of synapses onto motor cells that control the animat behavior. In these studies, different from our model, hippocampal place cells provide a code for space that is not stored explicitly in a cognitive map. Burgess et al. (1994) use metric information, such as distances to identified visual cues, as the exclusive and direct input to their system. In contrast, place units in our model codify the integration of visual and kinesthetic information, and we interpret metric properties of visual cues by means of neurons sensitive to specific landmarks information patterns. Our model builds and maintains a topological map storing place cell population activity and enabling navigation between locations in the environment. Brown and Sharp (1995) associate place cells with motor responses, and perform a trace-like learning rule where a given animat behavior immediately followed by reward has a higher probability of occurrence the next time the animat experiments the same situation. In contrast, our model implements reinforcement learning through an Actor-critic architecture enabling the animat to learn as well as unlearn reward locations.

Actually, prior use of this reinforcement learning method to solve goal-search tasks has been documented (Guazzelli et al., 1998; Foster et al., 2000). The work by Foster et al. (2000), for instance, implements an Actor-Critic architecture to enable a simulated rat to solve the reference memory task in a Morris water maze providing the animat with a reward at any given time of the experiment. If the animat moves towards the hidden platform (goal) in the water maze, the reward is set to 1; otherwise, it is set to 0. The animat is able to learn correct

actions precisely because such a reward signal is given as input at each time step. On the contrary, the animat in our model does not receive any information related to its physical proximity to the hidden goal during the exploration of the environment, but it is the animat who traces the correctness of its followed route once it finds the goal location. This route learning is achieved by propagating reward information through nodes in the topological map after every trial in the given experiment.

Our work is partially inspired on the computer-simulated model by Guazelli et al. (1998). One of our main extensions to this system includes a map exploitation process to enable goal-directed navigation in a mobile robot. Their original model endowed the simulated rat with the ability to learn goal locations from a fixed departure position within mazes that included just one decision point from where the goal was visible, and the animat was unable to find the target in more complex mazes having two or more decision points, and also to reach it from arbitrary starting positions. Specifically, in the model by Guazzelli et al., the action selection process considers the addition of reward expectations derived from affordances with reward expectations from their world graph model. As reward expectations from affordances were computed over the assumption of having different rotation possibilities at every decision point in the environment, the simulated rat fails to find the goal when executing the same reinforced rotation at two or more decision points offering the same rotation possibilities. We also extend the original model by providing a map adaptation process that permits on-line representations of changes in the physical configuration of the environment perceived by the robot (Barrera & Weitzenfeld, 2007b).

Our proposal differs from the model of Gaussier et al. (2002) in that they employ only visual information as input, hippocampal cells do not encode places but transitions between states, and the place recognition process is carried out by the entorhinal cortex rather than by the hippocampus. As in our model, they build a topological space representation; however, nodes in this map do not correspond to places, but to transitions between states. They implement a sort of map exploitation to determine sequences of transitions between states that lead to a goal location. Nonetheless, they do not model the animal's motivation and the prediction of reward expectations.

The main components of the neural architecture proposed by Arleo et al. (2004) are similar to those found in our model: integration of allothetic (visual) information and idiothetic (kinesthetic) signals at the level of hippocampal representation, use of Hebbian learning to correlate these inputs, mapping of the place cell population activity into spatial locations, and application of reinforcement learning to support goal-oriented navigation. We add to this model the use of affordances information instead of population vector coding to map the ensemble dynamics of place cells into spatial locations, an explicit construction of a topological map of places and their metric relations, and the implementation of an Actor-Critic reinforcement architecture that predicts, adapts and memorizes reward expectations during exploration to be exploited during goal-oriented navigation.

The focus of our approach differs from the one followed by Milford et al. (2007). Whereas they are concerned with the effectiveness of the hippocampus models in mobile robot applications exploring large environments with natural cues, our interest consists on endowing mobile robots with spatial cognition abilities similar to those found in rodents in order to produce comparable behavioral results and eventually provide experimental neuroscience with valuable feedback. Nevertheless, our model coincides with Milford et al.'s in some aspects related to mapping, map exploration and map adaptation, and

contrasts with it in map exploitation. Specifically, in the model by Milford et al., a topological map of experiences is built with each experience representing at a given time a snapshot of the activity within pose cells, which codify physical localization and orientation, and local view cells that encodes for visual information. In this map, transitions between experiences are associated with locomotion information. Nodes in our topological map also represent associations between visual-kinesthetic information patterns and the place cell population activity. Transitions between nodes are associated with metric information derived from the animat locomotion. During the exploration of the environment, previously unused movement behaviors are implemented by the animat in trying to experience new routes. Similarly, motion in our model is partially influenced by the curiosity to execute rotations not yet explored by the animat. Additionally, the map of experiences can be adapted to physical changes in the environment, which involves the elimination/creation of experiences and the update of transitions between experiences. We have demonstrated that the map built by our system is adapted on-line to represent changes in the physical configuration of landmarks (Barrera & Weitzenfeld, 2007b). The map exploitation process involves the storage of temporal information in the experiences map to find the fastest route to the goal, and the storage of behavioral information in transitions between experiences to navigate to the goal. In contrast, our model provides map exploitation based on reward expectations of locations predicted and adapted during environment exploration and used to guide the animat towards the goal from any given departure point.

3. A Bio-inspired Computational Model of Spatial Cognition

The spatial cognition model comprises distinct functional modules, shown in Fig. 2 and described below, that capture some properties of rat brain structures involved in learning and memory. A detailed mathematical depiction of each module is presented in (Barrera & Weitzenfeld, 2008).

3.1 Motivation

Motivation module relates to the rat lateral hypothalamus (LH) computing the value of hunger drive and producing the primary reward (r) the animat gets by the presence of food (goal). The reward r depends on the animat internal drive experimented at any given time.

3.2 Kinesthetic Processing

Rats carry out path integration processes using kinesthetic information derived from two systems: (i) vestibular organs in the semicircular canals of the inner ear, and (ii) feedback from muscles controlling movement. During path integration, rats update the position of their point of departure each time they move in relation to their current position (Mittelstaedt & Mittelstaedt, 1982; Etienne, 2004). 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 (Parron & Save, 2004) involving also the retrosplenial cortex (RC) (Cho & Sharp, 2001).

In our model, we attribute to PPC the representation of the updated animat position in relation to its point of departure 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 (Hebb, 1949). *PI* is defined as a matrix of activation values registered by all neurons in PIFDL at any given time.

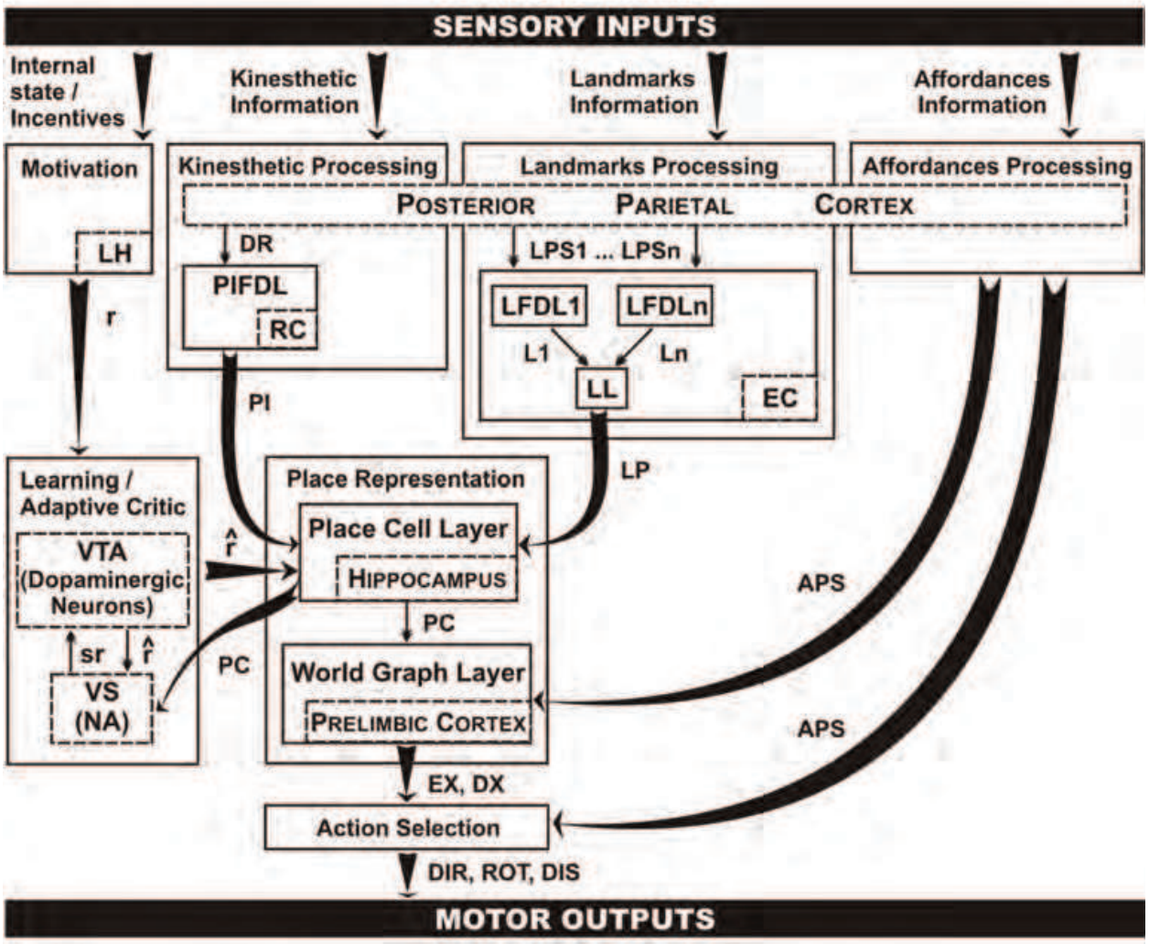

Fig. 2. The modules of the spatial cognition model and their interaction. Glossary: LH – Lateral Hypothalamus; RC – Retrosplenial Cortex; EC – Entorhinal Cortex; VTA – Ventral Tegmental Area; VS – Ventral Striatum; NA – Nucleus Accumbens; PIFDL – Path Integration Feature Detector Layer; LFDL – Landmark Feature Detector Layer; LL – Landmarks Layer. Inputs/Outputs: r= primary reinforcement; sr= secondary reinforcement; $\hat{r}$ = effective reinforcement; DR= dynamic remapping perceptual schema; LPS= landmark perceptual schema; APS= affordances perceptual schema; PI= kinesthetic information pattern; L1, Ln= information pattern for landmark 1, landmark n; LP= landmarks information pattern; PC= place information pattern; EX= maximum reward expectations and their corresponding directions (DX); DIR= next animat direction; ROT= animat rotation; DIS= next animat moving displacement.

3.3 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 and Touretzky (1997), 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 environment. In this way, *LP* is defined as a matrix of activation values of all neurons in LL representing the egocentric view from the animat at any given time.

3.4 Affordances Processing

McNaughton et al. (1994b) suggested that preceding the rat 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. 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° in 45° intervals at any given time from any given animat location (see Fig. 1(a)).

3.5 Place Representation

As shown in Fig. 3(a), place representation module receives input from kinesthetic (*PI*), landmark (*LP*) and affordances (*APS*) modules in addition to its interaction with the learning module ($\hat{r}$, later defined in Section 3.6). Place representation module comprises a place cell layer (PCL) and a world graph layer (WGL).

Place cell layer (PCL) corresponds to the rat hippocampus encompassing regions CA3, CA1 and 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:

$$PC(t) = PI(t)\ W_{PI}(t) + LP(t)\ W_{LP}(t), \tag{1}$$

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.

Associations between overlapping place fields and physical areas are represented by world graph layer (WGL) through a topological map enabling navigation between 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 (Granon & Poucet, 2000).

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. 3(b). 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*).

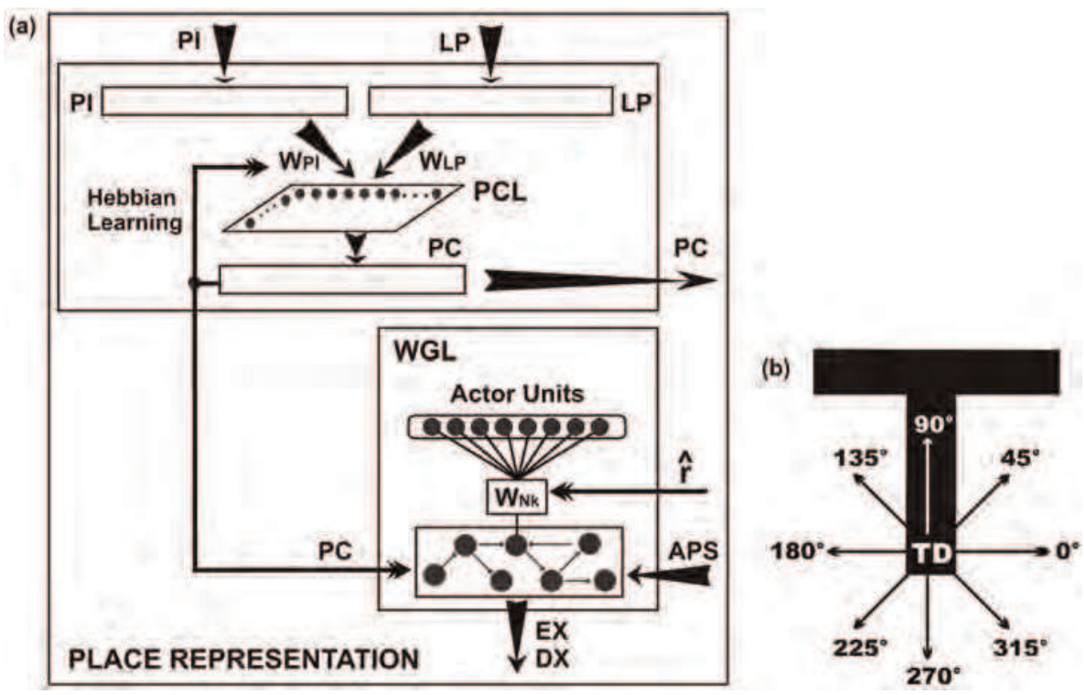

Fig. 3. (a) The place representation module of the spatial cognition model. Glossary: PCL= place cell layer; WGL= world graph layer; PI= kinesthetic information pattern; LP= landmarks information pattern; W_{PI}= connection weights between PIFDL and PCL; W_{LP}= connection weights between LL and PCL; 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 maximum reward expectations and their directions (DX). (b) Allocentric (global) 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.

Map Creation

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 *pat* as follows:

$$SD = \sum_{i=1}^{n} \min(pat_{1,i}\ , PC_{1,i}) \Big/ \sum_{i=1}^{n} PC_{1,i} \quad , \tag{2}$$

where *pat* and *PC* are matrices of *1×n* activation values registered by the collection of *n* neurons in PCL, *i* is the matrix column index, and *min* is a function that computes the minimum value between its two arguments.

The model distinguishes among two cases:

(i) If at least one *SD* exceeds certain threshold close to 1, the Actor unit storing the activation pattern with the biggest *SD* is considered the winner.

(ii) 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 (see Fig. 4(a)).

(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 (see Fig. 4(b)).

(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 (see Fig. 4(c-d)).

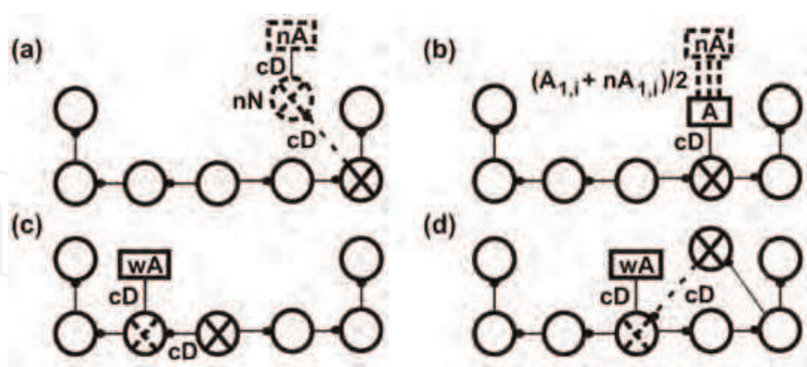

Fig. 4. Creation/activation of nodes in the map. Dotted lines illustrate new components, and the crossed node is the active one. Glossary: nA= new Actor unit; nN= new node; cD= rat's current direction; A= existing Actor unit; wA= winner Actor unit. (a) The creation of a new node. (b) The average between the activation patterns of two Actor units. (c) The activation of an existing node. (d) The connection between two existing nodes.

Map Adaptation

During exploration, an existing topological map may be adapted according to kinesthetic and visual information perceived by the animat in its current orientation due to two reasons:

(i) Late recognition of already visited locations. When the animat is exploring a given maze location represented by a certain map node *nX* in a particular direction *d1*, then the animat executes a rotation orienting itself to direction *d2* and an existing node *nY* is activated at that time, an integration process occurs between both active nodes *nX* and *nY* since they are representing the same maze location in two different directions. This integration process may or may not involve the creation of a new node in the map:

a) If node *nX* was created at time *t-1* and the activation of node *nY* occurs at time *t*, the arc pointing to node *nX* in direction *d1* is now updated to point to node *nY*, Actor unit associated to direction *d1* of node *nX* is assigned to node *nY*, node *nX* is eliminated, and node *nY* prevails as the only one active (see Fig. 5(a)).

b) If node *nX* was not created but just activated in direction *d1* at time *t-1* and the activation of node *nY* occurs at time *t* in direction *d2*, a new node *nZ* is created at this time integrating all input/output arcs to/from nodes *nX* and *nY*, as well as all Actor units of both original nodes. Merged nodes *nX* and *nY* are removed from the map, and node *nZ* becomes the active one (see Fig. 5(b)).

(ii) Dynamic changes in the physical configuration of the environment. When the environment is modified during the exploration process, the animat stops recognizing locations from where a different view is now obtained. In this case, the animat adds new map nodes establishing appropriate connections, and removes all those nodes that are no longer linked to the map (see Section 4.3 for detailed examples).

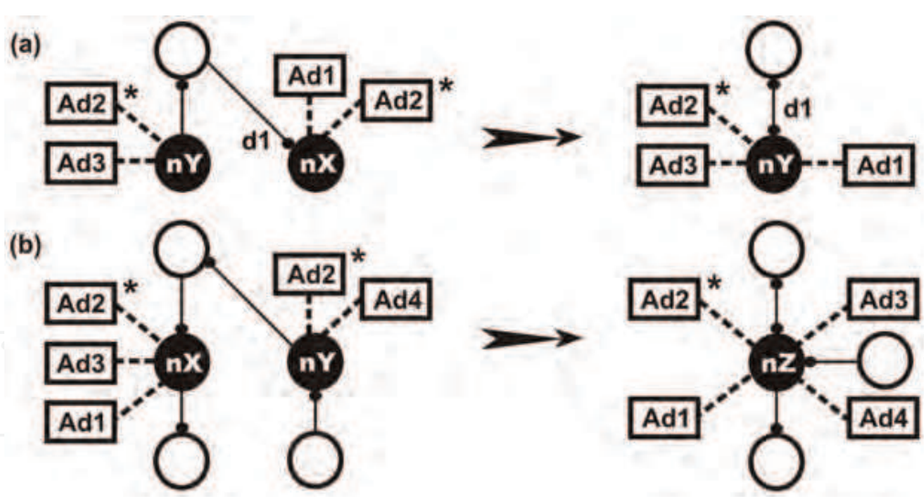

Fig. 5. Illustration of map adaptation resultant from late recognition of already visited locations. Glossary: Ad1 – Ad4= existing Actor units of different nodes; *= active node in direction *d2*. (a) The incorporation of node *nX* created at time *t-1* into the existing node *nY* activated at time *t* in animat direction *d2*. (b) The integration of existing nodes *nX* and *nY* into the new node *nZ* with both, *nX* and *nY*, active at time *t* in animat direction *d2*.

3.6 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. As illustrated in Fig 6, the Adaptive Critic (AC) includes a Prediction Unit (PU) that estimates the future reward value of any particular 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 (3):

$$P(t) = PC(t) \cdot W_{PCL}(t), \tag{3}$$

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.

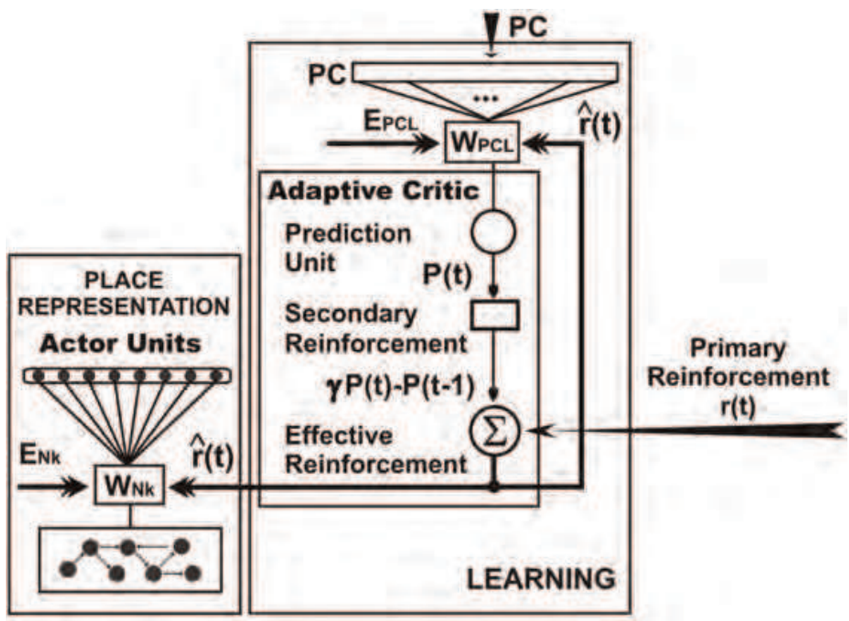

Fig. 6. The learning module of the spatial cognition model. Glossary: PC= current activation pattern in the PCL layer; W_{PCL}= connection weights between PCL and Prediction Unit; E_{PCL}= connection eligibility traces between PCL and Prediction Unit; *P(t)* and *P(t-1)* correspond to predictions of the future value of *PC* at time *t* and *t-1* respectively; W_{Nk}= connection weights between any given map node *k* and its corresponding Actor units; E_{Nk}= connection eligibility traces between any given map node *k* and its corresponding Actor units.

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 (4):

$$\hat{r}(t) = r(t) + \gamma P(t) - P(t-1). \tag{4}$$

The effective reinforcement is used to update the connection weights between PCL and PU in AC (i.e., 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 use

$$W_{PCL}(t+1) = W_{PCL}(t) + \beta \hat{r}(t) E_{PCL}(t), \tag{5}$$

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 use

$$W_{Nk}(t+1) = W_{Nk}(t) + \beta \hat{r}(t) E_{Nk}(t) \quad \forall \, map\ node\ k, \tag{6}$$

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 (5) and (6), both learning rules depend on the eligibility of the connections. At the beginning of every 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 only when the action executed by the animat at time t-1 allowed it to perceive the goal:

$$E_{PCL}(t) = E_{PCL}(t-1) + \chi PC(t), \tag{7}$$

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 map node na and the Actor unit corresponding to the current animat orientation dir is increased by τ as described by (8):

$$e_{na}^{dir}(t) = e_{na}^{dir}(t-1) + \tau. \tag{8}$$

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 (9):

$$\begin{aligned} E_{PCL}(t) &= \lambda E_{PCL}(t-1) \\ E_{Nk}(t) &= \sigma E_{Nk}(t-1) \quad \forall \, map\ node\ k. \end{aligned} \tag{9}$$

The use of the Actor-Critic architecture enables the estimation of reward expectation values of different locations in the environment, where maximum expectations correspond to locations from where the goal is perceptible and to orientations needed to be performed at those locations to reach the goal. Since the model allows the animat to recognize the goal just one step away from it in order to prevent a taxon or guidance navigation strategy, and the animat does not receive any information related to its physical proximity to the hidden goal during the exploration of the environment, it is the animat who traces the correctness of its followed route once it finds the goal location. This route learning is achieved by

implementing backward reward propagation through nodes in the topological map after every trial in the given experiment.

The backward reward propagation involves updating eligibility traces of Actor units in the direction of the arcs connecting the map nodes that represent the complete route followed by the animat. The strategy involved in this process is based on the existence of a factor referred to as goal gradient by Hull (1932), according to which the reinforcement effect is the most at the goal location and diminishes progressively as the animal moves backward through the environment. Specifically, in case the animat finds the goal at the end of the path, each eligibility trace is updated in a given positive amount of reinforcement divided by the amount of steps the animat performed to move from one node to the next one. Likewise, in case the animat does not find the goal at the end of the path, the amount of reinforcement used is negative. The reinforcement is initialized to a certain amount at the beginning of every training trial in the experiment, and this amount decreases as the distance from a node to the goal increases.

The update on eligibility traces that occurs when the given trial is concluded at time t renders a corresponding update on the connection weights between map nodes (representing the complete route) and Actor units (associated to animat orientations) at time $t+1$ according to (6), enabling the animat to learn as well as unlearn performed routes.

The positive update registered when the target is found resembles the activation pattern of rat striatal cells. Mulder et al. (2004) showed that after having learnt to locate the reward in a maze, striatal cells of the rat respond continuously for the complete path followed by the animal, implying that striatum cells are able to modify their synaptic weights to the inputs received from the place cells responding to locations in the traveled route.

3.7 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 to execute from current location and orientation (*APS*),
(ii) one random rotation between possible affordances (*RPS,* internally computed),
(iii) curiosity to execute rotations not yet explored (*CPS,* internally computed), and
(iv) maximum reward expectation (*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 vectors. Specific positions distributed throughout a vector correspond to particular relative rotations between -180° and +180° 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.

4. Robot Experimentation

The rat cognitive model was designed and implemented using the NSL system (Weitzenfeld et al., 2002). The computational model interacts with a real robotic environment through an external visual processing module getting as input three non-overlapping snapshots (0°, +90°, -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 (Barrera & Weitzenfeld, 2008) for further detail on the robotic implementation of the model.

This section presents results from the evaluation of the model through robot experimentation in a single T-maze, an 8-arm radial maze, and multiple T-mazes, performing spatial reversal behaviors, place recognition and goal-oriented navigation, and map creation and adaptation.

4.1 Reversal Behavior in a T-maze and in an 8-arm Radial Maze

The experiment carried out with the robot in the T-maze and in the 8-arm radial maze is inspired on the reversal task implemented by O'Keefe (1983) and described in Section 2.3, and was performed in both mazes separately.

In the T-maze and in the 8-arm radial maze shown in Fig. 7, the robot departs from location TD and navigates to any arm extreme. This process is repeated in every experiment's trial, at the end of which the robot is manually placed again at the departure location TD.

During the training phase of the experiment, the goal is placed at the end of the arm oriented to 180° (see Fig. 7). As abovementioned in Section 3.6, the system allows the robot to recognize the goal just one step away from it in order to prevent a taxon or guidance navigation strategy. At the beginning, the decisions of the robot at the choice location are determined by the curiosity for unexecuted rotations. After having visited each arm once, the curiosity level decreases notably, thus the robot executes random rotations visiting different maze arms. Eventually, the robot meets the learning criterion once the expectation of finding reward in orienting to 180° at the choice point becomes bigger than the random factor. When this event occurs, the training phase ends.

The average duration of the training phase obtained from experimenting separately with six robots in both mazes was 12 trials performed in less than 20 minutes.

When the testing phase begins, the goal is moved to the arm oriented to 0° (see Fig. 7). This situation constitutes, in the words by O'Keefe, "a reversal discrimination problem that involves an unlearning process giving up a previously correct hypothesis and switching to a new one" (O'Keefe and Nadel, 1978).

During reversal, the expectation of future reward for the arm oriented to 180° decreases continuously each time the robot reaches the end of that arm not finding reward, since this event is coded as frustrating by the learning module of the model. When the expectation of reward becomes smaller than the random factor, the robot starts visiting other arms. Each time it visits the 0° arm that provides reward, the expectation for this arm increases. Eventually, the robot meets the learning criterion when the expectation of reward for turning right at the choice point is large enough to avoid exploring any other maze arm.

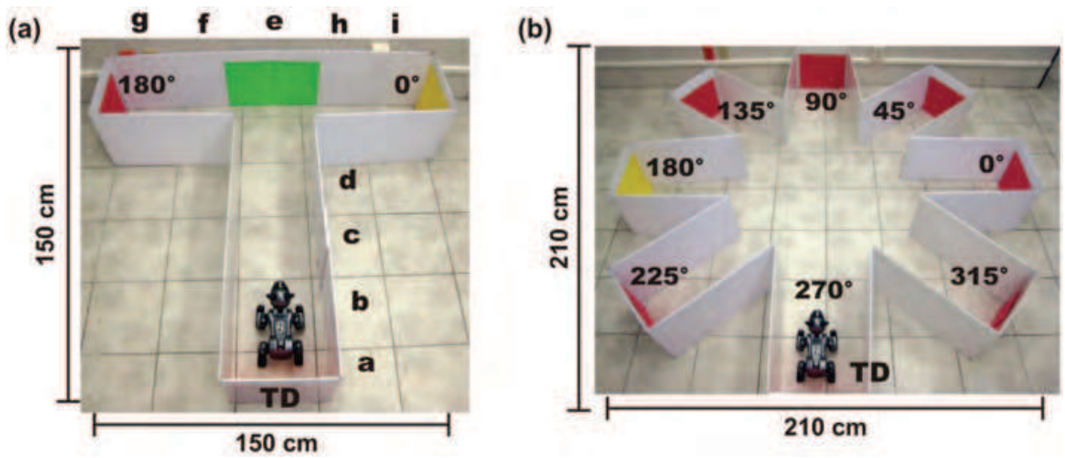

Fig. 7. (a) The physical T-maze used in the reversal task. Different locations are labeled with letters. (b) The physical 8-arm radial maze used in the reversal task. The picture shows the allocentric (global) directions of the arms in the maze. In (a) and (b), the AIBO robot is located at the starting position TD.

The average performance of six robots during reversal in the T-maze is shown in Fig. 8(a), where the robot's reversal behavior is expressed in terms of percentage of correct choices at the T-junction. Likewise, the average performance of those six robots during reversal in the 8-arm maze is shown in Fig. 8(b), which presents the robot's orientation decisions at the choice location of the maze in egocentric (local) terms (see Fig. 1(a)).

As a result from training in the T-maze, 100% of the robot's choices were correct (tag "control" in Fig. 8(a)). The graph in Fig. 8(a) shows 32 testing trials. As can be seen, the robot takes 12 trials to unlearn (i.e., decrease reward expectations) the previously correct hypothesis (i.e., arm oriented to 180°) (tag "criterion" in Fig. 8(a)). From trial 13, the robot has learnt (i.e., has increased reward expectations) the new one (i.e., arm oriented to 0°). In this way, the percentage of correct choices shifts from 36% in trial 12 to 95% in trial 16 and 100% from there. On the other hand, as a result from training in the radial maze, the robot chose consistently to turn left at the center location of the maze (tag "control" in Fig. 8(b)). During reversal, the robot's orientation did not reveal any systematic shift. As in the T-maze, the criterion occurred around trial 12, when the average orientation crosses the midline of the graph in Fig. 8(b).

Comparing our results with those reported by O'Keefe with normal rats in (O'Keefe, 1983), we can appreciate a behavioral similitude with robots in the T-maze. O'Keefe presented the average results obtained from four rats. His graph also shows 32 testing trials and a control measure of 100% of correct choices at the T-junction after training. In this case, rats reached the criterion in trial 20, where the percentage of correct choices was between 20% and 40%. By trial 24, rats chose the new reward arm in more than 90% of the times until performing 100% of correct decisions. As described by O'Keefe (1983), we could confirm that there was an abrupt change from the incorrect to the correct arm.

We also appreciate in our results a similitude with those reported by O'Keefe with four normal rats in the radial maze (O'Keefe, 1983). He also presented a graph showing 32 testing trials, where rats reached the criterion around trial 20. He explained that the abrupt shift from turning left to turning right in the T-maze was revealing the moment when the average

orientation crosses the midline in the radial maze. According to our tests, the same fact applies to the robot's behavior.

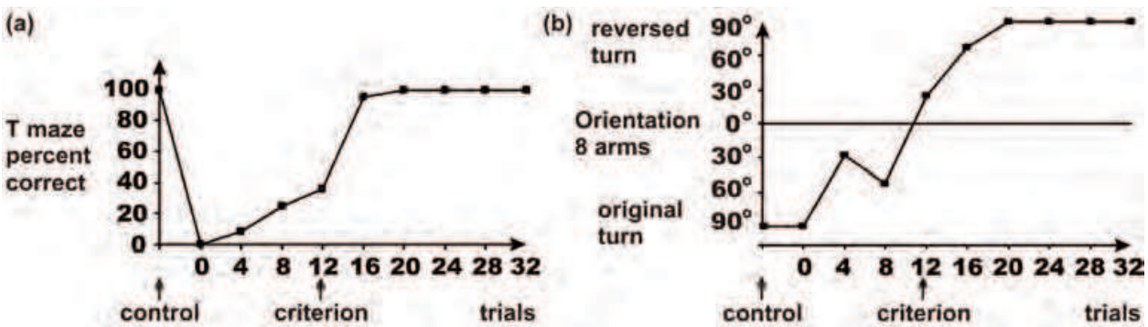

Fig. 8. The performance of six robots during the reversal task. Each graph was obtained by averaging the graphs of the individual robots. The graph in (a) shows the percentage of correct choices in the T-maze averaged over periods of four trials. The graph in (b) presents robot decisions in the radial maze also averaged over periods of four trials. To compare these results with those obtained by O'Keefe with four normal rats, refer to (O'Keefe, 1983). In our graphs, as in O'Keefe's, the abrupt shift from turning left to turning right in the T-maze reveals the moment (criterion) when the average orientation crosses the midline in the radial maze.

During the experiment, the robot builds and maintains the topological spatial representations shown in Fig. 9 for the T-maze and the 8-arm radial maze. The ensemble activity of the neurons found in the place cell layer of the model was determined only by the use of kinesthetic information, since no landmarks were available. The activity pattern of the collection of place cells registered when the robot is at any given maze location is stored in a particular Actor unit associated with the current robot's orientation and linked to the active map node.

As described in Section 3.5., the relevance of locations in a maze relies on the presence of rewards or on affordances changes sensed by the robot during exploration. In Fig. 7(a), for instance, the robot considers locations "a", "b", "e", "f", "g", "h" and "i" as relevant, and that is why the map in Fig. 9(a) includes seven nodes. When the robot reaches location "b" in direction 90°, the current ensemble activity of the place cell layer is stored in node 2, and although the activity pattern could slightly vary at locations "c" and "d", the affordances sensed by the robot did not change from "b" to "c" or "d", thus activity patterns registered in these locations are averaged and stored in the same node 2, defining in this way its physical extension. It should be pointed out that, different from rats, the robot was programmed to avoid 180° rotations when there exist other possible rotations (0°, +90°, -90°), thus optimizing the exploration process to find the goal. For this reason, arc directions between nodes in the map are one way.

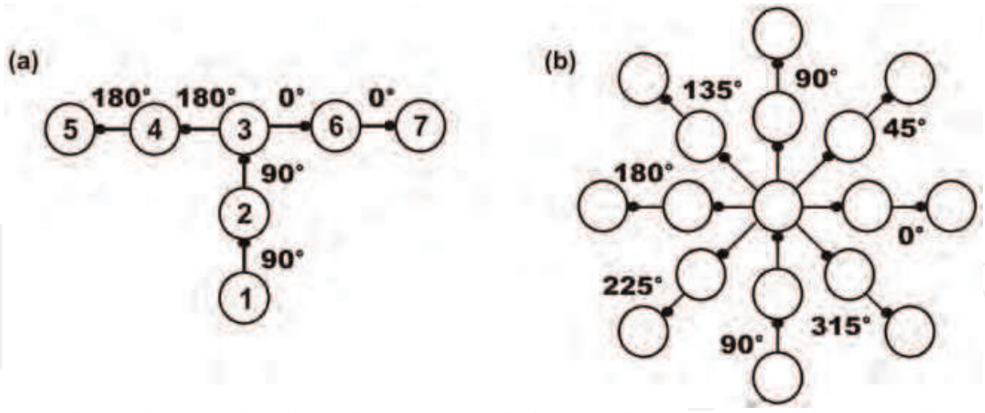

Fig. 9. Topological maps built and maintained during the reversal task in (a) the T-maze, and (b) the 8-arm radial maze. In (a) nodes are numbered in order of creation. In (a) and (b) arcs are associated to the robot direction when it moved from one node to the next one.

4.2 Place Recognition and Goal-oriented Navigation in Multiple T-Mazes

In (Barrera & Weitzenfeld, 2007a) we describe a variant of the classical Morris water maze procedure (Morris, 1981). On the basis of the results presented by Hollup et al. (2001), we supported our decision of using a land-based maze, and not implementing open-field navigation to exploit the affordances module of the spatial cognition model. Therefore, we designed and employed in this task a maze formed by multiple Ts surrounded by three colored cylinders representing landmarks as shown in Fig. 10(a).

In the reversal experiment just described in Section 4.1, the primary objective was to test the ability of the robot to learn the correct route to the goal from a fixed location by using just kinesthetic information, and to unlearn that route while learning the new one that leads to the target. In the task presented in this section, the objective is two-fold: (i) to test the place recognition process carried out by the robot employing not only kinesthetic but also visual information while exploring the maze, and (ii) the goal-directed navigation to find the goal from different starting locations.

The training phase proceeds as in the previous experiment in the T-maze and in the radial maze; i.e., in any given trial, the robot explores the maze until it finds the goal or the end of a corridor. Different from the usual Morris procedure, where animals start from a different location each training trial, the robot in our task, instead, starts from the same fixed location each time (TD in Fig. 10(a)). This variation has also been implemented in other existing studies such as that by Eichenbaum et al. (1990) with successful results.

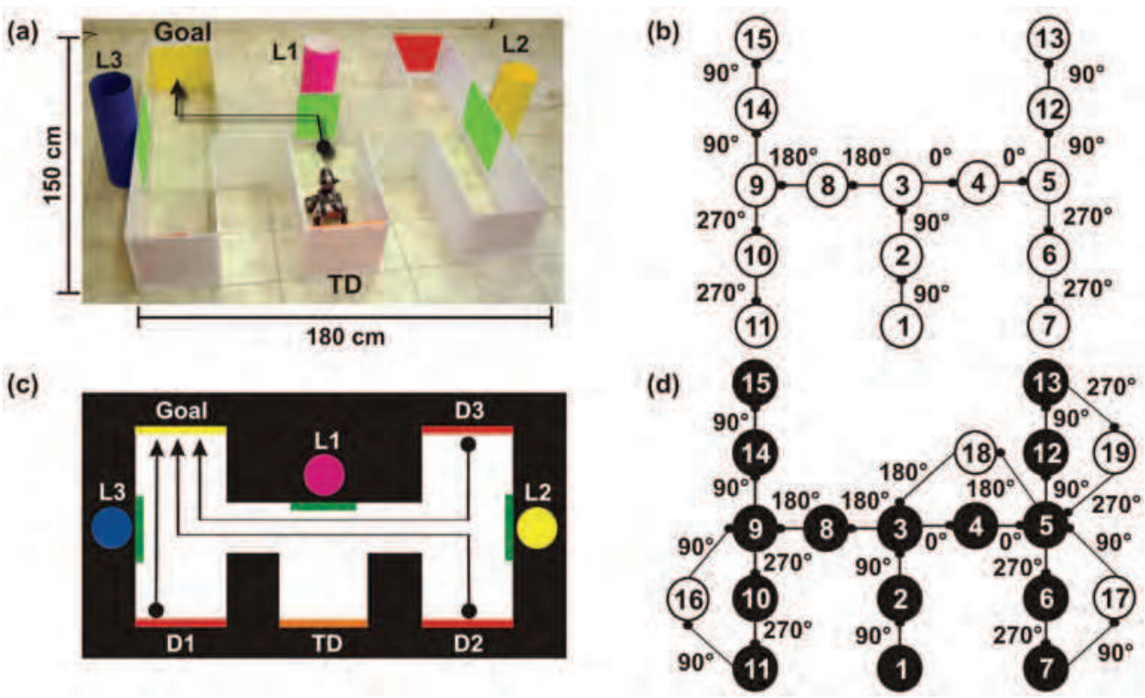

Fig. 10. (a) The physical maze formed by multiple Ts surrounded by landmarks L1, L2 and L3. The AIBO robot is located at the training starting position (TD), and the path to the goal is marked with an arrow. (b) The map built by the robot during training. (c) Routes followed by the robot in three testing trials departing from locations D1 to 90°, D2 to 90° and D3 to 270°. (d) The map updated by the robot during tests, showing existing nodes in black and new nodes in white. In (b) and (d) nodes are numbered in order of creation and arcs are associated to the robot direction when it moved from one node to the next one.

During maze exploration, the robot built a map similar to the one shown in Fig. 10(b), and took in average five training trials reaching the goal to learn the route from location TD. For six robots, the average duration of the training phase was 9 trials, i.e. 23 minutes approximately.

During testing trials, we placed the robot at different departure locations and orientations: D1, D2 and D3 in orientations 90°, 90° and 270°, respectively (see Fig. 10(c)). All six robots found the goal successfully from all starting positions tested, following (in most cases) direct routes as shown in Fig. 10(c), and updating the map of the environment as illustrated by Fig. 10(d). The average duration of trials starting at D1, D2 and D3 were 50 seconds, 120 seconds and 140 seconds, respectively.

To reach the target from any given location, the robot performs place recognition and map exploitation processes. Let's consider the case starting at location D2 (see Fig. 10(c)). The robot begins the trial being oriented to 90°. Since no landmarks were visible from there, neurons in the place cell layer of the model respond according to kinesthetic information only. The current ensemble activity of the layer is searched within the nodes in the map and found in Actor unit 270° of node 7. Although the robot's current direction is not 270°, both activity patterns are similar since none of them encodes for visual information. Thus, node 7 is activated in the map, indicating in this way that the robot recognizes location D2 as previously visited. Then, the robot moves forward and perceives some part of landmark 2

(L2). This time, the current activity pattern of the collection of place cells is not similar to any of those found in the map. This is because the robot had previously visited that location being oriented only to 270° and not to 90°, thus visual information sensed in both directions is different. As a result, the current activity pattern is stored in a new Actor unit 90°, which is associated with the new number 17 node. When the robot reaches the choice point of the corridor, the current ensemble activity of place cells is found within Actor unit 90° of node 5, thus this node is activated. At this location, the robot chooses to turn left by curiosity since reward expectation values are not significant. The current activity pattern of the collection of place cells is stored in a new Actor unit 180°, which is associated with node 5. Being oriented to 180°, the robot then moves forward not recognizing the places represented by node 4 in the map since they were previously visited only in direction 0°, thus node 18 is created. When reaching the choice location at the center of the maze, the current ensemble activity of place cells is found within Actor unit 180° of node 3, thus this node is activated and the link from node 18 to node 3 is added. At this location, the expectation of future reward in direction 180° was increased during training in an important manner; therefore, the robot exploits this information when decides to go ahead instead of turning left. From this location, the robot moves forward recognizing the places represented by nodes 8 and 9 in the map. At this point, the expectation of future reward in direction 90° was increased during training in an important manner; hence, the robot exploits this information when decides to turn right instead of turning left. From this location, the robot moves forward recognizing the places represented by node 14 in the map until it reaches the target, whose location is represented by node15.

As Morris (1981) and Eichenbaum et al. (1990) did with normal rats, we could confirm that to find a "hidden" goal in the maze independently of the starting location, the robot requires to exploit its cognitive map by recognizing reward expectations that make it navigate towards the target location successfully.

4.3 On-line Map Adaptation in Multiple T-mazes

In order to evaluate the on-line adaptation of the cognitive map carried out by the model when the environment is altered, we tested two different scenarios of the experiment performed in the maze formed by multiple Ts (see Fig. 10(a)) when landmark configuration is modified after having trained the robot to find the goal. Specifically, in (Barrera & Weitzenfeld, 2007b), we describe results from removing and interchanging landmarks in the maze.

<u>Removing one landmark</u>

The training phase of the task proceeds as in the original experiment just described in Section 4.2. After the robot learns to find the goal departing from the fixed location TD, there are 12 testing trials where one particular landmark was removed (L1, L2 or L3) and the robot departs from locations D1, D2, D3 and TD.

Considering tests from locations D2 and TD when landmark L1 was removed, Fig. 11(a) shows trajectories followed by one of the robots, whereas Fig. 11(b) depicts the map built during training and adapted during both tests.

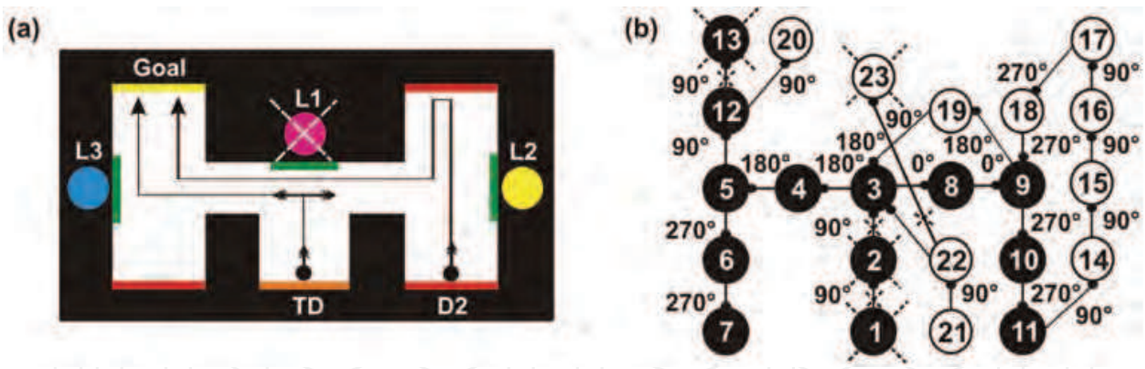

Fig. 11. (a) Robot trajectories traveled during tests departing from locations D2 and TD in direction 90° when landmark L1 was removed. (b) The map built and adapted by the robot during the experiment. Black nodes were created during training, whereas white nodes were added during both tests. Nodes are numbered in order of creation. Nodes and arcs marked as "X" were eliminated.

When starting at location D2, node 10 and 9 are not recognized in direction 90° since corresponding locations in that corridor were previously explored in direction 270°, hence the robot adds new nodes 14 and 15. Being at location represented by new node 15 oriented to 90°, the robot is curious for executing any possible rotation, thus it chooses to go ahead until the end of the corridor while creating nodes 16 and 17. While returning in direction 270°, the robot does not recognize node 16 thus adding node 18. Then, the robot recognizes node 9 and defines link 18-19 in direction 270°. At this time, the robot decides to turn right orienting itself to 180° and moves forward in that direction creating node 19 since node 8 was previously explored in direction 0°. When reaching the choice location at the center of the horizontal corridor in direction 180°, the robot still recognizes node 3 in spite of the absence of L1 because the visual pattern stored in its Actor unit 180° includes mainly information about L3 and the similarity degree between this visual pattern and the current one results close to 1. Exploiting the reward expectation value associated to Actor unit 180° of node 3, the robot goes ahead recognizing nodes 4 and 5, and exploits the reward expectation value associated to direction 90° while deciding to turn right at node 5. When reaching the goal location, the current visual information pattern is different from the one stored in Actor unit 90° of node 13 due to the absence of L1, thus the robot creates the new number 20 node, defines the new link between nodes 12 and 20, and removes connection 12-13 and node 13 since it is not feasible to reach two different places in the same direction 90° from the location represented by node 12 in the map.

On the other hand, during the test starting at location TD in direction 90°, the robot does not recognize map nodes 1, 2 and 3 due to the absence of L1, thus new nodes 21, 22 and 23 are added and connected in direction 90°. Being at the central choice location of the horizontal corridor now represented by new node 23, the curiosity for exploring the place in any orientation makes the robot turn right orienting itself to 0°. At this point, the robot is able to recognize that the current visual information pattern is quite similar to the one stored in Actor unit 0° of node 3 since the position of L2 has remained constant. The activation of node 3 while node 23 is also active involves the redundant representation of the same physical location, and hence the integration of both nodes. As node 23 has just been created, there is no need to add a new node in the map. Actor unit 90° of node 23 is assigned to node

3, arc 90° pointing to node 23 is updated to point to node 3, and node 23 is eliminated as well as connection 2-3, node 2, connection 1-2 and node 1. After the integration process of redundant nodes, the robot orients itself to 180° at the location represented by node 3 due to the large reward expectation value associated to its Actor unit 180°. From node 3, the robot follows directly the correct route to the goal exploiting reward expectation values of recognized map nodes 4, 5 and 12.

Interchanging two landmarks

After performing the training phase of the task as in the original experiment, the robot was tested during 12 trials departing from locations D1, D2, D3 and TD where landmarks L2 and L3 were interchanged. Fig. 12(a) shows trajectories followed by one of the robots during the test starting at D2, whereas Fig. 12(b) depicts the map built during training and adapted during the test.

Due to the new landmarks configuration, the robot does not recognize the previously explored nodes 5 (in direction 90°), 3 (in directions 180° and 90°), 8 (in direction 180°), and 9 (in direction 180°), thus it creates new nodes 17, 19, 22 and 23. In particular, relevant decision locations now corresponding to nodes 19 and 23 are not associated with reward expectations; therefore, robot motor actions are determined by the curiosity and randomness factors during the complete testing trial. Finally, the robot recognizes the goal location represented by node 11 from where L1 is perceptible and has remained in the same location during the entire experiment.

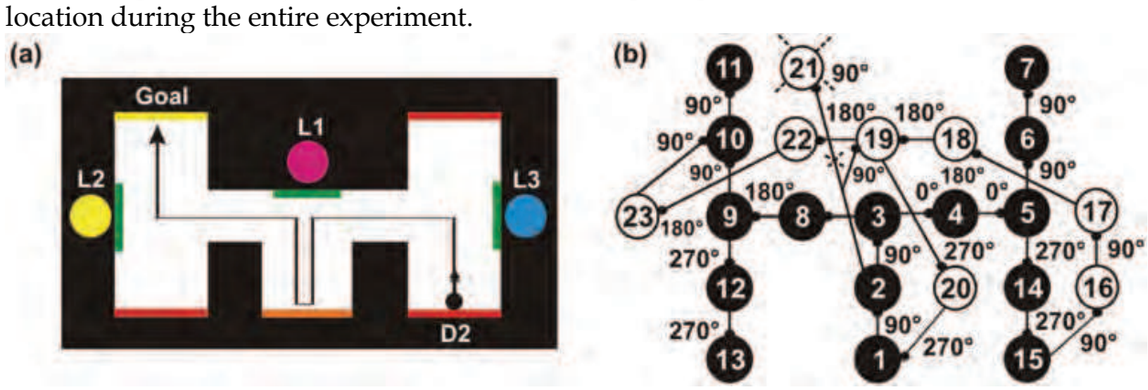

Fig. 12. (a) Robot trajectories traveled during test departing from location D2 in direction 90° when landmarks L2 and L3 were interchanged. (b) The map built and adapted by the robot during the experiment. Black nodes were created during training, whereas white nodes were added during the test. Nodes are numbered in order of creation. Nodes and arcs marked as "X" were eliminated.

5. Conclusions and Discussion

This chapter has presented a computational system-level model of rat spatial cognition whose performance has been demonstrated through mobile robot experimentation during diverse learning and memory tasks within different a priori unknown environments.

Discrimination experiments carried out in the T-maze and in the 8-arm radial maze showed that robots, after appropriate training, are not only able to learn the target location and the route to reach it directly from a fixed starting point, but also to "unlearn" both, the previously correct target location and the corresponding route towards it, when the target is moved to another location in the maze. During this "unlearning" behavior, the robot adapts

the cognitive map built during training by decreasing the expectation of future reward for the previously correct hypothesis while increasing the reward expectation for the new hypothesis until meeting the criterion of only traveling the appropriate route towards the goal.

We reported the behavioral similitude between results from robots and those obtained by O'Keefe (1983) with normal rats performing the classical reversal task. As stated within the manuscript, both subjects showed an abrupt shift in their response from turning left to turning right in the T-maze revealing the moment when their average orientation crosses the midline in the radial maze from the original quadrant to the reversed one.

In the spatial task performed in the maze composed of multiple Ts, allocentric information derived from landmarks was used by robots firstly to build the cognitive map and then to recognize locations previously explored. During training in the task, robots learned the route leading to the target location by reinforcing motor actions executed at every place belonging to that route. In locating the goal from any given departure location, robots navigated until recognizing a place belonging to the previously learnt route, and exploited properly the information stored in Actor units to reach the goal location by executing rotations associated with maximum reward expectations.

Even though all tested robots found the goal successfully from all starting positions used, we should point out that two of them took longer paths from locations D2 and D3 since they did not explore the environment exhaustively during training. Specifically, one of these robots did not explore the route from location TD to D2, thus not recognizing the decision location at the center of the corridor when departing from D3 oriented to 270°. The opposite occurs in the case of the other robot. It did not explore the route from location TD to D3, thus not recognizing the decision location at the center of the corridor when departing from D2 oriented to 90°. Therefore, an efficient performance during tests would be obtained by extending the number of training trials enabling robots to visit all possible direct routes from the fixed departure location to the goal or the end of each corridor.

An appropriate comparison between our results and those reported by Morris (1981) and Eichenbaum et al. (1990) with normal rats turns out difficult since we experimented with a land-based maze having corridors instead of an open field arena. Nevertheless, as both authors did with rats, we could confirm that to find a hidden goal in the maze independently of the starting location, robots require to exploit their cognitive map by recognizing reward expectations that make them navigate towards the target location successfully.

By means of testing scenarios where landmarks are removed or interchanged in the maze formed by multiple Ts after having trained the robot to find the goal, it was feasible to evaluate the robot's ability to adapt, in real-time, the map previously built during training by adding and/or eliminating nodes in order to reveal changes in the physical configuration of the maze.

As described within the manuscript, variations implemented in the landmarks configuration of the maze may alter robot place recognition and successful navigation to the goal, rendering the generation of a partially new cognitive map. Repeating the training process in such modified environment would enable the robot to relearn appropriate reward expectations and store them in the new map.

From a robotics perspective, this research constitutes one attempt to develop an integrated computational system addressing, from a biological approach, not only the spatial mapping

problem, but also the problems of map adaptation and exploitation during successful navigation.

From a biological perspective, our aim is to provide the computational 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 hypotheses in the robotic model and obtain results in a few hours instead of spending weeks experimenting directly with animals. Nevertheless, to achieve this objective, our model needs to be sophisticated by incorporating aspects such as the function of head-direction cells (Ranck, 1984) providing information about the rats' rotation magnitude and movement direction, and the functional differences between hippocampal substructures CA1, CA3 and DG relative to their capabilities of pattern completion and pattern separation (Leutgeb & Leutgeb, 2007). Extending the model in these directions will enable robots to navigate in different spatial contexts (i.e., mazes with corridors as well as open field environments), and improve its adaptive capability in conditions where environmental changes promote new navigational behaviors.

6. References

Arleo, A., Smeraldi, F., and Gerstner, W. (2004). Cognitive navigation based on nonuniform Gabor space sampling, unsupervised growing networks, and reinforcement learning. *IEEE Transactions on Neural Networks* 15 (3): 639-652.

Barrera, A., Weitzenfeld, A. (2007a). Bio-inspired Model of Robot Spatial Cognition: Topological Place Recognition and Target Learning, *Proceedings of the 7th IEEE International Symposium on Computational Intelligence in Robotics and Automation – CIRA*, Jacksonville, Florida, USA.

Barrera, A., Weitzenfeld, A. (2007b). Rat-inspired Model of Robot Target Learning and Place Recognition, *Proceedings of the 15th Mediterranean Conference on Control and Automation – MED*, Athens, Greece.

Barrera, A., Weitzenfeld, A. (2008). Biologically-inspired Robot Spatial Cognition based on Rat Neurophysiological Studies. *Autonomous Robots*, Vol 25, No. 1-2, pp. 147-169.

Barto, A. G. (1995). Adaptive critics and the basal ganglia, In: *Models of information processing in the basal ganglia*, Houk, J. C., Davis, J. L., Beiser, D. G. (Eds.), pp. 215-232, MIT Press, Cambridge, MA.

Bosse, M., Newman, P., Leonard, J., Teller, S. (2004). SLAM in large-scale cyclic environments using the Atlas Framework. *International Journal on Robotics Research* – 23(12):1113–1139.

Brown, M. A., Sharp, P. E. (1995). 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.

Burgess, N., Recce, M., and O'Keefe, J. (1994). A model of hippocampal function. *Neural Networks* 7 (6-7): 1065-1081.

Cho, J., and Sharp, P. (2001). Head direction, place, and movement correlates for cells in the rat retrosplenial cortex. *Behavioral Neuroscience* 115 (1): 3-25.

Eichenbaum, H., Stewart, C., Morris, R. G. M. (1990). Hippocampal representation in place learning. The Journal of Neuroscience 10(11): 3531-3542.

Etienne, A., Jeffery, K. (2004). Path integration in mammals. *Hippocampus* 14(2): 180-192.

Foster, D. J., Morris, R. G. M., Dayan, P. (2000). A Model of Hippocampally Dependent Navigation, Using the Temporal Difference Learning Rule. *Hippocampus* 10: 1-16.

Franz, M. O., Schölkopf, B., Mallot, H. A., Bülthoff, H. (1998). Learning view graphs for robot navigation. *Autonomous Robots* – 5: 111-125.

Frese, U. (2006). A discussion of Simultaneous Localization and Mapping. *Autonomous Robots* – 20: 25-42.

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, pp. 15-28.

Granon, S., and Poucet, B. (2000). Involvement of the rat prefrontal cortex in cognitive functions: A central role for the prelimbic area. *Psychobiology* 28 (2): 229-237.

Guazzelli, A., Corbacho, F. J., Bota, M. and Arbib, M. A. (1998). Affordances, motivation, and the world graph theory. *Adaptive Behavior* 6 (3-4): 435-471.

Guivant, J., Nebot, E., Nieto, J., Masson, F. (2004). Navigation and mapping in large unstructured environments. *International Journal of Robotics Research* – 23(4): 449-472.

Hähnel, D., Burgard, W., Wegbreit, B. and Thrun, S. (2003). Towards lazy data association in SLAM, *Proceedings of the 11th International Symposium of Robotics Research (ISRR)*, Sienna, Italy.

Hartley, T., Burgess, N. (2005). Complementary memory systems: competition, cooperation and compensation. *Trends in Neuroscience* 28(4):169-170.

Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*, Wiley-Interscience, New York.

Hollup, S. A., Molden, S., Donnett, J. G., Moser, M. and Moser, E. I. (2001). Place fields of rat hippocampal pyramidal cells and spatial learning in the watermaze. *European Journal of Neuroscience* 13: 1197-1208.

Houk, J. C., Adams, J. L., and Barto, A. G. (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement, In: *Models of information processing in the basal ganglia*, Houk, J. C., Davis, J. L., Beiser, D. G. (Eds.), pp. 249-270, MIT Press, Cambridge, MA.

Hull, C. L. (1932). The goal gradient hypothesis and maze learning. *Psychological Review* 39: 25-43.

Jeffery, K. J., and O'Keefe, J. M. (1999). Learned interaction of visual and idiothetic cues in the control of place field orientation. *Experimental Brain Research* 127: 151-161.

Kelley, A. (2004). Ventral striatal control of appetitive motivation: role in ingestive behavior and reward-related learning. *Neuroscience and Biobehavioral Reviews* 27 (8): 765-776.

Kuipers, B., Modayil, J., Beeson, P., MacMahon, M., Savelli, F. (2004). Local metrical and global topological maps in the Hybrid Spatial Semantic Hierarchy, *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, New Orleans, USA.

Leutgeb, S., Leutgeb, J.K. (2007). Pattern separation, pattern completion, and new neuronal codes within a continuous CA3 map. *Learning and Memory* 14(11):745-57.

McNaughton B. L., Knierim J. J., and Wilson, M. A. (1994a). Vector encoding and the vestibular foundations of spatial cognition, In: *The cognitive neurosciences*, Gazzaniga, M. (Ed.), pp. 585-595, MIT Press, Boston.

McNaughton, B., Mizumori, S., Barnes, C., Leonard, B., Marquis, M., and Green, E. (1994b). Cortical representation of motion during unrestrained spatial navigation in the rat. *Cerebral Cortex* 4: 27-39.

Milford, M., Wyeth, G. (2007). Spatial mapping and map exploitation: a bio-inspired engineering perspective, In: *Spatial Information Theory*, Winter, S., Duckham, M., Kulik, L. and Kuipers, B. (Eds.), pp. 203-221, Springer-Verlag, Germany.

Mittelstaedt, M., Mittelstaedt, H. (1982). Homing by path integration in a mammal, In: *Avian Navigation*, Papi, F. and Wallraff, H. G. (Eds.), pp. 290-297, Springer Verlag, Berlin.

Morris, R. G. M. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation* 12: 239 – 260.

Moser, E. I., Kropff, E., Moser, M-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annu. Rev. Neurosci.* 31: 69-89.

Movarec, H. P., Elfes, A. (1985). High resolution maps from wide angle sonar, *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 116-121.

Mulder, A. B., Tabuchi, E., Wiener, S. I. (2004). Neurons in hippocampal afferent zones of rat striatum parse routes into multi-pace segments during maze navigation. *European Journal of Neuroscience*, Vol. 19, pp. 1923-1932.

O'Keefe, J. (1983). Spatial memory within and without the hippocampal system, In: *Neurobiology of the Hippocampus*, Seifert, W. (Ed.), pp. 375-403, Academic Press, NY.

O'Keefe, J., Conway, D. H. (1978). Hippocampal place units in the freely moving rat: why they fire where they fire. *Experimental Brain Research* 31: 573–590.

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): 171 – 175.

O'Keefe, J., Nadel, L. (1978). *The hippocampus as a cognitive map*, Oxford University Press.

Parron, C., Save, E. (2004). Evidence for entorhinal and parietal cortices involvement in path integration in the rat. *Experimental Brain Research* 159 (3): 349-359.

Poucet, B. (1993). Spatial cognitive maps in animals: new hypotheses on their structure and neural mechanisms. *Psychological Review* 100 (2): 163–182.

Quirk, G. J., Muller, R. U., Kubie, J. L. (1990). The firing of hippocampal place cells in the dark depends on the rat's recent experience. *Journal of Neuroscience* 10(6): 2008-2017.

Ranck, J. B., Jr. (1984). Head-direction cells in the deep layers of dorsal presubiculum in freely moving rats. *Soc. Neurosci. Abstr.* 10: 599.

Redish, A., Touretzky, D. (1997). Cognitive maps beyond the hippocampus. *Hippocampus* 7(1): 15-35.

Risold, P., Thompson, R., and Swanson, L. (1997). The structural organization of connections between hypothalamus and cerebral cortex. *Brain Research Reviews* 24 (2-3): 197-254.

Roberts, W. A. (1998). *Principles of animal cognition*, McGraw Hill, USA, pp. 201 – 230.

Schultz, W., Tremblay, L., and Hollerman, J. (1998). Reward prediction in primate basal ganglia and frontal cortex. *Neuropharmacology* 37 (4-5): 421-429.

Stentz, A., Hebert, M. (1995). A complete navigation system for goal acquisition in unknown environments. *Autonomous Robots* 2(2): 127-145.

Thrun, S., et al. (2000). Probabilistic algorithms and the interactive museum tour-guide robot Minerva. *The International Journal of Robotics Research* 19(11): 972-999.

Tolman, E. (1948). Cognitive maps in rats and men. *Psychological Review* 55: 189-208.

Weitzenfeld, A., Arbib, M., Alexander, A. (2002). Neural Simulation Language, MIT Press.

Zivkovic, Z., Bakker, B., Kröse, B. (2005). Hierarchical map building using visual landmarks and geometric constraints, *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7-12, Edmonton, Canada.

Mobile Robots Navigation

Edited by Alejandra Barrera

ISBN 978-953-307-076-6
Hard cover, 666 pages
Publisher InTech
Published online 01, March, 2010
Published in print edition March, 2010

Mobile robots navigation includes different interrelated activities: (i) perception, as obtaining and interpreting sensory information; (ii) exploration, as the strategy that guides the robot to select the next direction to go; (iii) mapping, involving the construction of a spatial representation by using the sensory information perceived; (iv) localization, as the strategy to estimate the robot position within the spatial map; (v) path planning, as the strategy to find a path towards a goal location being optimal or not; and (vi) path execution, where motor actions are determined and adapted to environmental changes. The book addresses those activities by integrating results from the research work of several authors all over the world. Research cases are documented in 32 chapters organized within 7 categories next described.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Alejandra Barrera (2010). Robot Topological Mapping and Goal-Oriented Navigation Based on Rat Spatial Cognition, Mobile Robots Navigation, Alejandra Barrera (Ed.), ISBN: 978-953-307-076-6, InTech, Available from: http://www.intechopen.com/books/mobile-robots-navigation/robot-topological-mapping-and-goal-oriented-navigation-based-on-rat-spatial-cognition

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821