Biol Cybern (2010) 103:299–317
DOI 10.1007/s00422-010-0400-z

ORIGINAL PAPER

Biological
Cybernetics

Path planning versus cue responding: a bio-inspired model of switching between navigation strategies

Laurent Dollé · Denis Sheynikhovich · Benoît Girard · Ricardo Chavarriaga · Agnès Guillot

Received: 8 January 2010 / Accepted: 21 June 2010 / Published online: 9 July 2010

Abstract In this article, we describe a new computational model of switching between path-planning and cue-guided navigation strategies. It is based on three main assumptions: (i) the strategies are mediated by separate memory systems that learn independently and in parallel; (ii) the learning algorithms are different in the two memory systems—the cue-guided strategy uses a temporal-difference (TD) learning rule to approach a visible goal, whereas the path-planning strategy relies on a place-cell-based graph-search algorithm to learn the location of a hidden goal; (iii) a strategy selection mechanism uses TD-learning rule to choose the most successful strategy based on past experience. We propose a novel criterion for strategy selection based on the directions of goal-oriented movements suggested by the different strategies. We show that the selection criterion based on this "common currency" is capable of choosing the best among TD-learning and planning strategies and can be used to solve navigational tasks in continuous state and action spaces. The model has been successfully applied to reproduce rat behavior in two water-maze tasks in which the two strategies were shown to interact. The model was used to analyze competitive and cooperative interactions between different strategies during these tasks as well as relative influence of different types of sensory cues.

Keywords Computational model · Spatial navigation · Strategy switch · Parallel memory systems · Action selection

1 Introduction

An increasing number of behavioral research studies focus on the capacity of animals to switch between different navigation strategies when it is required by the environmental circumstances (see Franz and Mallot 2000; White 2004; Arleo and Rondi-Reig 2007; Khamassi 2007, for reviews). The majority of these articles explore the interactions between response- and place-based strategies (Packard and McGaugh 1996; Devan and White 1999; Roberts and Pearce 1999; Gibson and Shettleworth 2005; Rich and Shapiro 2009). Response-based strategies are thought to learn associations between sensory cues and actions linked with reward, whereas place-based strategies use a form of spatial representation to store the goal position and plan a path to it. Experimental evidence in support of such a separation between navigational strategies comes from lesion's studies that gave rise to the theory of parallel memory systems in the brain of the rat (Packard et al. 1989; McDonald and White 1993; Devan and White 1999; Kim and Baxter 2001; White and McDonald 2002; White 2004; Burgess 2008). According to this theory, the dorsolateral striatum (DLS) is involved in the control of response-based strategies by means of a slow and inflexible "trial and error" learning, whereas place-based strategies are mediated by the hippocampus (Hc) and other neural structures to which it projects, such as prefrontal

Laurent Dollé, Denis Sheynikhovich—First authorship shared.

L. Dollé (✉) · D. Sheynikhovich · B. Girard · A. Guillot
Institut des Systèmes Intelligents et de Robotique, UPMC CNRS
UMR 7222, 4 Place Jussieu, 75252 Paris Cedex 05, France
e-mail: laurent.dolle@upmc.fr

Present Address:
D. Sheynikhovich
Neurobiology of Adaptive Processes UPMC,
CNRS UMR 7102,
9 quai St. Bernard, 75005 Paris, France

R. Chavarriaga
Defitech chair on Non-Invasive Brain-Computer Interface (CNBI),
Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne,
Switzerland

cortex (PFC) (Mizumori 2008; Jankowski et al. 2009; White 2009). Learning in the Hc-dependent pathway is considered to be rapid and flexible (Granon and Poucet 1995; Yin and Knowlton 2004; Grahn et al. 2008).

The existence of two (or more) parallel memory systems mediating different behavioral strategies raises a question of when one or other strategy takes control over behavior. Experimental evidence suggests that different memory systems favor separate sets of sensory cues: DLS-mediated system mostly uses proximal cues (e.g., visible platform in the Morris Water Maze, or intra-maze landmark signaling the platform position), whereas Hc-mediated system encodes configurations of distal cues (like extra-maze landmarks and environmental boundaries) (McDonald et al. 2004; Hartley and Burgess 2005; Doeller and Burgess 2008; Doeller et al. 2008; Leising and Blaisdell 2009; Blaisdell 2009; Pearce 2009). Distal cues and environmental boundaries can be used to form a spatial representation encoded in the activities of location selective neurons (termed "Place Cells") residing in the Hc (O'Keefe and Dostrovsky 1971; O'Keefe and Nadel 1978; Redish 1999; Save and Poucet 2000; Kelly and Gibson 2007). The question raised by these studies is how different types of sensory cues influence ongoing behavior, including strategy selection.

Interactions between multiple navigation strategies when two or more of them can be used at the same time is often analyzed in terms of competition and cooperation. *Competition* between two memory systems (and hence, the corresponding strategies) is demonstrated when a lesion of one of the systems entails an improvement of the learning of the other, while *cooperation* implies that such a lesion leads to the impairment of the other system's performance (Kim and Baxter 2001; Gold 2004). In the spatial domain, competition or cooperation between navigational strategies are respectively observed when one of the strategies perturbs (Packard and McGaugh 1992; Pearce et al. 1998; Chang and Gold 2003; Canal et al. 2005) or facilitates (McDonald and White 1994; Hamilton et al. 2004; Voermans et al. 2004) the other one for reaching the goal. The analysis of switching between place- and response-based strategies suggests that they can interact both across and within experimental trials (Pearce et al. 1998; Devan and White 1999). Moreover, depending on the training protocol, the strategies can be switched immediately after the appearance or disappearance of relevant sensory cues (Devan and White 1999), or learned progressively across trials to prefer one type of cues over another (Pearce et al. 1998). In summary, although these and other behavioral and lesion's studies provide valuable information concerning the influence of sensory cues on behavior and the types of interactions between strategies, the mechanism of the strategy selection is not clear.

In this article, we propose a bio-inspired computational model of selection between response- and place-based strategies applied for navigation in continuous space. This model is based on three key assumptions. The first one is that these strategies are mediated by separate memory systems that can learn independently and in parallel (as in the computational models of, e.g., Guazzelli et al. 1998; Girard et al. 2005; Chavarriaga et al. 2005; Daw et al. 2005). The second assumption is that learning algorithms within the two memory systems are of different types: while response-based strategy relies on a slow and stereotyped "trial-and-error" learning implemented as a temporal-difference (TD) learning procedure, learning in the place-based strategy is fast and flexible and is based on a graph-search algorithm for finding a goal (as in Guazzelli et al. 1998; Girard et al. 2005; Daw et al. 2005). The third assumption is that the selection mechanism is not fixed but continuously updates its estimates of the relative "goodness" of different strategies (as in Chavarriaga et al. 2005; Daw et al. 2005). The novelty of our approach is in the proposed "common currency" allowing the comparison of strategies that use different learning algorithms for reaching the goal. This common currency is defined as the direction of the goal-oriented movement proposed by each strategy. We show below that the selection criterion based on this common currency, is capable of choosing the best among TD-learning and planning strategies and can be used to solve navigational tasks in continuous state and action spaces.

We use our model to reproduce and analyze rat behavior in two experimental protocols in which response- and place-based strategies were shown to interact with each other (Pearce et al. 1998; Devan and White 1999) with the aim of answering the following questions: (i) what is the mechanism of strategy selection that can result in competition and cooperation between strategies across and within experimental trials? (ii) What is the possible selection criterion, i.e., how can the performance of different strategies (with potentially different learning mechanisms) be compared so that the best strategy is chosen to take control over behavior? and (iii) How different types of sensory cues influence strategy selection? The rest of the article is structured as follows: Section 2 describes the model of strategy selection; Sections 3 and 4 describe the results of computer simulations aimed at reproducing animal data; in Sect. 5, we discuss the results of this study in relation to the previous questions and to other available experimental and theoretical studies; Finally, we conclude in Sect. 6 with the outlook on future study.

2 The model

In the model of navigation under this study, response- and place-based strategies are implemented by two "experts", referred to as *Taxon expert* and *Planning expert* in this article. They represent DLS and Hc–PFC memory systems,

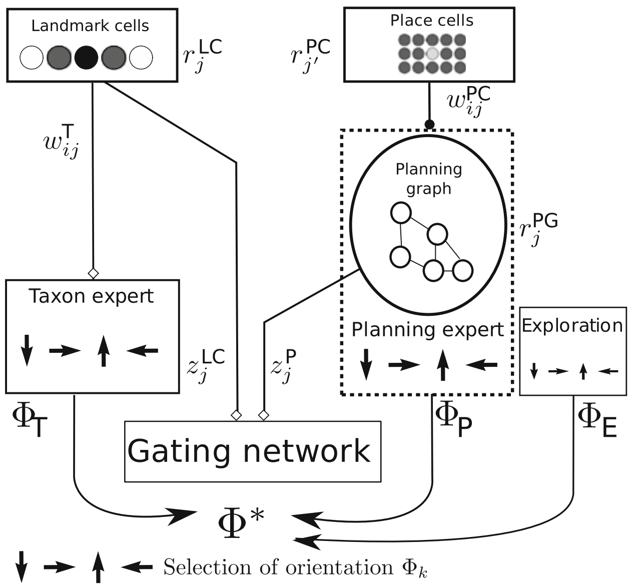

Fig. 1 Model overview (see text for details). *LC* Landmark Cells, *PC* Place Cells, *PG* Planning Graph, *T* Taxon expert, *P* Planning expert, *E* Exploration expert. Φ^* is the direction of the next movement resulting from the selection process

respectively. During navigation, these experts propose a direction for the next movement according to either visual input (Taxon expert) or the estimated location (Planning expert). In addition, the third, *Exploration expert*, proposes a direction of movement randomly chosen between 0 and 2π. The actual movement, performed by the simulated rat (henceforth referred to as "animat"), is determined by the selection module (the gating network) which selects one of the experts to take control over behavior on the basis of previous performance (Fig. 1).

2.1 Taxon expert

The Taxon expert implements response-based strategy in the model. In particular, we consider two kinds of response-based strategies: approaching a visible target (sometimes referred as "beacon learning") and approaching a hidden target marked by a landmark located on a certain distance from it (i.e., "guidance" in terms of O'Keefe and Nadel (1978)). Information about the landmark (or the visible target) is encoded by the activities of N_{LC} Landmark Cells (see Table 1 for parameter values) which code the presence or the absence of the landmark in a particular direction $\phi_i^{\mathrm{LC}} = \frac{2\pi i}{N_{\mathrm{LC}}}$. The activity of LC i is given by:

$$r_i^{\mathrm{LC}} = \exp\left(-\frac{\Delta\Phi_i}{2(\sigma_{\mathrm{LC}}/\Delta_{\mathrm{R}\to\mathrm{L}})^2}\right), \qquad (1)$$

where $\Delta\Phi_i = \Phi^{\mathrm{L}} - \phi_i^{\mathrm{LC}}$ is the angular distance between the direction of the landmark Φ^{L} and the cell's preferred direction, and $\Delta_{\mathrm{R}\to\mathrm{L}}$ is the distance from the animat to the landmark in centimeters (see in, e.g., Brown and Sharp (1995), Touretzky and Redish (1996), for similar modeling of sensory input). The width of the Gaussian centered at the landmark direction increases as the animat approaches the landmark, expressing the fact that the landmark image takes up a larger part of the view field if the animat is close to the landmark.

In the model of our study, the Taxon expert can work in either allocentric or egocentric directional reference frames. The allocentric reference frame is fixed with respect to distal (room) cues and is assumed to be supported by the head direction network involving the anterodorsal nucleus of thalamus (Taube et al. 1990). In this reference frame, the direction to the landmark Φ^{L} is given with respect to the zero direction that is defined at the first entry to the environment (see Fig. 2) and remains fixed thereafter. In the second, egocentric reference frame, Φ^{L} is given relative to the zero direction that coincides with the current gaze direction of the animat.

The motor response of the Taxon expert to the landmark stimulus is encoded by $N_{\mathrm{AC}} = 36$ *Action Cells* (AC), so that each AC i receives input from all LCs and codes for movement direction $\phi_i^{\mathrm{T}} = \frac{2\pi i}{N_{\mathrm{AC}}}$ in the corresponding reference frame. Its activity represents the value of moving in the corresponding direction and is computed as follows (note that superscript T in the following text denotes Taxon expert and not matrix transposition):

$$a_i^{\mathrm{T}}(t) = \sum_{j=1}^{N_{\mathrm{LC}}} r_j^{\mathrm{LC}}(t) w_{ij}^{\mathrm{T}}(t). \qquad (2)$$

The activity in the AC population is interpreted as a population code for the continuous direction Φ^{T} of the next movement of the animat, proposed by the Taxon expert (Strösslin et al. 2005; Chavarriaga et al. 2005):

$$\Phi^{\mathrm{T}}(t) = \arctan\left(\frac{\sum_i a_i^{\mathrm{T}}(t)\sin(\phi_i^{\mathrm{T}})}{\sum_i a_i^{\mathrm{T}}(t)\cos(\phi_i^{\mathrm{T}})}\right). \qquad (3)$$

Learning of the weights is performed by the TD-based Q-learning algorithm (Sutton and Barto 1998). We consider the activity $a_i^{\mathrm{T}}(t)$ of an AC i to be the Q-value of the corresponding state–action pair, giving rise to the following formula for the weight update (Strösslin et al. 2005; Chavarriaga et al. 2005):

$$\Delta w_{ij}^{\mathrm{T}} = \eta\delta^{\mathrm{T}}(t)e_{ij}^{\mathrm{T}}. \qquad (4)$$

where η is the learning rate, $\delta^{\mathrm{T}}(t)$ is the reward prediction error, and e_{ij}^{T} is the eligibility trace. The reward-prediction error is defined as the difference between the current and previous estimates of the discounted future reward (Sutton and Barto 1998):

$$\delta^{\mathrm{T}}(t) = R(t+1) + \gamma \max_a a_i^{\mathrm{T}}(t+1) - a^{\mathrm{T}}(t), \qquad (5)$$

Table 1 Parameters of the experts

Name	Value	Description
Taxon expert and gating network		
N_{LC}[1]	100	Number of Landmark Cells
σ_{LC}[1]	27.5°	Normalized landmark width
$N_{\mathrm{AC}}^{\mathrm{T}}$[2]	36	Number of action cells
σ^{T}[2]	22.5°	Standard deviation of the generalization profile
η[3]	0.001	Learning rate
λ[2]	0.76	Eligibility trace decay factor
γ[2]	0.8	Future reward discount factor
ξ[2]	0.01 / 0.05	Learning rate of the gating network (depending on the experiment)
Planning expert		
θ^{PC}[3]	0.3	Activity threshold for place-cells node linking
θ^{P}[3]	0.3	Activity threshold for node creation
α[3]	0.7	Decay factor of the goal value
N_{PC}[4]	1681	Number of simulated Place Cells
σ_{PC}[4]	10 cm	Place field size

[1] Set to give sufficient detailed representation; [2] adapted from Chavarriaga et al. (2005); [3] hand-tuned; [4] set to give a sufficient overlap between place fields

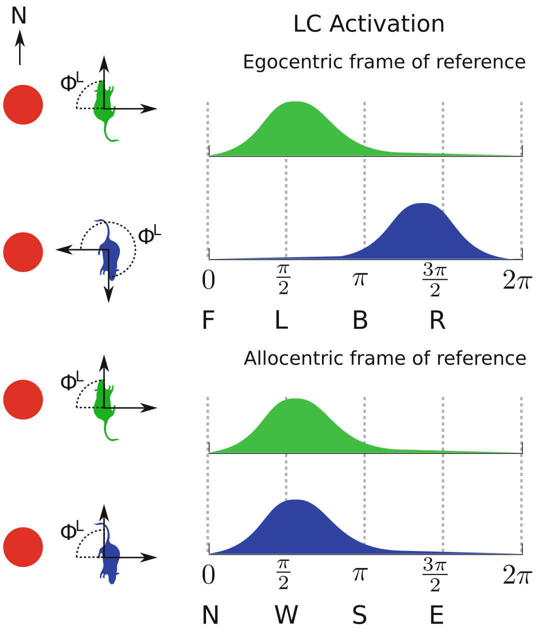

Fig. 2 Internal representation of the landmark (*black dot*) in the egocentric (*top*) and allocentric (*bottom*) spatial reference frames. In the egocentric reference frame, the landmark seen by the animat oriented toward north (marked by *light grey*) or toward south (marked by *dark grey*) will be represented by highly active Landmark Cells at egocentric directions $\Phi^{\mathrm{L}} = \pi/2$ (i.e., on the left side relative to the animat's head direction) and $\Phi^{\mathrm{L}} = 3\pi/2$ (i.e., on the right side), respectively (see Eq. 1). In the allocentric reference frame, the landmark will be represented by highly active cells at the allocentric direction $\Phi^{\mathrm{L}} = \pi/2$ in both cases, since the landmark is located in the western direction from the animat (here, the north direction was chosen as the zero direction of the allocentric reference frame). *F* front, *L* left, *B* back, *R* right; *N* north, *W* west, *S* south, *E* east

where $R(t)$ is the reward delivered at time t, $0 < \gamma < 1$ is the future reward discounting factor, and $a^{\mathrm{T}}(t)$ is the Q-value of the action performed at time t, estimated by the Taxon expert. The eligibility trace e_{ij}^{T} in Eq. 4 speeds up learning by remembering the state–action pairs experienced in the past:

$$e_{ij}^{\mathrm{T}}(t+1) = r_j^{\mathrm{LC}}(t) r_i(t) + \lambda e_{ij}^{\mathrm{T}}(t), \tag{6}$$

where $\lambda < 1$ is the eligibility trace decay rate, $r_j^{\mathrm{LC}}(t)$ is given by Eq. 1 and r_i^{AC} is given by:

$$r_i^{\mathrm{AC}}(t) = -\exp\left(\frac{\phi_i^{\mathrm{T}} - \Phi^{\mathrm{T}}(t)}{2\sigma^{\mathrm{T}2}}\right). \tag{7}$$

This term represents the activity of action cells in the *generalization phase* (Strösslin et al. 2005) and allows the actions which are close to the actually performed action Φ^{T} (Eq. 3) to update their weights in the same direction. The use of a generalization phase for action learning, together with the use of Eq. 3 for action selection results in the ability of the Taxon expert to work in a continuous action space (Strösslin et al. 2005).

We note that the learning algorithm described above does not depend on the spatial reference frame (i.e., allocentric or egocentric, see Fig. 2) that is used. The information about the reference frame is implicitly encoded by the landmark information. However, the learned behavior of the animat in some tasks can be different, depending on what reference frame is used as illustrated in the results (Sect. 3.2.5).

The calculation of the reward-prediction error (Eq. 5) and the corresponding weight update (Eq. 4) are performed on each time step independently from the identity of the expert (i.e., Taxon, Planning or Exploration) that generated the last action. Moreover, reward signal $R(t)$ is shared between all the experts at each time step. Therefore, goal-oriented actions performed under the control of, e.g., the Planning expert, help the Taxon expert to adjust its weights. This way, the *cooperation* between strategies is implemented in the model, in addition to the *competition* between strategies, governed by the selection network (see Sect. 2.3 below for the competitive selection algorithm).

2.2 Planning expert

The Planning expert uses a simple graph-search algorithm to find the shortest path to the goal (Martinet et al. 2008). During an unrewarded *map building* phase, the Planning expert builds a graph-like representation of space based on the activities of simulated *Place Cells*. During a reward-based *goal planning* phase, this representation is used to plan and execute goal-directed path. Since extra-maze cues are stable in the experiments that we will simulate, we use a simple model of Place Cells as described later (see Arleo and Gerstner 2000; Sheynikhovich et al. 2009 for more detailed models of Place Cells that integrate information from distal cues and path integration). The population of Place Cells in our model is created before the learning is started, and the activity of place cell j is given by

$$r_j^{\mathrm{PC}} = \exp\left(-\frac{\Delta^2_{\mathrm{A}\to j}}{2\sigma^2_{\mathrm{PC}}}\right), \tag{8}$$

where $\Delta_{\mathrm{A}\to j}$ is the distance between the animat and the center of firing field of place cell j (i.e., place field), and σ_{PC} is the width of the place field. Place field centers are distributed uniformly in the environment.

Given the Place Cells activity, the *Planning Graph* is built during unrewarded movements by the following algorithm. When a new node N_i is created, it is connected to place cell j with connection weights w_{ij}^{P}:

$$w_{ij}^{\mathrm{P}} = r_j^{\mathrm{PC}}\mathcal{H}\left(r_j^{\mathrm{PC}} - \theta^{\mathrm{PC}}\right), \tag{9}$$

where $\mathcal{H}(x) = 1$ if $x > 0$, $\mathcal{H}(x) = 0$ otherwise. The activity r_i^{P} of node i is then computed by

$$r_i^{\mathrm{P}} = \sum_j r_j^{\mathrm{PC}} w_{ij}^{\mathrm{P}}. \tag{10}$$

A new node is added on each time step unless at least one existing node is active above threshold θ^{P}. The overlap between PCs, threshold values θ^{PC}, and θ^{P} have been chosen to guarantee that, when the condition for the creation of a new node is met (i.e., no node activity above θ^{P}), there is always at least one PC, whose activity is above θ^{PC}. Thus, any newly created graph node has at least one connection weight to the PCs that is non-zero.

A link between nodes N_i and N_j stores the allocentric direction of movement required to pass from one node to the other:

$$\Phi^{\mathrm{P}}(t) = \widehat{\vec{x}\,\overrightarrow{N_i N_j}}, \tag{11}$$

where x is the zero angle of the allocentric reference frame. This link is created only when the animat travels between two nodes, with no intermediary node having already been present. This means that when a new node is created, it is already connected to the node previously visited by the animat. Therefore, a node i will be connected to the node j if and only if there is no node k such that

$$\widehat{\overrightarrow{N_i N_j}\overrightarrow{N_i N_k}} < \epsilon \quad \text{and} \quad \|\overrightarrow{N_i N_k}\| < \|\overrightarrow{N_i N_j}\|, \tag{12}$$

where ϵ is dependent on the moving and rotation speeds of the animat. This insures that graph nodes are only connected to their closest neighbors.

Given the Planning Graph, the optimal path to the goal is determined by the activation–diffusion mechanism (Burnod 1991; Hasselmo 2005), based on the Dijkstra's algorithm for finding the shortest path between two nodes in a graph (Dijkstra 1959). More specifically, during goal planning, the Planning expert first determines its location using a *position value* and then calculates the direction toward the goal using *goal value*. The position value corresponds to the activity r_i^{P} of the node (Eq. 10). The goal value $G_i = 0$ when no goal position is known. In this case, the strategy proposes a random movement direction among the different possible actions from the current node. In contrast, when the goal position is found (using the actions generated by any expert), the goal value of the closest (goal) node is set to $G_{i^*} = 1$, and is propagated to all the adjacent nodes, decreased by a decay factor $\alpha < 1$. The goal value G_i of a node i of distance n from the goal node (measured as the number of nodes between the goal node and the node i) is given by $G_i = \alpha^n$. The next movement direction is given by the link to the adjacent node with the highest goal value.

2.3 Strategy selection

During goal learning, the model has to select out of the three experts, Taxon, Planning, and Exploration experts (T, P, and E, respectively), which one takes control over behavior, i.e., chooses the next action. The gating network learns to select experts on the basis of the "common currency" defined as the direction of movement proposed by each expert. After learning, the expert that proposes directions of movements that are closest to the true direction to the goal is considered the best at each time step. We use only three experts at present, but the selection network can work with any number of experts

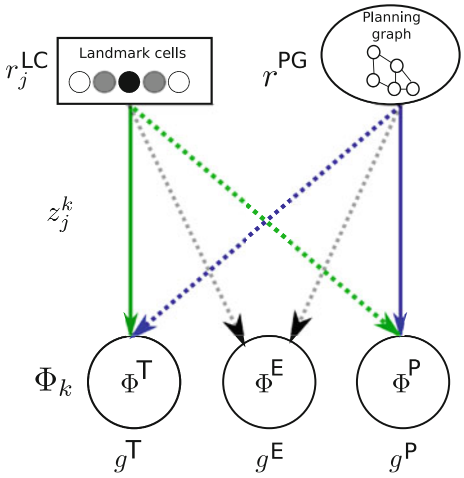

Fig. 3 Gating network. The inputs of the Taxon and Planning experts (LC and PG) are linked to the units in the gating network. The gating values g^k are weighted sums of the input values r_j with weights z_j^k. One of the three experts is selected according to a winner-take-all scheme

as long as they provide a direction of movement toward the goal as their output.

In the present model, the gating network consists of three units $k \in \{\mathrm{T, P, E}\}$, each corresponding to a separate expert. The activity g^k of the unit k is called "gating value" of the corresponding expert. The input to the units in the gating network is provided by the activities of the LC population and the nodes of the Planning Graph (Fig. 3). The gating values g^k are calculated as

$$g^k(t) = \sum_{j=1}^{N_{\mathrm{LC}}} z_j^k(t) r_j^{\mathrm{LC}}(t) + \sum_{j=N_{\mathrm{LC}}+1}^{N_{\mathrm{LC}}+N_{\mathrm{P}}} z_j^k(t) r_j^{\mathrm{P}}(t), \qquad (13)$$

where z_j^k is the connection weight between the unit k of the gating network and input unit j of the experts. As described in the previous sections, at each time step experts propose candidate directions Φ^k of the next movement. The gating values are used to choose the next movement direction Φ^* to be taken by the animat using a winner-take-all scheme:

$$\phi^k(t);\, k = \mathrm{argmax}_i(g^i(t)) \qquad (14)$$

Similar to the learning in the Taxon expert, the connection weights for the Taxon and Planning gating values are randomly initialized between 0 and 0.01 and adjusted using a Q-learning algorithm. The weight update in this case is given by

$$\Delta z_j^k = \xi^{\mathrm{G}} \delta^{\mathrm{G}}(t) e_j^k(t), \qquad (15)$$

where ξ^{G} is the learning rate of the gating network, and $\delta^{\mathrm{G}}(t)$ is the reward-prediction error:

$$\delta^{\mathrm{G}}(t) = R(t+1) + \gamma \max_k \left(g^k(t+1) \right) - g^{k^*}(t), \qquad (16)$$

Here, $R(t)$ is the reward delivered at time t, γ is the future reward discount factor of the gating network, and g^{k^*} is the gating value of the expert, chosen at time step t (i.e., the time step that corresponds to the direction of movement in Eq. 14).

As for the Taxon strategy, the eligibility trace e_j^k of expert k allows the gating network to reinforce the experts selected in the past:

$$e_j^k(t+1) = \Psi(\Phi^*(t) - \Phi^k(t)) r_j^k(t) + \lambda e_j^k(t), \qquad (17)$$

where λ is the eligibility trace decay factor. The term $\Psi(\Phi^*(t) - \Phi^k(t))$ can be considered as a discrete version of the action generalization in the Taxon expert, where

$$\Psi(x) = \exp(-x^2) - \exp(-\pi/2). \qquad (18)$$

This term insures that the closer the orientation is from the selected one, the higher the corresponding strategy will be reinforced ($\Psi(x)$ is maximum when $x = 0$). In contrast, two strategies that proposed two opposite orientations will have opposite reinforcements. The selection between experts is performed at each time step, unless the Exploration expert is chosen, in which case the chosen orientation is taken during three subsequent time steps. This was done to avoid the animat being stuck in a particular location due to random weight initialization. Since exploration actions are pseudo-random, their weight will *decrease* with learning relative to the weight associated with strategies that direct the animat toward the goal (since the gating network assigns higher weights to strategies that maximize reward). This situation does not change when the weights start to converge, since exploration strategy will not predict rewards better at the end of training; its actions remain always pseudo-random.

3 Simulation I: Experiment of Pearce et al. (1998)

In this experiment, two groups of rats (Control and Hippocampal-lesioned) learned to find the location of a hidden platform in a circular water maze. A visible landmark was located in the pool at a certain distance and allocentric direction from the platform. At the start of an experimental session, the platform and the landmark were moved to one of eight predefined locations in the pool (Fig. 4a), and remained fixed for four trials, after which a new session started. The principal observed results of this experiment (see Fig. 3 in their article) consisted in the observations that (i) both the lesioned and intact rats learned to swim to the hidden platforms at the end of training, and (ii) escape latencies of Hc-lesioned rats were significantly shorter than Control rats in the first trials of intermediate sessions, while they were significantly longer than Control rats in the last trials of each session. From these results, the authors concluded that the intact rats used two competing navigation strategies to locate the goal: a

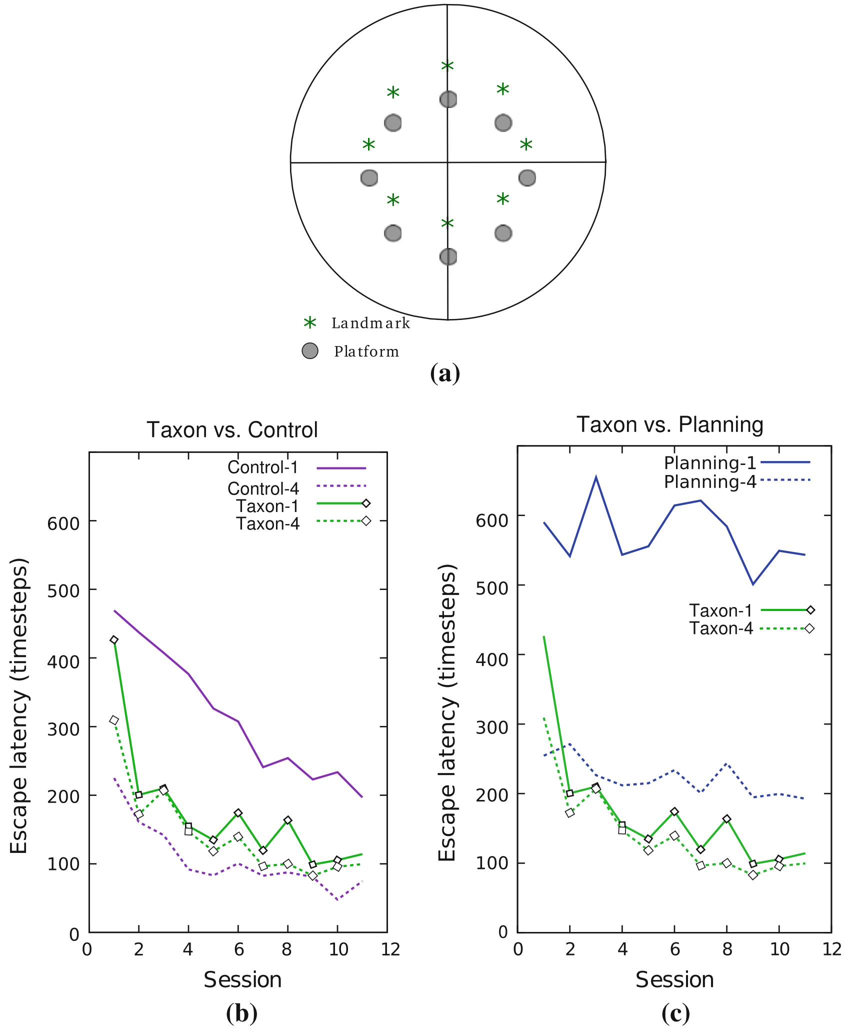

Fig. 4 **a** Experimental setup of Pearce et al. (1998). Mean escape latencies of simulated rats across sessions. **b** Control versus Taxon group. **c** Planning versus Taxon group. *Solid*, and *dotted lines* correspond, respectively to first-trial and last-trial latencies

Hc-dependent strategy that remembered the goal location with respect to distal extra-maze cues; and a Hc-independent strategy (termed "heading vector strategy" by the authors) that remembered the allocentric direction from the landmark to the goal.

3.1 Simulation procedure and data analysis

The simulated water maze, rat, and landmark were represented by circles of 200, 15, and 20 cm in diameter, respectively. The reward location of 10 cm in diameter was always located 20 cm south from the landmark. At the start of a session, the platform and the associated landmark were randomly moved to one of the eight positions, as shown in Fig. 4a. At the beginning of each trial, the animat was placed in one of the four cardinal positions near the wall, with a random initial orientation. The starting locations were pseudorandomly avoiding two consecutive trials with the same start location. The moving speed of the animat was set to 18 cm/s, with a simulation time step corresponding to 1/3 s. If the animat was not able to reach the platform in 200 s, it was automatically guided to it along a direct path to the target, similarly to the real rats in this experiment. Reaching the goal was rewarded by $R = 1$, and wall hits were punished by $R = -0.5$ (see Eq. 5 and 16).

The intact rats were simulated by a full model (Control group), including Taxon, Planning, and Exploration experts. Two lesion groups were simulated: animats in the Taxon group used only Taxon and Exploration experts, while animats in the Planning group used only Planning and Exploration experts. The Taxon group corresponded to the Hc-lesioned animals of the original experiment. The allocentric version of the Taxon version was used in this simulation (see Sect. 5.3.2).

In all simulations now being discussed, the results were averaged over 100 animats (noise in the system was due to the random initialization of weights and random choice of starting position). Both across and within sessions, performance of Control, Taxon and Planning groups were statistically assessed by comparison of their mean escape latencies—the number of time steps per trial—in the first and the fourth trials of a session, using signed-rank Wilcoxon test for matched-paired samples. Between-group comparison was performed using a Mann–Whitney test for non-matched-paired samples. Animat behavior was characterized by three measures: *Goal occupancy rate* of a goal location, defined as the number of times the animat visited a rewarded zone, divided by the total trajectory length; *Goal selection rate* of an expert, calculated as the number of times this particular expert was chosen within a square zone of 0.4 m^2 around the goal, divided by the total number of times the animat visited this zone; *Trial selection rate* of an expert, defined as the number of times the expert was selected over the total length of the trajectory.

The competitive interaction between strategies was estimated by the negative correlation (Pearson's product-moment coefficient) of their selection rates x and y calculated as $\rho_{x,y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$, where σ_{xy} is the covariance, and σ_x, σ_y are the standard deviations of the selection rates x and y, respectively.

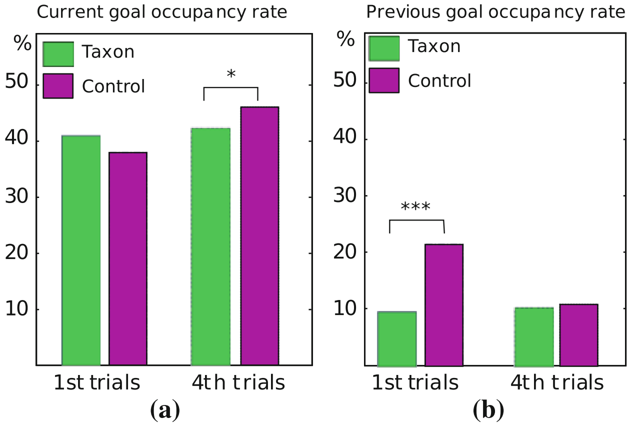

Fig. 5 Occupancy rates for **a** the current and **b** previous goal locations for simulated Taxon and Control groups. *** and * correspond respectively to significance levels $P < 0.001$ and $P < 0.05$

3.2 Simulation results

3.2.1 Learning across and within sessions

Both the simulated Control and Taxon groups were able to learn the location of a hidden platform, as shown by the decrease of their escape latencies (Fig. 4b; $P < 0.001$ for all groups). Moreover, in contrast to the Taxon group, animats in the Control group decreased significantly their escape latencies within all the sessions (Control-1 vs. Control-4 in Fig. 4b).

A comparison of two simulated lesion groups (Taxon and Planning groups) shows that the Taxon expert was responsible for decreasing escape latencies across sessions, while the place-based expert was responsible for learning within sessions (Fig. 4c). Moreover, the Control group found the platform more quickly in the fourth trials (dotted line in Fig. 4b) than both the Taxon and Planning groups (dotted lines in Fig. 4c), suggesting that the two strategies cooperated during learning. This was also assessed by their current goal occupancy rate that increases in fourth trials (Fig. 5a).

Similar to real rats, simulated Control group had greater escape latencies than Taxon group in the first trials (Fig. 4b). Pearce et al. (1998) suggest that this might be explained by the preferential use of the Hc-based strategy at the end of a session, so that, at the beginning of a new session (when the platform has moved to a new location), this strategy led the animal to the previous (thus wrong) platform location. In order to check whether this is the case in our model, we calculated goal occupancy rates near previous and current goal locations for simulated Control and Taxon groups. The results show that indeed, the Control group had a significant bias toward the previous goal location on first trials (Fig. 5b, first trials), while this bias disappeared after the Planning expert had learned the new goal location (Fig. 5b, fourth trials). The reason for this is that the Planning expert of our model was not able to notice that platform and landmark have been moved to a new location at the start of a session, in contrast to the Taxon expert.

Thus, the overall performance of the model in this task is consistent with that reported by Pearce et al. (1998). The advantage of the modeling approach applied here is that we can go further in our analysis of behavior and explore the interactions between behavioral strategies *within* experimental trials. Such an analysis is usually hard to perform in animal experiments like that of Pearce et al. (but is possible for simpler tasks, like e.g., Hamilton et al. 2004). Such a complementary analysis allows us to get insights into (i) the importance of different types of sensory cues for different strategies and (ii) competitive and cooperative interactions between trials across and within experimental sessions.

3.2.2 Influence of sensory cues

In order to analyze the importance of landmark versus spatial cues on learning, we compared the synaptic weights between the connections from Landmark Cells (that encode the landmark) and nodes of Planning Graph (that encode location) to the units of the gating network, which encode the two strategies in the model. The observed increase in the average weights for all connections suggests that all types of cues played a role in the selection process (Fig. 6). However, weights from Landmark Cells to both the Taxon and Planning gating units grew significantly faster with learning, than those from Planning Graph nodes ($P < 0.01$, see caption of Fig. 6). These results suggest that, in our model, the landmark exerted progressively stronger influence on strategy selection than spatial cues, which is consistent with the fact that this task could be solved only by paying attention to the landmark. Nevertheless, the spatial cues were also learned, although

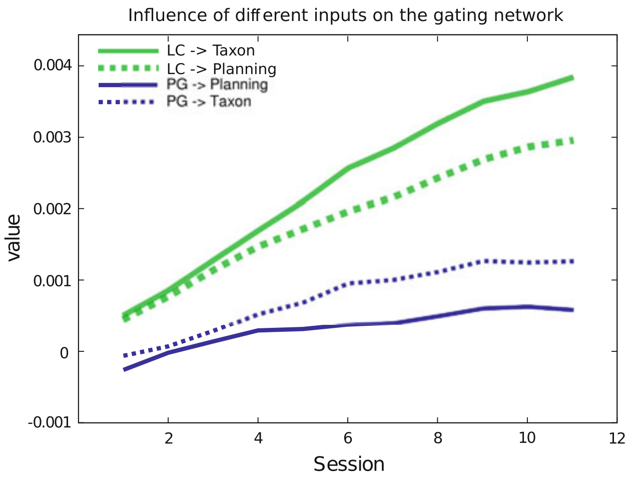

Fig. 6 Evolution of the average synaptic weights between inputs of the gating network and gating units of different strategies. *Thick lines* represent *straight links* (LC → Taxon, PG → Planning). *Dotted lines* represent cross links (LC → Planning, PG → Taxon). A linear regression test on these slopes indicates that LC → Taxon weights grow 5.4 times faster than PG → Taxon weights. Accordingly, LC → Planning weights grow 2.3 times faster than PG → Planning weights

with a smaller rate, and so could influence selection when Planning expert becomes more efficient.

3.2.3 Competition between strategies across experimental sessions

Next, we analyzed the competitive interaction between experts in the Control group across training sessions by comparison of their goal and trial selection rates. Pearce et al. (1998) suggest that, at the beginning of each session, the place-based strategy was in competition with the heading-vector strategy, the latter being the winner of the competition by the end of training. We checked whether our model is consistent with this hypothesis.

At the start of a new session, the Planning expert was not able to detect the change in the platform location and hence its goal selection rate did not change significantly from earlier to later sessions (Fig. 7a, first trials). Accordingly, the first trial selection rate of the Planning strategy did not change significantly across sessions (Fig. 7b). In contrast, the Taxon expert learned to track the changes in landmark position, as suggested by the progressive increase of its trial selection rate across experimental sessions (Fig. 7b, first trials), and by the significant increase in its goal selection rate in the later sessions relative to earlier sessions (Fig. 7c). The competitive interaction between the Taxon and Planning experts is illustrated by the typical trajectory of the simulated animal at the beginning of a session (Fig. 8a). The Planning expert led the animat toward the previous platform location, while the Taxon expert led it toward the current one.

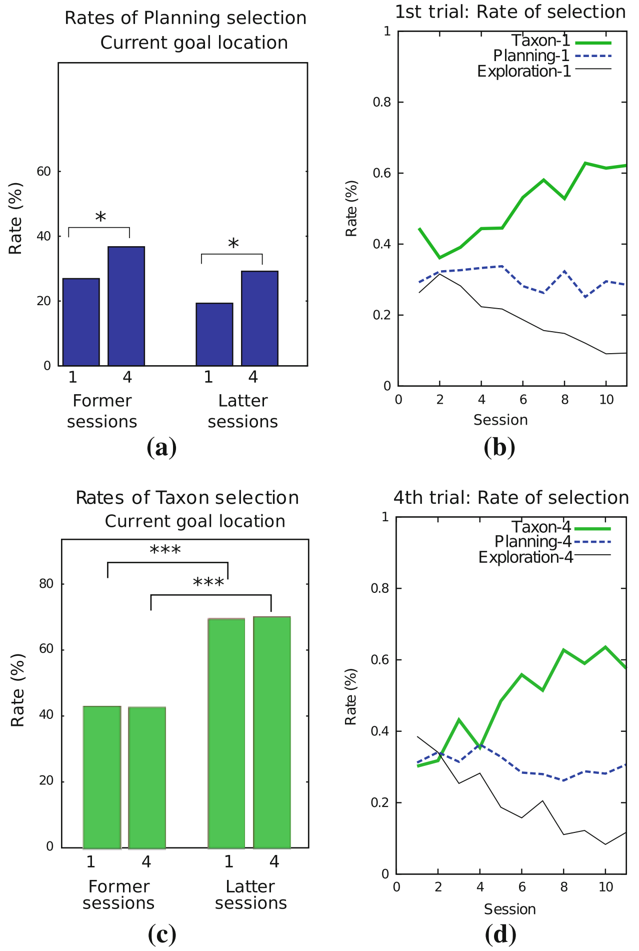

Fig. 7 **a** Selection rates of the Planning expert near the current goal location, across and within sessions. **b** Strategy selection rates across sessions in first trials. **c** Taxon strategy selection rate near the current goal location. **d** Strategy selection rate across the sessions in fourth trials. *** and * correspond respectively to significance levels $P < 0.001$ and $P < 0.05$

Interestingly, the decrease in the trial selection rate of the Exploration expert was almost opposite in magnitude to the increase in the Taxon selection rate (correlation coefficient $r = -0.96$). This result suggests that the preferential use of the Taxon strategy at the end of training corresponds to a decrease in exploratory behavior, rather than a decrease in place-based strategy (Fig. 7b).

3.2.4 Cooperation between strategies within a session

As shown above, the competitive interactions between Taxon and Planning strategies were due to the fact that these two strategies encoded different goal locations at the start of a session. However, this situation changed by the end of session when both strategies had learned the true goal location. In both early and late sessions, the Planning expert was selected significantly more often near the current goal location in

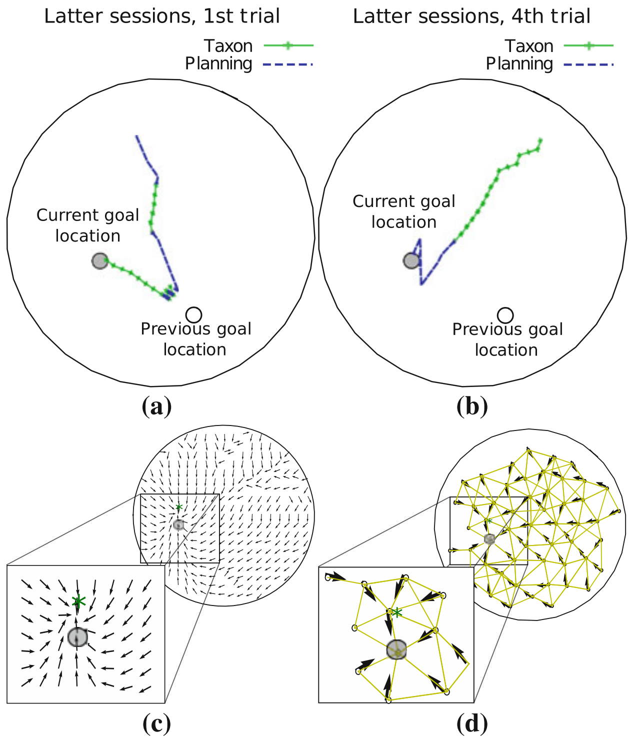

Fig. 8 Control group **a** example of typical trajectories in last sessions of the first trial, **b** example of typical trajectory in the last sessions of the fourth trials and the associated navigational maps around the goal location of **c** Taxon and **d** Planning strategies

fourth trials than in first trials (Fig. 7a), whereas Taxon expert was selected near the current goal location as much often in fourth trials as in first trials in both early and late sessions (Fig. 7c). The increase in Planning selection rate near goal, without provoking a decrease of the Taxon selection rate, and superior performance of Control group over other groups in the fourth trials (Fig. 5a) suggests a cooperative interaction between both experts. Such a cooperative interaction is illustrated by a typical trajectory in the fourth trial (Fig. 8b). Here, both strategies led to the correct goal location and the choice of a particular strategy depended on the accuracy of the corresponding expert at different locations along the trajectory. Examples of navigational maps of the two experts near the goal location are shown in Fig. 8c, d. In these maps, arrows corresponding to the learned directions of movement for each sample location (Taxon expert) or for each spatial node (Planning expert), show that the Taxon expert points southward the landmark, and the Planning expert toward the platform location.

3.2.5 Allocentric Taxon strategy as a heading-vector navigation

In the simulation shown above, we used an allocentric version of the Taxon expert to reproduce the rat behavior attributed by Pearce et al. (1998) to heading-vector navigation. They defined the heading-vector strategy as follows: rats "might use a heading vector that specifies the direction and distance of the goal from a single landmark." Here, we show that the allocentric Taxon expert suits well this definition.

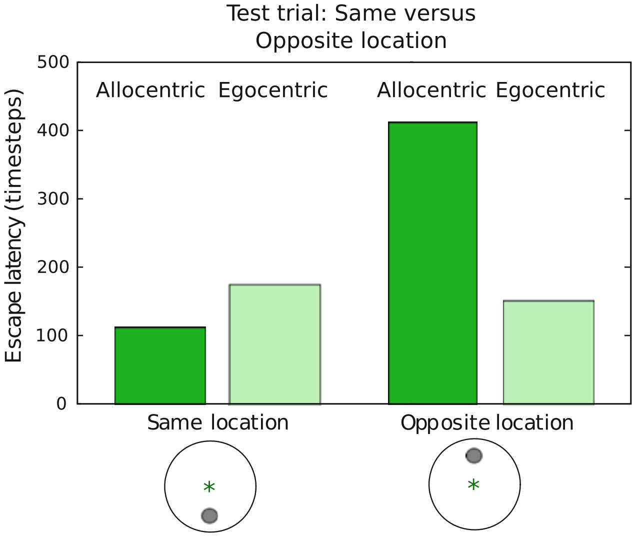

Fig. 9 A correspondence between the allocentric Taxon strategy in the model and the heading-vector strategy (Pearce et al. 1998). The plot shows the mean escape latency to find the platform hidden in the same location relative to the landmark as during training (same location), or in the location opposite to it (opposite location). Contrary to the egocentric Taxon expert, the allocentric Taxon expert had difficulty in finding the platform in the opposite location, since it "remembers" the allocentric direction from the landmark to the hidden goal

In order to demonstrate that the allocentric taxon strategy in the model is similar to the "heading-vector" strategy observed in rats, we performed behavioral test similar to that used in the original experiment. After training the Taxon group in 11 sessions of the main experiment, the landmark was placed at the center of the pool. In the case of half the number of the animats in the simulated Taxon group, the platform was located south of the landmark and at the same distance as before, while for the other half, the platform was located north of the landmark. We compared the performance of the allocentric and egocentric versions of the Taxon expert in the model. Similar to the Hc-lesioned animals, animats with allocentric Taxon expert for which the platform was located north of the landmark took significantly longer to locate the platform than the other group (Fig. 9). This is explained by the fact that the allocentric taxon strategy relies on the remembered allocentric direction from the landmark to the goal, while the egocentric taxon strategy cannot use this information, and hence searches randomly around the landmark (see Sect. 2.1). From these results we conclude that the allocentric Taxon expert is a suitable model of the heading-vector strategy observed by Pearce et al.

In summary, our results support the hypothesis of Pearce et al. (1998) that, at the beginning of the training sessions, place- and response-based strategies were in competition with each other. However, on the basis of results of this study, we propose that, at the end of a session, a cooperation between strategies takes place. In addition, we propose that the improvement of the rat performance by the end of training is not due to the decrease in the use of place-based strategy, but rather due to the decrease in the number of exploratory actions. We stress here that in the model described, the trade-off between exploration and exploitation is not fixed, but learned during training (see Sect. 5).

4 Simulation II: Experiment of Devan and White (1999)

In this experiment, sham-operated, fornix-lesioned and DLS-lesioned groups rats were trained for nine days to remember the location of a platform in a water maze. On days 3, 6, and 9 the platform was hidden, whereas it was visible on the other days. During a competition test on day 10, the visible platform was placed in a novel location (Fig. 10a).

Four principal findings from the original experiment were related to the issue of interaction between place- and response-based strategies (see Fig. 2 in their article). First, sham-operated rats, rats with fornix/fimbria lesions and rats with DLS lesions were equally fast in learning the visible platform location, suggesting that either strategy can be used to approach a visible goal. Second, rats with fornix/fimbria lesions were slower than both sham-operated and DLS-lesioned rats during the hidden platform sessions, suggesting that Hc-dependent strategy, and not the DLS-dependent strategy, is required to locate the hidden platform. Third, on the competition test, rats with fornix/fimbria lesions escaped faster from the pool than either sham-operated or DLS-lesioned groups, suggesting a competition between the two strategies. Fourth, the authors identified two groups of sham-operated animals during the final test day: "place-responders" were approaching the place where the hidden platform was in the previous trial, discarding information from the visible platform in a new place; "cue-responders" headed toward the visible platform and were not biased by the hidden platform location in the previous trials.

4.1 Simulation procedure and data analysis

The experimental setup was similar to that used in Simulation I, except that the diameter of the water maze was set to 172 cm to be consistent with the original protocol. On days 1, 2, 4, 5, 7, 8, the visual landmark 10 cm in diameter (representing the visible platform) was placed into the center of the southwest quadrant of the environment (its position coincides with the reward zone). On days 3, 6, and 9, the landmark was absent, but the reward zone remained in the same location. On day 10, the landmark together with the reward zone were moved to the center of the northeast quadrant of the environment. Starting positions were chosen as in Simulation I. On the competition test the starting position equidistant from both landmark locations was chosen.

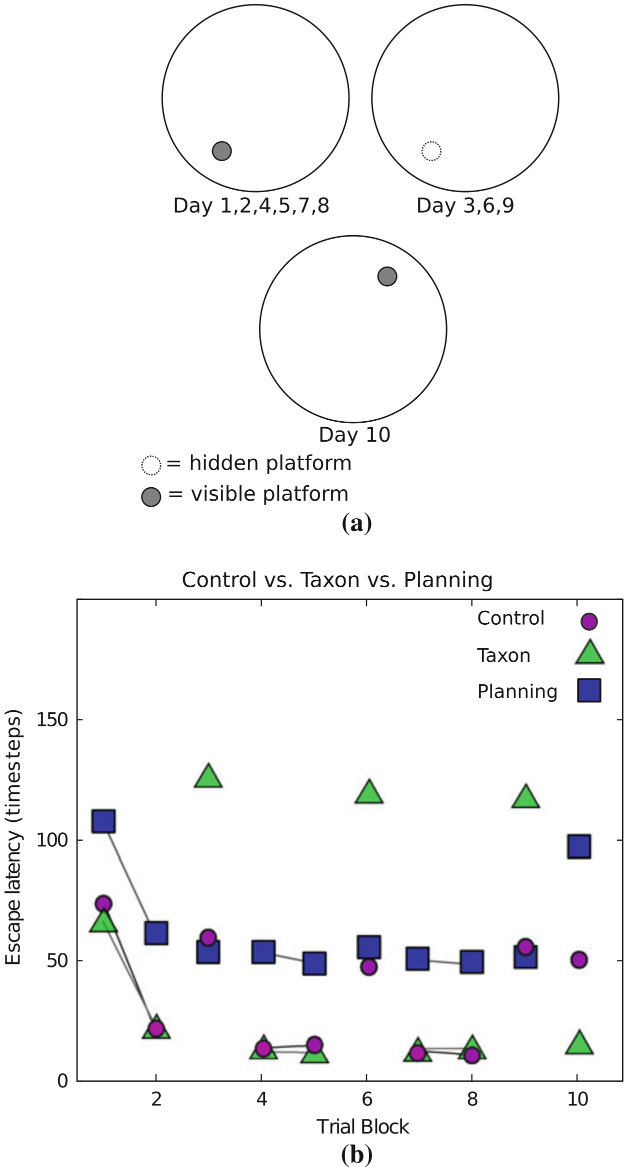

Fig. 10 **a** Protocol of the experiment. **b** Mean escape latencies of simulated rats in Control, Taxon, and Planning groups across sessions with visible (connected plot, days 1, 2, 4, 5, 7, and 8) and hidden (days 3, 6, and 9) platform. Competition test was conducted on day 10 (see text)

Sham-operated, fornix-lesioned and DLS-lesioned groups were respectively simulated by the Control, Taxon and Planning groups as in Simulation I. In this simulation we used the egocentric version of the Taxon expert (see Model and Sect. 5.3.2). The same statistical tests as in Simulation I were used to assess learning.

4.2 Simulation results

4.2.1 Parallel learning of navigational strategies

When the visible landmark was signaling the platform location, all groups of animats were successful in learning the goal location (Fig. 10b, trial blocks 1, 2, 4, 5, 7, and 8). The Planning group was longer than Taxon and Control groups. In this model, this is a consequence of the fact that the platform location did not usually coincide with a graph node, resulting in the lower precision of the Planning Graph compared to the visual input and elevated use of the Exploration expert. The performance of Control and Taxon groups was not different, suggesting that in this case, the behavior of the Control animats was controlled primarily by the Taxon expert.

When the reward location was not signaled by the landmark, the Taxon group had significantly longer escape latencies, that did not decrease with training, similarly to the rats with fornix/fimbria lesions (Fig. 10, trial blocks 3, 6, 9). The performance of the Control group was not different from that of Planning group, suggesting that, in these trial blocks, the behavior was controlled by the Planning expert.

On the competition test, animats from the Taxon group were significantly faster than those from either Control and Planning groups in reaching the new platform location ($P < 0.001$, Fig. 10, trial block 10). In addition, Control group was significantly faster than Planning group, whose performance did not differ from that in the first trial. This last difference was not observed in the original experiment, possibly due to the fact that DLS-lesions in rats may have spared some ability to approach a visible target moved to a new position, whereas our animats in the Planning group were not able to do so. Nevertheless, these results are consistent with the finding of Devan and White (1999) that rats with fornix/fimbria lesions performed significantly better on the competition test than both the Control and DLS-lesioned groups.

Taken together, these results show that our selection model is consistent with the rat behavior observed in this experiment. Similar to the analysis performed in Simulation I, in the next section, we focus on the influence of visual cues and on analysis of strategy interaction.

4.2.2 Influence of sensory cues

The evolution of the synaptic weights in the gating network reflected the irrelevance of the Taxon expert for the trials in which the goal is hidden (Fig. 11). This was expressed by the progressive decrease of the connection weights between spatial cues and the gating unit corresponding to the Taxon strategy. This is in marked contrast with the weight evolution in Simulation I (Fig. 6), where both types of cues were present throughout training and could be both used to find the goal.

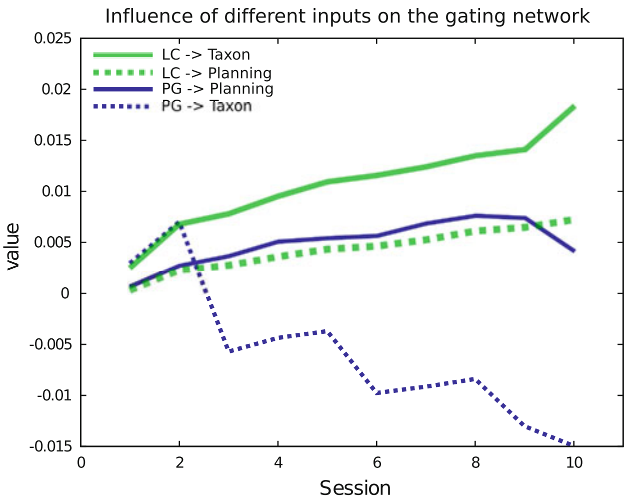

Fig. 11 Synaptic weights of the gating values in the gating network in Control group. *Thick lines* represent straight links (LC → Taxon, PG → Planning). *Dotted lines* represent cross links (LC → Planning, PG → Taxon)

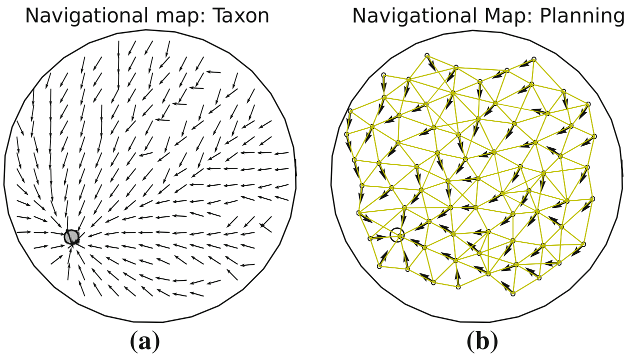

Fig. 12 Navigational maps of **a** Taxon and **b** Planning experts at the end of the trial blocks 8 and 9

4.2.3 The absence of cooperation between strategies during training

During training, both the Taxon and Planning experts learned to approach the fixed goal location. This is illustrated by the navigational maps learned by the two experts (Fig. 12). It can be observed that the map learned by the Taxon expert was more accurate than that of the Planning expert, due to the fact that in this experiment goal location coincided with the landmark. Hence, no cooperation with the Planning expert was necessary in this case. Indeed, trial selection rates of different experts show that the Taxon expert clearly controlled the behavior when the goal was visible (Fig. 13a, b).

In contrast, during trial blocks in which the goal was hidden, the Planning expert was progressively more selected than the Taxon expert (Fig. 13a, b). In addition, the role

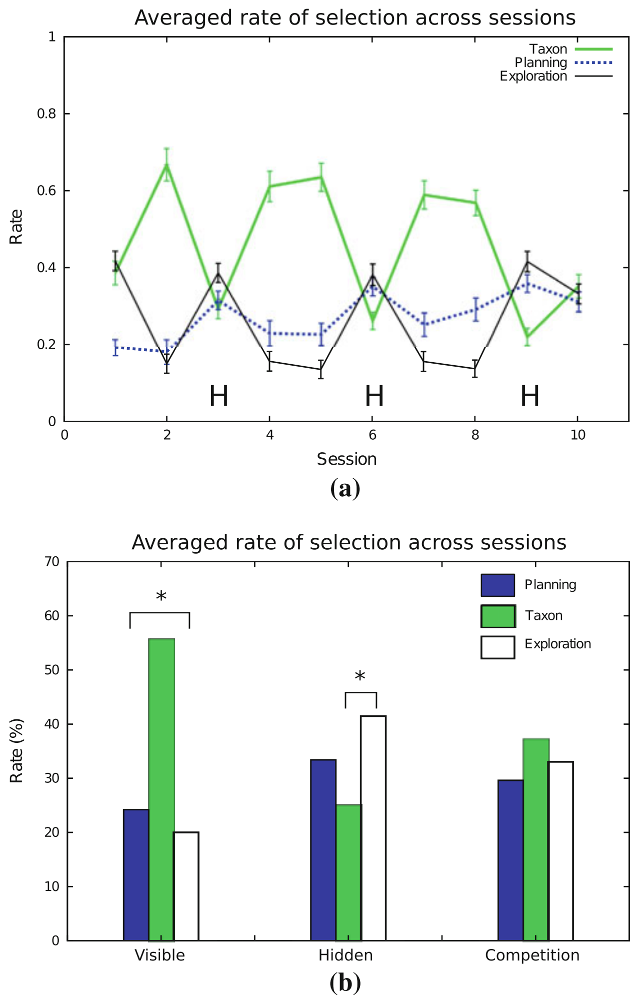

Fig. 13 **a** Selection rates of the three experts during training and competition test in Simulation II (H: Hidden Platform). **b** Summary plot, showing the average rate of selection rate of different experts during trial blocks with visible goal, hidden goal and during competition test

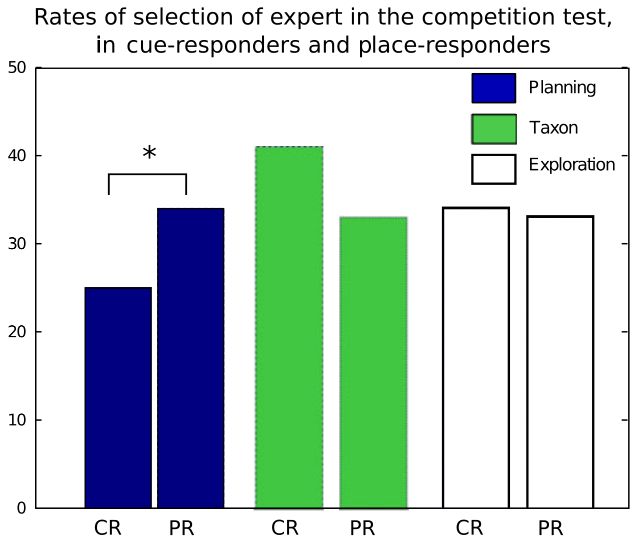

Fig. 14 Rates of selection of experts of cue-responders (CR) and place-responders (PR) in the competition test

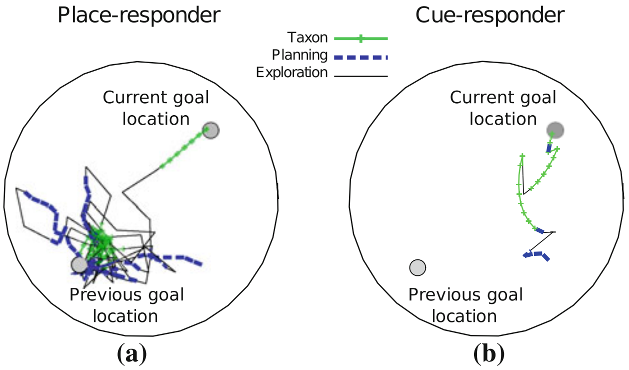

Fig. 15 **a**, **b** Typical trajectories of animats labeled as **a** "place-responders" and **b** "cue-responders"

of exploratory behavior was more prominent in these trials, compensating the relative inaccuracy of the Planning Graph.

4.2.4 Competition between strategies during test

In the competition test, the simulated Control group was able to select a cue-based strategy to reach the goal location, as suggested by escape latencies (Fig. 10) and the selection rate of the Taxon expert (Fig. 13b). However, a significantly better performance of the Taxon group in the competition test (Fig. 10) and a higher selection rate of the Taxon expert during training with visible goal (Fig. 13b) suggests that competition with other strategies slowed down the Control group relative to the Taxon group during the test.

Using the same labeling scheme as Devan and White (1999), Control animats could also be classified into "cue-responders" (59%) and "place-responders" (41%). This division qualitatively reproduced the division of Control rats into both groups of the original experiment (4 "cue-responders" and 6 "place-responders" over 10 animals). Analysis of the trial selection rates of Taxon and Planning experts showed that indeed, in the group of "place-responders" the Planning expert was selected significantly more often than for the group of "cue-responders" ($P < 0.05$), In contrast, in the group of "cue-responders," the Taxon expert was selected more often (although the difference does not reach the significance level, $P = 0.05$, Fig. 14a, b). The observation of typical trajectories of place- and cue-responders shows that place-responders were stuck near the previous goal location during competition test, while cue-responders went almost straight to the visible goal (Fig. 15).

In summary, these results suggest that the proposed selection criterion is flexible enough to deal with rapid strategy switches required when environmental cues drastically change. The Taxon expert in our model learned navigational maps that were more accurate than those of the Planning expert. The limited number of nodes of the Planning Graph was compensated by the high selection rate of the Exploration expert in the sessions with hidden goal (Fig. 13a). The role of the Exploration expert in our model was to find the exact goal location in an approximate goal area signaled by the "cognitive map" (represented by the Planning Graph), rather than to update the map, as is usually proposed (O'Keefe and Nadel 1978).

5 Discussion

We presented a computational model of switching between cue-guided and place-based strategies in the water maze. The main novel property of this model is that it is capable of learning to select between cue-guided and place-based strategies that use different learning algorithms and spatial reference frames to locate a goal. The place-based strategy uses a graph-search algorithm to find the shortest path to the goal. The graph is learned online using the activities of simulated Place Cells that encode spatial location of the animat in an allocentric reference frame. The cue-guided strategy uses a TD learning rule to approach either a visible goal, encoded in an egocentric reference frame; or a hidden goal marked by a landmark, encoded in an allocentric directional reference frame. The strategy selection is performed by a gating network that learns to predict, using a simple TD-learning rule, the most successful strategy, on the basis of the direction of movement that each expert offers at each time step, given all current sensory inputs.

The model was tested in two simulated water-maze tasks designed to investigate interactions between place- and response-based strategies in rats. Owing to the separation between cooperative (during action learning) and competitive (during action selection) interaction between strategies in the model, we were able to assess the relative contribution of different strategies within, as well as across experimental trials. The sections hereafter shall aim at answering the questions raised in the introduction.

5.1 Strategy selection mechanism

5.1.1 Relation to other models

Several models of strategy switching based on the theory of parallel memory systems were proposed earlier (Guazzelli et al. 1998; Daw et al. 2005; Girard et al. 2005; Chavarriaga et al. 2005). In the model of Guazzelli et al. (1998), the orientations proposed by egocentric taxon and allocentric planning strategies are, respectively, determined by current affordances and cognitive knowledge. The final movement is computed as a sum of these orientations that hand-tuned parameters adapt to the situation. A similar selection is also made in the basal-ganglia loops model of Girard et al. (2005). In these models, strategy switches occur in a set of situations a priori chosen by the modeler. In our earlier study (Chavarriaga et al. 2005; Dolle et al. 2008), the strategy-selection network is adaptive, but it is able to select only between strategies that use TD learning to learn the task. Indeed, the selection network uses TD reward-prediction error as a measure of success of different strategies and hence is not able to deal with other goal-navigation algorithms such as planning. Reinforcement learning framework is also used in the model of Uchibe and Doya (2005) to select between two navigational strategies, but does not handle strategies that are not learned by RL. Finally, the model of action selection in an operant conditioning (Daw et al. 2005) proposes another interesting mechanism of selection depending on the relative uncertainty of different experts. However, in this model, the tree-based computations performed by the experts only allow the model to work with rather small state spaces, and hence cannot be applied to navigation in continuous space. The advantage of the selection criterion proposed in this study is that it permits comparison between experts that use different learning rules and scales well with increasing number of exerts.

5.1.2 The role of random exploration

In the above model, exploration is implemented as a separate "strategy," i.e., during goal learning, it is chosen when its gating value is the highest among the gating values of all the strategies. It means that the need for exploring novel actions is learned during training and can depend on sensory input. This is in contrast to standard reinforcement learning algorithms in which exploration is chosen according to a predefined stochastic scheme. For example, Arleo and Gerstner (2000) and Chavarriaga et al. (2005) use an ϵ-greedy scheme, in which novel actions are tested with small probability ϵ on each time step, while Foster et al. (2000) use a soft-max selection where actions with high Q-values have a higher probability of being chosen. In robotic experiments (Cuperlier et al. 2007; Barrera and Weitzenfeld 2007), the exploration is chosen when the animat cannot associate its location with any existing node in its topological map. In Girard et al. (2005), the exploration is a random direction chosen among the other strategies, but the selection is not learned. We show here that the model in which the balance between exploitation and exploration is not predefined but learned with training can reproduce well the rat behavior in two real-world behavioral tasks. In agreement with standard RL algorithms, the exploration is mainly chosen at the beginning of the training and then decreases as the strategies are learned (Fig. 7). Our simulations also show that Planning strategy is associated with higher exploration rate (Fig. 13b, sessions 3, 6, and 9), which is explained by the lower accuracy of the cognitive map compared to visual input (due to a limited number of nodes). In the model proposed, the path to the goal derived from the cognitive map can only follow connections between nodes, thus producing paths which are close to optimal, but still deviating from the approximately straight paths generated by Taxon strategy.

The above mentioned model suggests that exploratory behavior may be governed by a separate brain network similarly to Taxon (DLS) and planning (Hc–PFC) networks. If so, then exploratory behavior can be potentially dissociated from other strategies using a specialized experimental paradigm. In support of this idea, several experimental addressed thigmotaxic (i.e., wall-following) behavior which can be considered as an exploratory (yet non-random) behavior (Devan and White 1999; Devan et al. 1999; Pouzet et al. 2002; Chang and Gold 2004).

5.2 The mechanism of selection can result in competition and cooperation between strategies, across and within trials

In the above model, the Taxon and Planning experts learn in parallel and in such a way that action–outcome pairs generated by one of the experts can be used by the other expert to update its action value estimates. Learning of an expert from the actions performed by another expert represents cooperation between strategies in our model, which fits well the definition of cooperation introduced by behavioral studies (see Sect. 1). In our simulations, the facilitating effect of cooperation is clearly seen by observing that performance of intact simulated animals is always better than or equal to that of lesioned simulated animals, when both strategies predict correct paths (Fig. 4b, Taxon-4 and Control-4).

On the other hand, the gating network will select an expert with the highest gating value at each time step, where the gating value corresponds to the total future reward predicted for this strategy. Such a reward-based selection of experts allows competition between strategies (see Sect. 1). Evidence for competition in our simulations is given by performance data showing that when two strategies suggest contradictory predictions about goal location, lesioned simulated animals outperform control ones (Fig. 4b, Taxon-1 and Control-1 and Fig. 10b, session 10). In summary, the presented model provides a rather simple strategy selection mechanism which implements cooperation as well as competition between the strategies within the same network.

5.3 Influence of sensory cues

5.3.1 Influence of intra-maze and extra-maze cues

A noticeable contribution of the model concerns the analysis of the influence of different types of sensory cues (intra versus extramaze) on strategy selection, which is hard to do in real life experiments. Within the gating network, the gating units of both Taxon and Planning strategies receive two types of sensory input provided by Landmark Cells (i.e., landmark information) and Planning Graph nodes (location information). Essentially this means that the availability of sensory cues at each moment in time determines the relative values of available strategies. Hence, by observing the evolution of synaptic weights between sensory inputs and gating units, it is possible to assess the relative contribution of different types of input on the behavior. From the weight analysis in our simulations we make two observations.

First, in both behavioral tasks, landmark information is more important than spatial cues for strategy selection, as shown by larger average weights of Landmark Cells compared to Place Cells (bold lines versus thin lines, respectively, in Figs. 6, 11). In Simulation I, this is due to a higher accuracy of landmark information over information provided by spatial cues, since the landmark signals the correct goal location at the beginning of a session. In Simulation II, this is due to the fact that the presence or the absence of the landmark determines whether the Taxon strategy can be used at all.

Second, in Simulation II, the input from the spatial cues (i.e., Planning Graph nodes) serves mainly to decrease the influence of Taxon expert in the trials with hidden goal by negative projection from Place Cells to the Taxon gating value (Fig. 11). However, this does not completely prevent this inappropriate expert from being selected in this situation (see Fig. 13a, showing that the Taxon is selected even when it cannot "see" the landmark). In the absence of a landmark, the Taxon expert proposes a randomly chosen action and is thus equivalent to the Exploration expert. Its selection rate on the trials without landmark decreases with learning, as can be seen from Fig. 13a, b.

5.3.2 Allocentric versus egocentric cue-based learning

There are two versions of Taxon strategy in the model. They use exactly the same learning algorithm, but the visual cues are represented in an allocentric directional reference frame for the allocentric Taxon expert, and in an egocentric reference frame for the egocentric Taxon expert (Fig. 2 and Sect. 2.1). The use of allocentric directional frame implicitly requires the use of stable extra-maze cues with respect to which such a frame is defined. Our model does not include the estimation of the allocentric head direction from extra-maze cues (see Skaggs et al. 1995; Zhang 1996), but it is assumed to be provided by the head direction network (Taube et al. 1990. In contrast, infromation from the intra-maze cues is sufficient for the egocentric Taxon expert to determine direction to the goal.

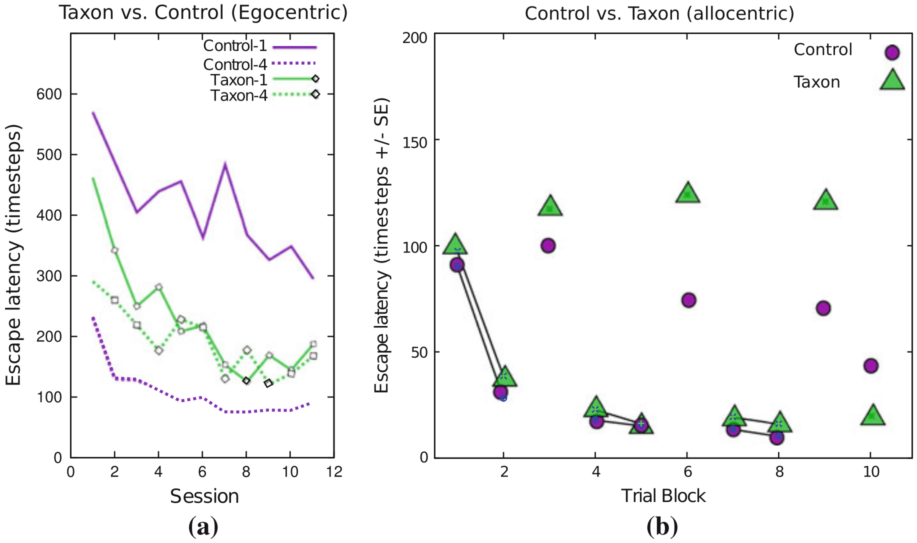

Fig. 16 **a** Results of Simulation I (Pearce et al.'s experiment) with an egocentric Taxon instead of an allocentric one. **b** Results of Simulation II (Devan and White's experiment) with an allocentric Taxon instead of an egocentric one

As shown in Fig. 9, in contrast to the egocentric Taxon strategy, the allocentric Taxon strategy reproduces the rat behavior attributed to the "heading vector" strategy observed by Pearce et al. (1998). This is because the allocentric Taxon strategy takes into account the current allocentric heading, and thus is able to tell whether the platform is located north or south of the landmark. When the platform position changes, the allocentric Taxon strategy fails to find the goal. For the egocentric Taxon strategy, the two cases are identical since the animat is using random search around the landmark in both cases.

We note here that our main results will not change if we use egocentric Taxon strategy in the simulation of the experiment of Pearce et al. (1998), as demonstrated in Figs. 4a and 16a. The use of the egocentric strategy simply slows down the performance of both Taxon and Control groups. Accordingly, the use of an allocentric Taxon strategy does not deeply change the results of Taxon and Control groups in the simulation of Devan and White (1999) when the platform is visible (Figs. 10b, 16b). However, Control group is much less efficient in hidden trials: in the sudden absence of the landmark, the allocentric Taxon, which has memorized the previous heading, helps to a lesser extent in finding the goal than does the egocentric Taxon which proposes a random orientation.

5.4 Neural substrates for the strategy-selection network

According to Ragozzino et al. (1999) and Rich and Shapiro (2009), the prelimbic–infralimbic areas (PL/IL) of the medial prefrontal cortex (mPFC) are not required for acquiring navigation strategies, but are responsible for switching between them. These data fit well to the model proposed here. Indeed, PL/IL areas receive afferents from Hc (e.g., Conde et al. 1995) and dorsomedial striatum (e.g., Groenewegen et al. 1991) which are the potential biological loci for the place- and cue-based learning, respectively. Moreover, PFC receives dopaminergic projection from the ventral tegmental area (e.g., Descarries et al. 1987). and so the reward information necessary for reward-based learning in the model may be available in the PFC.

On the neural level, Rich and Shapiro (2009) observed that different subpopulations of mPFC neurons code for different behavioral strategies. In the current model, gating values of different strategies can be considered as representing the activity of these subpopulations. Indeed, switches between strategies in the current model correspond to switch in relative gating values: if Taxon gating value is greater than Planning gating value, Taxon strategy takes the control of behavior, and vice versa (see, e.g., Figs. 6, 13). This switch between relative gating values corresponds to the switch between population activities in the recorded data of Rich and Shapiro (2009) (see Fig. 6a in their article).

Despite these similarities, however, the model cannot account for some other data in relation to the role of the mPFC in behavior. For example, it has been shown that mPFC is responsible for cross-modal but not intra-modal selection (i.e., reversal learning, Young and Shapiro 2009). In the current model, both strategy switching and reversal can be learned within the same network, since reversal in our model corresponds to simply changing the reward location. Other inconsistencies come from the study of Rich and Shapiro (2007), who have shown that mPFC is involved only during first strategy switches and it does not seem to play a role during subsequent switches. Our model cannot provide plausible explanation for these data. In summary, mPFC might be considered as a biologic locus for the selection network, but in this case (i) a separation of the gating network into at least two different parts is required to take into account the reversal data (Young and Shapiro 2009), and (ii) an extension to the model is required to explain how the strategy switching is performed after more than a few subsequent switches (Rich and Shapiro 2007).

6 Conclusion

This study proposes a mechanism of switching between procedural cue-based and cognitive place-based navigation experts in continuous environment. The cue-based expert uses visual input, while the place-based expert uses a topological representation of the environment built on the basis of Place Cells. Random exploration is considered as a separate strategy and participates in the strategy selection process. The selection between strategies is performed by estimating how successful the strategies are in predicting the reward, on the basis of the direction of movement they propose. The model is able to select between navigation strategies that are based on distinct learning mechanisms (i.e., procedural or cognitive), potentially operating in different spatial reference frames (i.e., allocentric or egocentric). As we demonstrated, the model can serve as a useful tool for analyzing interactions between navigational strategies in spatial learning and for prediction of behaviours of lesioned animals.

The model is intended to be extended to model experimental paradigms that add, change, or remove extra-maze landmarks. The current integration of a recent hippocampal model (Ujfalussy et al. 2008) will allow Place Cells to be learned on line and to express dynamic changes in the environment. The model will also be able to simulate paradigms using multiple intra-maze landmarks. Addition of a second landmark amounts to adding another Taxon expert (either egocentric or allocentric) tuned to the new landmark. No changes need to be implemented in the selection network. Such an extended model can potentially be used to address the issue of blocking and overshadowing effects between different types of cues (Rescorla and Wagner 1972; Chamizo 2003; Gibson and Shettleworth 2003, 2005; Stahlman and Blaisdell 2009). These effects are inherent to any learning algorithm which updates associative weights between cues and rewards so as to reduce reward prediction error (e.g., TD-learning) as is true for the selection network in our model.

Acknowledgment This research was granted by the EC Integrated Project ICEA (Integrating Cognition, Emotion and Autonomy, IST 027819).

References

Arleo A, Gerstner W (2000) Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. Biol Cybern 83(3):287–299

Arleo A, Rondi-Reig L (2007) Multimodal sensory integration and concurrent navigation strategies for spatial cognition in real and artificial organisms. J Integr Neurosci 6(3):327–366

Barrera A, Weitzenfeld A (2007) Bio-inspired model of robot spatial cognition: topological place recognition and target learning. In: CIRA, pp 61–66

Blaisdell A (2009) The role of associative processes in spatial, temporal, and causal cognition. In: Watanabe SB, Blaisdell AP, Huber L, Young A (eds) Rational animals, irrational humans. Keio University Press, Tokyo pp 153–172

Brown M, Sharp P (1995) Simulation of spatial learning in the Morris water maze by a neural network model of the hippocampal formation and nucleus accumbens. Hippocampus 5(3):171–188

Burgess N (2008) Spatial cognition and the brain. Ann N Y Acad Sci 1124:77–97

Burnod Y (1991) Organizational levels of the cerebral cortex: an integrated model. Acta Biotheor 39(3–4):351–361

Canal C, Stutz S, Gold P (2005) Glucose injections into the dorsal hippocampus or dorsolateral striatum of rats prior to T-maze training: modulation of learning rates and strategy selection. Learn Mem 12(4):367–374

Chamizo V (2003) Acquisition of knowledge about spatial location: assessing the generality of the mechanism of learning. Q J Exp Psychol 56(1):102–113

Chang Q, Gold PE (2003) Switching memory systems during learning: changes in patterns of brain acetylcholine release in the hippocampus and striatum in rats. J Neurosci 23(7):3001

Chang Q, Gold PE (2004) Inactivation of dorsolateral striatum impairs acquisition of response learning in cue-deficient, but not cue-available, conditions. Behav Neurosci 118(2):383–388

Chavarriaga R, Strösslin T, Sheynikhovich D, Gerstner W (2005) A computational model of parallel navigation systems in rodents. Neuroinformatics 3(3):223–242

Conde F, Maire-Lepoivre E, Audinat E, Crepel F (1995) Afferent connections of the medial frontal cortex of the rat. II. Cortical and subcortical afferents. J Comp Neurol 352(4):567–593

Cuperlier N, Quoy M, Gaussier P (2007) Neurobiologically inspired mobile robot navigation and planning. Front Neurorobotics 1: 1–15

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8(12):1704–1711

Descarries L, Lemay B, Doucet G, Berger B (1987) Regional and laminar density of the dopamine innervation in adult rat cerebral cortex. Neuroscience 21(3):807–824

Devan B, White N (1999) Parallel information processing in the dorsal striatum: relation to hippocampal function. J Neurosci 19(7):2789–2798

Devan B, McDonald R, White N (1999) Effects of medial and lateral caudate-putamen lesions on place-and cue-guided behaviors in the water maze: relation to thigmotaxis. Behav Brain Res 100(1–2): 5–14

Dijkstra E (1959) A note on two problems in connection with graphs. Numer Math 1(269–270):269–271

Doeller CF, Burgess N (2008) Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. Proc Natl Acad Sci USA 105(15):5909–5914

Doeller CF, King JA, Burgess N (2008) Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. Proc Natl Acad Sci USA 105(15):5915–5920

Dolle L, Khamassi M, Girard B, Guillot A, Chavarriaga R (2008) Analyzing interactions between navigation strategies using a computational model of action selection. LNAI 5248:71–86

Foster DJ, Morris RG, Dayan P (2000) A model of hippocampally dependent navigation, using the temporal difference learning rule. Hippocampus 10(1):1–16

Franz MO, Mallot HA (2000) Biomimetic robot navigation. Rob Auton Syst 30(1):133–153

Gibson B, Shettleworth S (2003) Competition among spatial cues in a naturalistic food-carrying task. Learn Behav 31(2):143–159

Gibson B, Shettleworth S (2005) Place versus response learning revisited: tests of blocking on the radial maze. Behav Neurosci 119(2):567–586

Girard B, Filliat D, Meyer J, Berthoz A, Guillot A (2005) Integration of navigation and action selection functionalities in a computational model of cortico-basal-thalamo-cortical loops. Adapt Behav 13(2):115–130

Gold P (2004) Coordination of multiple memory systems. Neurobiol Learn Mem 82(3):230–242

Grahn J, Parkinson J, Owen A (2008) The cognitive functions of the caudate nucleus. Prog Neurobiol 86(3):141–155

Granon S, Poucet B (1995) Medial prefrontal lesions in the rat and spatial navigation: evidence for impaired planning. Behav Neurosci 109(3):474–484

Groenewegen H, Berendse H, Meredith G, Haber S, Voorn P, Wolters J, Lohman A (1991) The mesolimbic dopamine system: from motivation to action. In: Willner P, Scheel-Kriiger J (eds) Functional anatomy of the ventral, limbic system-innervated striatum. Wiley, Chichester pp 19–59

Guazzelli A, Corbacho F, Bota M, Arbib M (1998) Affordances, motivation, and the world graph theory. Adapt Behav 6(3):435–471

Hamilton D, Rosenfelt C, Whishaw I (2004) Sequential control of navigation by locale and taxon cues in the morris water task. Behav Brain Res 154(2):385–397

Hartley T, Burgess N (2005) Complementary memory systems: competition, cooperation and compensation. Trends Neurosci 28(4):169–170

Hasselmo ME (2005) A model of prefrontal cortical mechanisms for goal-directed behavior. J Cogn Neurosci 17(7):1115–1129

Jankowski J, Scheef L, Hüppe C, Boecker H (2009) Distinct striatal regions for planning and executing novel and automated movement sequences. Neuroimage 44(4):1369–1379

Kelly D, Gibson B (2007) Spatial navigation: spatial learning in real and virtual environments. Comp Cogn Behav Rev 2:111–124

Khamassi M (2007) Complementary roles of the rat prefrontal cortex and striatum in reward-based learning and shifting navigation strategies. PhD thesis, University Paris 6

Kim J, Baxter M (2001) Multiple brain-memory systems: the whole does not equal the sum of its parts. Trends Neurosci 24(6):324–330

Leising K, Blaisdell A (2009) Associative basis of landmark learning and integration in vertebrates. Comp Cogn Behav Rev 4:80–102

Martinet LE, Passot JB, Fouque B, Meyer JA, Arleo A (2008) Map-based spatial navigation: a cortical column model for action planning. LNAI 5248:39–55

McDonald R, White N (1993) A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. Behav Neurosci 107(1):3–22

McDonald R, White N (1994) Parallel information processing in the water maze: evidence for independent memory systems involving dorsal striatum and hippocampus. Behav Neural Biol 61(3):260–270

McDonald R, Devan B, Hong N (2004) Multiple memory systems: the power of interactions. Neurobiol Learn Mem 82(3):333–346

Mizumori S (2008) Hippocampal place fields. Oxford University Press, USA

O'Keefe J, Dostrovsky J (1971) The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res 34(1):171–175

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

Packard M, McGaugh J (1992) Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: further evidence for multiple memory systems. Behav Neurosci 106(3):439–446

Packard M, McGaugh J (1996) Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiol Learn Mem 65(1):65–72

Packard M, Hirsh R, White N (1989) Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems. J Neurosci 9:1465–1472

Pearce J (2009) The 36th Sir Frederick Bartlett Lecture: an associative analysis of spatial learning. Q J Exp Psychol 62(9):1665–1684

Pearce J, Roberts A, Good M (1998) Hippocampal lesions disrupt navigation based on cognitive maps but not heading vectors. Nature 396(6706):75–77

Pouzet B, Zhang W, Feldon J, Rawlins J (2002) Hippocampal lesioned rats are able to learn a spatial position using non-spatial strategies. Behav Brain Res 133(2):279–291

Ragozzino M, Detrick S, Kesner R (1999) Involvement of the prelimbic-infralimbic areas of the rodent prefontal cortex in behavioral flexibility for place and response learning. J Neurosci 19(11):4585–4594

Redish A (1999) Beyond the cognitive map: from place cells to episodic memory. The MIT Press, Cambridge

Rescorla R, Wagner A (1972) A theory of pavlovian conditioning: the effectiveness of reinforcement and non-reinforcement. In: Black A, Prokasy W (eds) Classical conditioning II: current research and theory. Appleton-Century-Crofts, New York, pp 64–69

Rich E, Shapiro M (2007) Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. J Neurosci 27(17):4747

Rich E, Shapiro M (2009) Rat prefrontal cortical neurons selectively code strategy switches. J Neurosci 29(22):7208–7219

Roberts A, Pearce J (1999) Blocking in the Morris swimming pool. J Exp Psychol Anim Behav Process 25(2):225–235

Save E, Poucet B (2000) Involvement of the hippocampus and associative parietal cortex in the use of proximal and distal landmarks for navigation. Behav Brain Res 109(2):195–206

Sheynikhovich D, Chavarriaga R, Strösslin T, Arleo A, Gerstner W (2009) Is there a geometric module for spatial orientation? Insights from a rodent navigation model. Psychol Rev 116(3):540–566

Skaggs W, Knierim J, Kudrimoti H, McNaughton B (1995) A model of the neural basis of the rat's sense of direction. Adv Neural Inf Process Syst 7:173–182

Stahlman W, Blaisdell A (2009) Blocking of spatial control by landmarks in rats. Behav Processes 81(1):114–118

Strösslin T, Sheynikhovich D, Chavarriaga R, Gerstner W (2005) Robust self-localisation and navigation based on hippocampal place cells. Neural Netw 18(9):1125–1140

Sutton R, Barto A (1998) Reinforcement learning: an introduction. Bradford Book. The MIT Press, Cambridge

Taube JS, Muller RU, Ranck JBJr (1990) Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J Neurosci 10(2):420

Touretzky D, Redish A (1996) Theory of rodent navigation based on interacting representations of space. Hippocampus 6(3):247–270

Uchibe E, Doya K (2005) Reinforcement learning with multiple heterogeneous modules: a framework for developmental robot learning. In: The 4th international conference on development and learning. IEEE Computer Society Press, pp 87–92

Ujfalussy B, Eros P, Somogyvari Z, Kiss T (2008) Episodes in space: a modelling study of hippocampal place representation. LNAI 5040:123–136

Voermans N, Petersson K, Daudey L, Weber B, Van Spaendonck K, Kremer H, Fernández G (2004) Interaction between the human hippocampus and the caudate nucleus during route recognition. Neuron 43(3):427–435

White N (2004) The role of stimulus ambiguity and movement in spatial navigation: a multiple memory systems analysis of location discrimination. Neurobiol Learn Mem 82:216–229

White N (2009) Some highlights of research on the effects of caudate nucleus lesions over the past 200 years. Behav Brain Res 199(1):3–23

White N, McDonald R (2002) Multiple parallel memory systems in the brain of the rat. Neurobiol Learn Mem 77:125–184

Yin H, Knowlton B (2004) Contributions of striatal subregions to place and response learning. Learn Mem 11(4):459–463

Young J, Shapiro M (2009) Double dissociation and hierarchical organization of strategy switches and reversals in the rat PFC. Behav Neurosci 123(5):1028–1035

Zhang K (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. J Neurosci 16(6):2112