Neuroinformatics

1539-2791/05/223–242/$30.00
DOI: 10.1385/NI:3:3:223

Original Article

A Computational Model of Parallel Navigation Systems in Rodents

*Ricardo Chavarriaga, Thomas Strösslin, Denis Sheynikhovich, and Wulfram Gerstner**

Ecole Polytechnique Fédérale de Lausanne (EPFL), School of Computer and Communication Sciences and Brain Mind Institute, 1015 Lausanne, Switzerland

Abstract

Several studies in rats support the idea of multiple neural systems competing to select the best action for reaching a goal or food location. Locale navigation strategies, necessary for reaching invisible goals, seem to be mediated by the hippocampus and the ventral and dorsomedial striatum whereas taxon strategies, applied for approaching goals in the visual field, are believed to involve the dorsolateral striatum. A computational model of action selection is presented, in which different experts, implementing locale and taxon strategies, compete in order to select the appropriate behavior for the current task. The model was tested in a simulated robot using an experimental paradigm that dissociates the use of cue and spatial information.

Index Entries: Action selection; biomimetic agents; navigation strategies; reinforcement learning.

(Neuroinformatics DOI: 10.1385/NI:3:3:223)

Parallel Navigation Systems

Animals can adopt different navigation strategies depending on the environment and the task they have to solve (Redish, 1999). In some cases, the target is visible and can be reached by applying simple landmark guidance behavior. This kind of strategy is classified as *taxon navigation*. In other tasks, classified as *locale navigation*, the target cannot be identified by any single cue (or sequence of cues), requiring the use of a spatial representation. For rodents such a representation, whose anatomical locus seems to be the hippocampus (O'Keefe and Nadel, 1978), is the basis of the cognitive map theory (Tolman, 1948). A third type of strategy, called *praxic navigation*, involves the execution of a constant motor program. Both praxic and taxon strategies can be

*Author to whom all correspondence and reprint requests should be addressed.
E-mail: wulfram.gerstner@epfl.ch

understood as a stimulus–response-based navigation (Redish, 1999).[1]

Several studies have been designed to assess the use of spatial information as opposed to simple cue response in solving a navigational task, that is, the dissociation of a locale vs a taxon strategy (Packard et al., 1989; Packard and McGaugh, 1996; Pearce et al., 1998; Devan and White, 1999; Chang and Gold, 2003; Da Cunha et al., 2003). They have shown that hippocampal lesions impair the learning of spatial tasks, but have little or no effect in navigating toward visible goals. In contrast, lesions in the dorsal striatum impair the learning of a stimulus–response association (such as swimming toward a visible platform).

For example, Devan and White (1999) use a combined cue-place learning task to compare the effect of lesions in different areas implicated in navigation. They use the water maze task, which consists of a pool of colored water in which the rats are trained to swim to reach an escape platform which can be either below the surface (hidden version) or protruding above the water (visible version). During the first 2 d, the rats are placed in the visible version of the maze, which corresponds to a stimulus–response task. On the third training day, the visible platform is replaced by a submerged platform at the same location, requiring the use of spatial information to solve the task. The sequence is repeated two more times, for a total of six training days in the visible maze (days 1, 2, 4, 5, 7, and 8), and 3 d in the hidden maze (days 3, 6, and 9).

In order to dissociate the responses to the two types of information (visual cue vs spatial information), a competition trial is performed at day 10, where the platform is *visible* but at a *different* location. Animals were either intact (control group) or they had lesion in one of the following three areas implicated in navigation: (i) the hippocampus, (ii) the dorsolateral striatum, and (iii) the dorsomedial striatum.[2] During the competition trial animals with either hippocampal or dorsomedial lesions swam directly toward the visible platform (*cue response*), applying a taxon strategy to solve the task. In contrast, dorsolateral lesions produce a preference to the use of a locale strategy, having them swimming first toward the location where the platform was during the training phase (*place response*). All intact animals were able to solve the task, but 60% of them swam first to the former platform location before turning to the visible platform (i.e., exhibiting a place response and then switching to a taxon strategy). Figure 1 shows representative swimming paths of cue and place responders in the competition trial.

These results suggest that there exist at least two navigation systems in the rat brain working in parallel and mediating different forms of learning. One, including the dorsolateral striatum, mediates a form of learning in which a cue is associated with reward and triggers a guidance behavior (taxon strategy). The other system participates in the association between place and reward (locale strategy). The hippocampus seems to be one of the components of this system.[3] These systems work in parallel in a competitive way according to the situation in which learning occurs (White and McDonald, 2002).

This paper presents a model for animal navigation able to select between taxon and locale strategies to solve navigational tasks. Based on the hypothesis that multiple parallel systems

[1]In the present article the term taxon is often used to refer to stimulus–response behavior in the general sense.

[2]The striatum corresponds to one of the input structures of the basal ganglia. In rodents, the dorsal striatum is also referred to as caudato putamen.

[3]The nucleus accumbens in the ventral basal ganglia and, to some extent, the dorsomedial striatum (Devan and White, 1999) seem to be also involved in this navigation system.

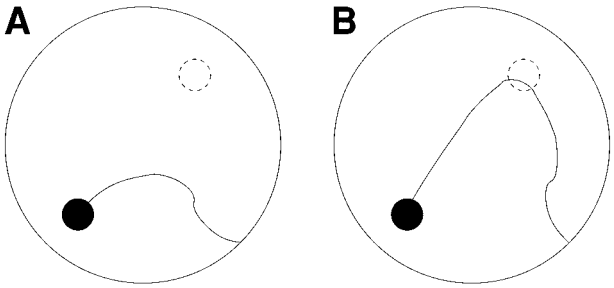

Fig. 1. Representative swimming paths of cue and place responders in the competition trial. The dashed circle shows the location of the platform during training. The filled circle shows the location of the visible platform in the competition trial. **(A)** Animals exhibiting cue response (taxon strategy) swim directly to the visible platform. **(B)** Some animals swim first to the location of the platform (place response), before going toward the visible cue. (Redrawn after Devan and White, 1999.)

control these strategies, we implement each strategy as a separate, independent module. The module implementing the locale strategy learns associations between the location of the agent and the actions required to reach the goal. A representation of space, encoded in a population of simulated place-sensitive cells, is used to achieve self-localization. A second module, implementing the taxon strategy, is based on the association of sensory input (e.g., signaling a cue) to a specific response. Action control is performed by selecting among the directions of movement provided by these two modules. The question we ask in this paper is whether it is possible to design a competition mechanism between the two modules that is flexible, learnable, and would automatically select locale or taxon strategies depending on the task it is required to solve. As we will see, the same competition mechanism will also change from one strategy to another, when the strategy that was chosen first fails to solve the task, in a similar way as place responder rats do in the experiment described above.

It is hypothesized that such a competition mechanism exists, but we do not want to speculate about its biological implementation, be it a local or distributed system. Even though the modeling of the competition mechanism is not biologically detailed, it serves to illustrate some key points of the interaction among the systems involved in rodent navigation, as well as provide some insight into the nature of the computation underlying the selection of navigation strategies in rodents.

Description of the Model

According to the hypothesis of multiple parallel systems, a full model for spatial navigation should contain two separate modules, one for the *locale* and the other for the *taxon* strategies, and provide a mechanism to select the one that is the most appropriate in the current context. In this section we first describe the modules for locale (in the subsection "Locale Navigation Strategy") and taxon (in the subsection "Taxon Navigation Strategy") strategies, and then we present in detail the selection (in the subsection "Strategy Selection") and learning mechanism (in the subsections "Updating the Modules" and "Updating the Gating Network").

Locale Navigation Strategy

The locale navigation strategy requires a representation of space, which has been suggested to reside in the hippocampus (O'Keefe and Nadel, 1978). Our model of locale navigation is similar to the model proposed by Arleo and Gerstner (2000), except that we use a simplified representation of visual input as explained below. Both the models of Arleo and Gerstner (2000) and the current one build an incremental population of place-sensitive cells by combining external sensory input with internal (idiothetic) information. The population of place cells can then be used to learn about an association between states (places) and actions (directions of movement) allowing the model to perform navigation toward hidden goals.

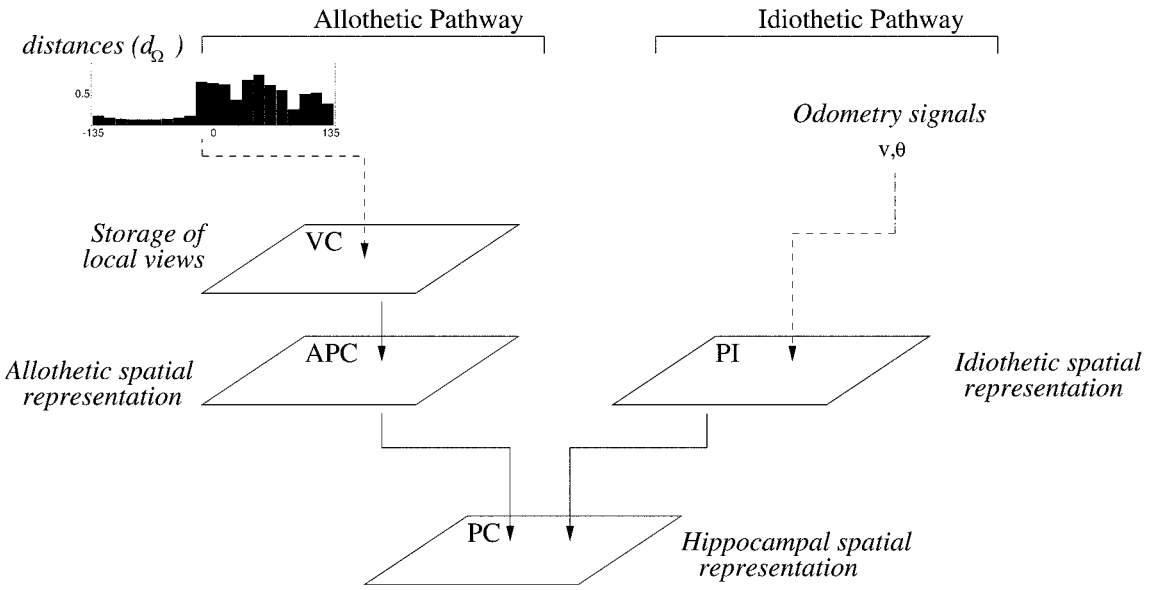

Fig. 2. Functional diagram of the hippocampal model. Dashed lines denote algorithmic transformation of the sensory information, solid lines denote projections between populations. The population of hippocampal PCs yield a representation of space used by the locale strategy. PI denotes a representation of space based on PI using proprioceptive odometry signals. APC denotes a spatial representation by allothetic PCs, driven by VCs encoding the local view.

See Arleo and Gerstner (2000) for a detailed description of the hippocampal model.

Figure 2 presents a functional diagram of the module for locale navigation. External stimuli are encoded in a population of view cells (VCs), which project to another population (referred to as *allothetic place cells* [APCs]) in which a representation of space based on external input is built. The transformation from external input to a first representation of space corresponds to the allothetic pathway leading to the hippocampus. A second pathway is based on idiothetic (i.e., proprioceptive) information. We refer to the ability of rodents to navigate using self-motion information (i.e., odometry and vestibular inputs) as *path integration* (PI) (McNaughton et al., 1996; Etienne et al., 1998; for a review *see* Etienne and Jeffery, 2004). It allows the rat to navigate in darkness or in absence of visual cues. In our model, we simply assume that such a representation based on PI exists. Both allothetic (APC) and idiothetic (PI) populations project onto the hippocampal population of place cells (PCs). The components of the locale system will be described in more detail in the following paragraphs.

Idiothetic Input to the Hippocampus

In our model, the idiothetic representation of space, encoded in the PI population, is implemented by an uniformly distributed population of cells with preconfigured metric relations. If the agent is moving at speed $v(t)$, we estimate its position $p^{\mathrm{PI}}(t) = \left[x(t), y(t)\right]$ by integration starting from the previous estimate $p^{\mathrm{PI}}(t-1)$,

$$x(t) = x(t-1) + \int_{t-1}^{t} v(t')\cos[\theta(t')]\mathrm{d}t'$$

$$y(t) = y(t-1) + \int_{t-1}^{t} v(t')\sin[\theta(t')]\mathrm{d}t' \qquad (1)$$

Proprioceptive information provides the speed (v) of the agent, and we assume that vestibular

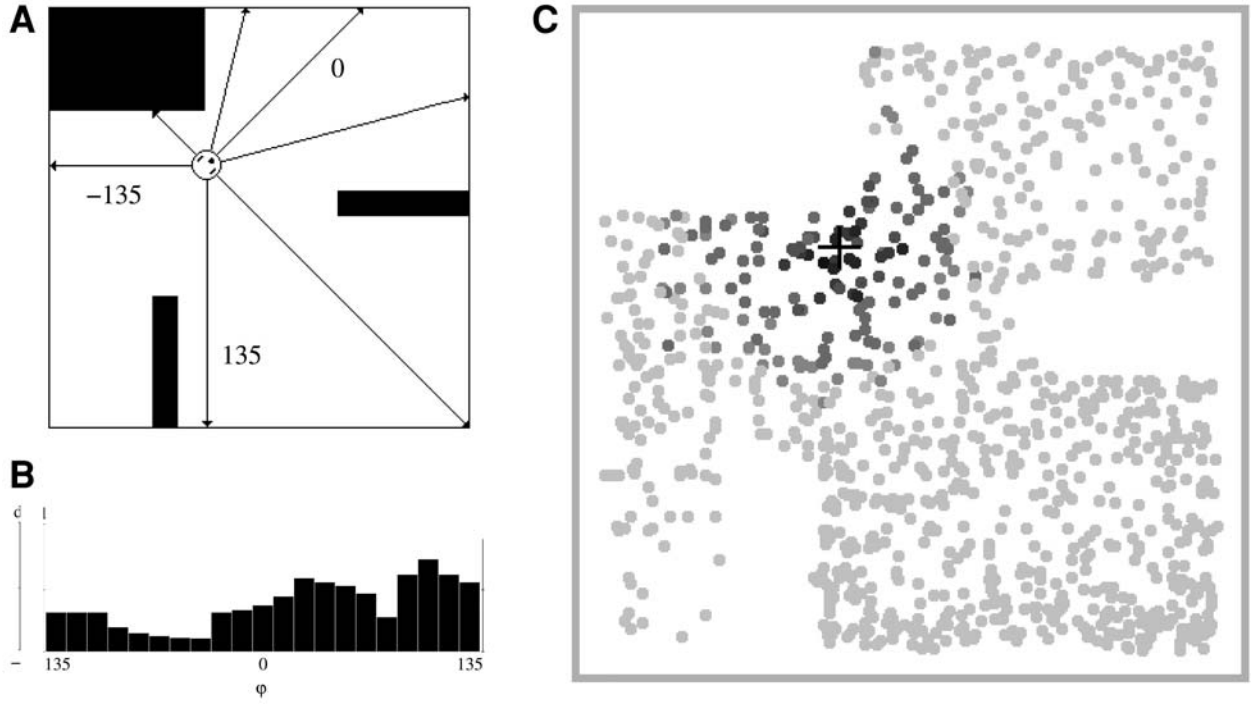

Fig. 3. **(A)** Agent situated in a simulated environment with three obstacles (black boxes). The large arrow marked "0" denotes the current direction of gaze. **(B)** Local view as perceived by the agent. The features (d_ψ) stored by the VCs are the normalized distances to the walls in 21 different directions ψ in the visual field (270°). **(C)** Activity of the hippocampal population. Each cell is represented by a dot located at the center of its place field. Highly activated cells are represented by dark dots. The cross marks the position of the robot.

information can be used to obtain the angle θ[4]. After a movement, the activity r_j^{PI} of each cell j in the PI module is updated according to the following:

$$r_j^{\mathrm{PI}} = \exp\left[-\frac{\left(p^{\mathrm{PI}} - p_j^{\mathrm{PI}}\right)^2}{2\sigma_{\mathrm{PI}}^2}\right] \qquad (2)$$

where p_j^{PI} is the center of the field of the cell j, σ_{PI} its width, and $p^{\mathrm{PI}} = \left[x(t), y(t)\right]$ is the updated estimate based on path intersection (Eq. 1).

[4]A set of neurons in the hippocampal formation, termed head direction cells, codes for directional information and can be seen as an allocentric compass of the animal (Taube et al., 1990; Blair and Sharp, 1995; Sharp et al., 1995). A neural model for the head-direction system has been implemented previously by Arleo and Gerstner (2001).

Allothetic Input to the Hippocampus

Behavioral and physiological experiments suggest that the hippocampal spatial representation is sensitive to visual cues (Muller and Kubie, 1987). The geometric properties of the environment, extracted by visual information, play an important role in the establishment of the spatial representation (O'Keefe and Burgess, 1996). Similar to the model proposed by Burgess et al. (2000), we encode the external input to the hippocampal model in a population of VCs using the distances (d_ψ) to the closest wall in $N_{\mathrm{dir}} = 21$ different directions (ψ) in the visual field.

At each location the agent takes a view of its environment and stores it in a newly recruited VC. The features of the local view as perceived by the agent in a simulated environment are presented in Fig. 3. Encoded local

views are aligned according to an allocentric frame of reference (which can be provided by the head direction system [Arleo and Gerstner, 2001]). The response of a VC cell m depends on the comparison between the features of the current local view (d) after alignment, and the stored features (d^m). As views are taken in different directions, only the $N_\Omega \leq N_{\text{dir}}$ features in the overlapping region (Ω) of the view field are taken into account for the comparison

$$r_m^{\text{VC}} = \exp\left[-\frac{\left(\frac{1}{N_\Omega}\sum_{\psi\in\Omega}|d_\psi^m - d_\psi|\right)^2}{2\sigma_{\text{VC}}^2}\right] \quad (3)$$

VCs project downstream onto the population of APCs that code for the location using only visual information. The firing rate of a postsynaptic APC i is computed as

$$r_i^{\text{post}} = g\left(\frac{\sum_j w_{ij} r_j^{\text{pre}}}{\sum_j w_{ij}}\right) \quad (4)$$

where g is a piecewise linear function $g(x) = x$ for $0 < x < 1$; $g(x) = 1$ for $x > 1$ and zero otherwise.

A two-step Hebbian learning procedure is applied to the connections from VCs to APCs. First, each time a VC is recruited, a new APC is selected and set to maximal rate $r_i^{\text{post}} = 1$. Connections from all active VCs' j with activity $r_j^{\text{pre}} > \theta$ to APC i are updated to a value $w_{ij} = r_j^{\text{pre}}\, r_i^{\text{post}}$. Second, after initialization unsupervised Hebbian learning is applied to the projection weights w_{ij} in order to allow the integration of information from several local views into a single APC. Specifically, a synapse from a presynaptic VC j to a postsynaptic APC i changes according to the following:

$$\Delta w_{ij} = \eta r_i^{\text{post}}(r_j^{\text{pre}} - w_{ij}) \quad (5)$$

Synapses with weight $w_{ij} = 0$ are considered as nonexisting and are not updated.

Combining Allothetic and Idiothetic Information

The two different representations of space discussed above, driven by external and proprioceptive inputs, are located in APC and PI populations, respectively. These two populations project onto the hippocampal population, in which the visual information and the proprioceptive information are combined.

During exploration, new PCs are recruited and connected to simultaneously active AP and PI cells. The activity of a hippocampal cell r_i^{PC} is computed using Eq. 4, with presynaptic cells in both APC and PI populations. Those connections are modified by means of the Hebbian learning rule (Eq. 5) in order to integrate both idiothetic and allothetic information.

Action Selection in the Locale Strategy

PCs in our model of the hippocampus project to a population of $N_{\text{AC}} = 36$ Action cells (ACs), with activity

$$a_i^{\text{L}} = \sum_j w_{ij} r_j^{\text{PC}} \quad (6)$$

where w_{ij} is the connection weight and r_j^{PC} is the firing rate of the hippocampal cell j. We think of each cell i as representing a direction of movement $\varphi_i = 2\pi i/N_{\text{AC}}$, and the population activity of all 36 cells encodes the direction (Φ_{L}) the agent should take based on a pure locale strategy. The learning mechanism (i.e., updating of weights w_{ij}) will be detailed later (the subsection "Updating the Modules").

Taxon Navigation Strategy

Taxon strategies rely on associating a cue to a specific response. As in the case of the hippocampal model, actions are encoded in a population of 36 ACs, driven in this case by a population of sensory input (SI) cells. SI consists of a horizontal one-dimensional grayscale image (I). Each sensory cell i has a narrow receptive field pointing in direction φ_i and its activity corresponds to the normalized grayscale

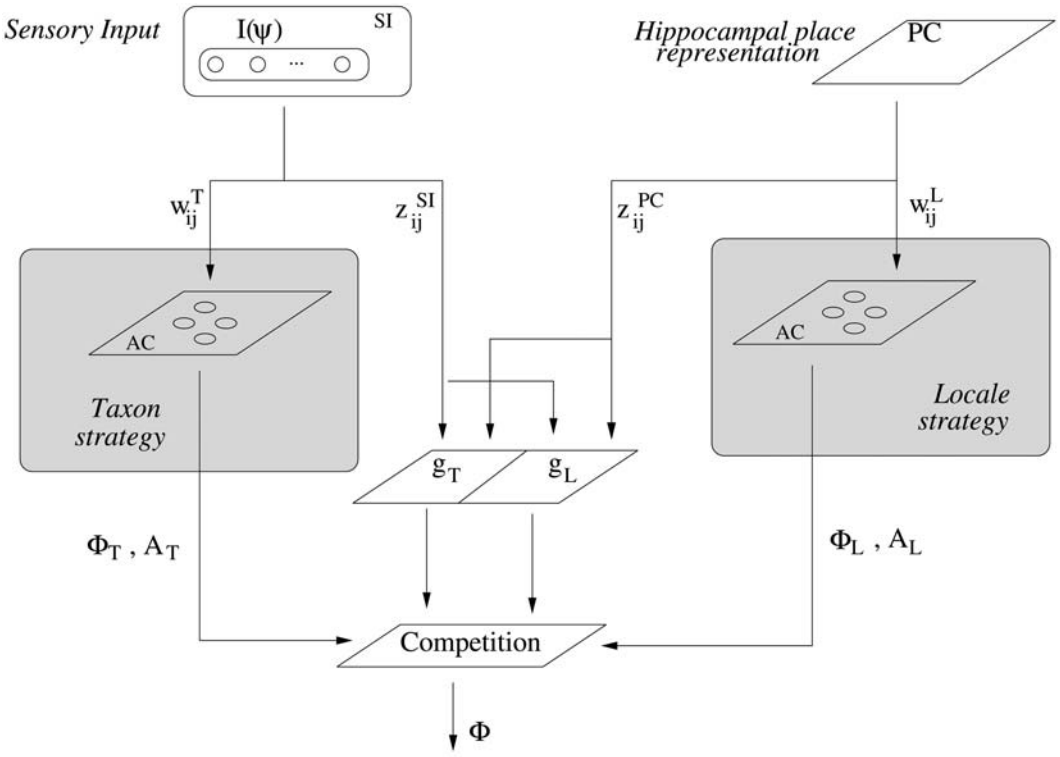

Fig. 4. Architecture of the model of navigation. PC = Spatial information coming from place cells in the hippocampus. SI = Sensory input (I_ψ). AC = Action cells. A_T = Action value (predicted future reward) of the movement Φ_T selected by the taxon strategy. A_L = Action value of the direction of movement Φ_L selected by the locale strategy. g_T,g_L = Gating values for the taxon and locale strategy, respectively. Φ = Direction of movement selected by the gating network.

value in that direction, $r_i^{SI} = I(\varphi_i)$. A visual cue in direction φ_i is represented as a dark spot in the image, that is, $I(\varphi_i) = 1$.

Actions based on the taxon strategy are then computed in analogy with Eq. 6,

$$a_i^T = \sum_\varphi w_{i\varphi} r_\varphi^{SI} \qquad (7)$$

Now we proceed to present the gating system and the learning algorithm for both modules in detail.

Strategy Selection

The two modules described above implement locale (L) and taxon (T) strategies, respectively. Here we describe the selection mechanism that decides which strategy will be taken by the agent and, according to that decision, modulates the learning process in these modules. The selection mechanism consists of two experts, each expert corresponding to one of the strategies, and a gating network is used to select the most appropriate strategy according to the external input and the internal state of the system.

The selected action of the system is the result of a competition among the experts. For an expert $k \in \{L, T\}$ (for locale and taxon strategies, respectively), the probability of being selected is a function of a gating value g_k. Both the experts and the gating network modify their parameters by means of reinforcement learning.

We think of one of the experts as being located in the ventral and dorsomedial striatum driven by the hippocampal activity. The

activity of the hippocampal PCs represents the current state *s*, and the expert learns a mapping between this hippocampal spatial representation (PC) and a set of actions, implementing a place-based strategy (i.e., locale strategy). The other expert receives external sensory signals (SI) and encodes simple stimulus–response behaviors (taxon strategy). The biological locus for this expert may be the dorsolateral striatum.

As described in the subsection "Action Selection in the Locale Strategy," actions are encoded in a population of ACs for each expert. An AC i in population k represents a direction of movement φ_i^k and its activity a_i^k corresponds to the action value or Q-value in the sense of reinforcement learning (Sutton and Barto, 1998). The action values for both PC-driven (a^{L}) as stimulus-driven (a^{T}) strategies are computed in the following way:

$$a_i^k = \sum_j w_{ij}^k r_j^{\mathrm{pre}} \tag{8}$$

where r_j^{pre} corresponds to the firing rate of cells coding for the external SI or the hippocampal PCs for k being the taxon or the locale module, respectively. Equation 8 repeats Eqs. 6 and 7, with slightly different notation.

For each expert k, the direction of the population vector Φ_k represents the continuous direction of movement, which predicts the maximum reward.

$$\Phi_k = \arctan\left(\frac{\sum_i a_i^k \sin(\varphi_i)}{\sum_i a_i^k \cos(\varphi_i)}\right) \tag{9}$$

As the Q-value is only available for discrete actions a^k, the action value A_k for the continuous direction of movement Φ_k is computed by linear interpolation of the two nearest discrete actions.

In order to select the direction Φ of the movement to be performed by the agent, we use a gating scheme such that probabilities depend not only on the Q-values of the actions (A_k) but also on a gating value (g_k):

$$P(\Phi = \Phi_k) = \frac{g_k A_k}{(g_{\mathrm{L}} A_{\mathrm{L}}) + (g_{\mathrm{T}} A_{\mathrm{T}})} \tag{10}$$

The gating system allows a module to be preferred even if it predicts a smaller future reward. In the other sense, it allows taking *opportunistic* choices such as selecting actions with small gating value, but predicting high rewards. Gating values g_k depend on both the hippocampal input PC and the sensory input SI according to the following:

$$g_k = \sum_{j \in \mathrm{PC}} (z_{kj}^{\mathrm{PC}} r_j^{\mathrm{PC}}) + \sum_{j \in \mathrm{SI}} (z_{kj}^{\mathrm{SI}} r_j^{\mathrm{SI}}) \tag{11}$$

where z_{kj}^{PC} and z_{kj}^{SI} are the weights of a connection from a presynaptic cell j to a gating unit k.

The system described so far is able to select among different strategies according to its perceptual input. At each timestep, every expert proposes an action on the basis of its afferents and gating value. However, a strategy should be applied during several timesteps in order to exhibit a coherent behavior and better assess its suitability for the current task. To allow this, instead of imposing a competition at every timestep, a chosen strategy will continue till its accumulated prediction error (since the moment it was chosen) reaches some threshold. Then, a new competition among the experts is performed (allowing, but not forcing, the selection of a different strategy).

Updating the Modules

After an action has been selected (according to Eq. 10) and performed, the weights for each module (w_{ij}) as well as the gating network (z_{ij}) are updated according to the reward prediction error δ_k. In order to allow simultaneous learning of both strategies, the two modules are updated in a way that modules with high probability of being selected have more significant changes in their weights, as well as those modules with small reward prediction error.

In standard reinforcement learning, weight updates are proportional to the reward prediction error δ_k. According to the above considerations, the proportionality factor has to be scaled by an extra factor h_k. This factor depends on the gating value g_k and the prediction error for the module k (Baldasarre, 2002):

$$h_k = \frac{g_k c_k}{\sum_i (g_i\ c_i)} \tag{12}$$

where $c_k = \exp(-\rho\delta_k^2)$ $(\rho > 0)$. Weights in each module (w^{L} and w^{T}) in Eq. 8 are updated using a variation of TD (λ) algorithm (Sutton and Barto, 1998).

$$\Delta w_{ij}^k = \eta^k\ \delta_k\ h_k\ e_{ij}^k \tag{13}$$

with

$$\delta_k(t) = R_t - A^k(t-1) + \gamma A^k(t) \tag{14}$$

Here R_t is the reward received at time t, η^k is the learning rate of expert k, δ_k is the reward prediction error from Eq. 14, and e_{ij}^k is the eligibility trace. This eligibility trace can be interpreted as a memory of temporal pre- and postsynaptic coincidences.

$$\begin{aligned} e_{ij}^{\mathrm{L}}(t+1) &= \gamma\lambda e_{ij}^{\mathrm{L}}(t) + r_j^{\mathrm{PC}} a_i^{\mathrm{L}} \\ e_{ij}^{\mathrm{T}}(t+1) &= \gamma\lambda e_{ij}^{\mathrm{T}}(t) + r_j^{\mathrm{SI}} a_i^{\mathrm{T}} \end{aligned} \tag{15}$$

where $0 < \lambda < 1$, is the trace-decay factor and $e_{ij}^k(0) = 0$.

The use of the scale factor h_k in Eq. 13 assures that the weight update is more significant for those experts that have consistently small reward prediction errors ($c_k \approx 1$) and have a high probability of being selected ($g_k > g_i$). Note that the experts modify their weights such that, even if an expert is not selected, it can improve its performance in the current task.

Updating the Gating Network

The weights $z_{kj}^{\mathrm{PC,SI}}$ in Eq. 11 are updated such that $g_k \to h_k$.

$$\Delta z_{kj}^{\mathrm{PC,SI}} = \xi(h_k - g_k) r_j^{\mathrm{PC,SI}} \tag{16}$$

with a learning rate ξ. This will increase the gating value of the module showing smaller reward prediction error, and since $\Sigma_k h_k = 1$ (from Eq. 12), after learning the sum of gating values will also tend to 1.

Results

We tested the model on a simulated Kephera robot with a visual field of 270°. The robot has a diameter of 5.2 cm, and at each timestep it moves 6 cm in the direction Φ provided by the action controller. Other parameters (e.g., collision detectors) were set according to the specifications of the Kephera robot. The simulation environment provides external SI (linear vision and distances to the walls) and odometry information from the agent. Before each experiment all the weights are randomly initialized. Four sets of experiments were performed in a square environment of 120 × 120 cm. The target has a diameter of 12 cm. We test the model capabilities to (i) solve a task using a taxon strategy (i.e., the visible water maze), (ii) solve a task requiring a locale strategy (i.e., hidden water maze), and (iii,iv) develop both strategies simultaneously (i.e., combined cue-place learning). These two last tests follow experimental paradigms reportedly applied to rats (Pearce et al., 1998; Devan and White, 1999).

All the experiments consist of 10 to 11 blocks, with four trials per block. In the combined cue-place learning task, the 10th block corresponds to the competition trials as described by Devan and White (1999). Each trial starts with the agent located at a random position at a minimum distance of 70 cm from the center of the goal (requiring the robot to make more than 10 steps to get at the goal location). A trial is finished when the agent reaches the target location; at that moment a positive reward in given. If the agent is not able to reach the goal within 100 timesteps, it is guided to the platform,

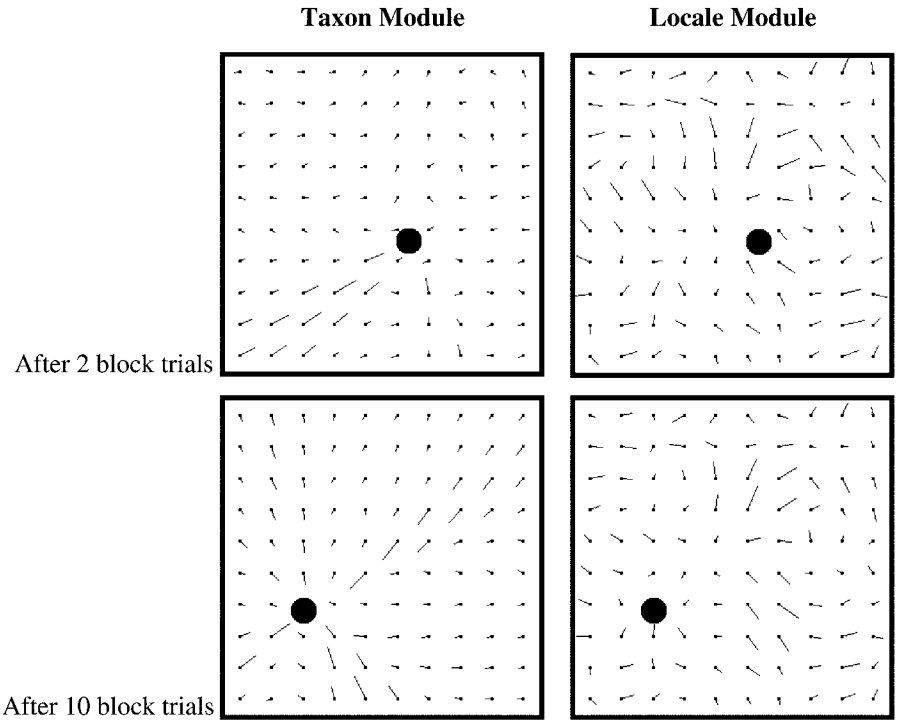

Fig. 5. Navigation maps after 2 (*top*) or 10 (*bottom*) blocks of training in the visible version of the water maze. *Left*: Taxon module. *Right*: Locale module. The filled circle marks the location of the goal.

analogous to the experimental procedure with rats (Devan and White, 1999).

The same set of parameters was used in all the simulations. The learning rates for the locale and taxon modules are set to $\eta^L = \eta^T = 0.01$. The parameters of the TD (λ) rule (Eqs. 13–15) are $\gamma = 0.8$ and $\lambda = 0.95$. The learning rate for the gating network is set to $\xi = 0.3$ (Eq. 16). A value of $\rho = 1.0$ was used to compute the factor c_k in Eq. 12.

In the first set of experiments, the visible version of the water maze is modeled by placing the agent in an environment with white walls and a dark, visible cue in the place where a reward is provided. The location of the platform is changed in every trial such that spatial information cannot be used to solve the task. The navigation maps for both the locale and the taxon module after learning are presented in Fig. 5. At each point the lines show the selected direction Φ_k for that specific module, the length of the line is proportional to the action value A_k associated to a movement in that direction. The lines show that the taxon module is able to learn the association between the SI and the appropriate action to solve the task. As the spatial information is not useful to reach the visible goal, the locale module is not able to build a proper association between location and action.

The second set of experiments tests the ability of the system to develop locale strategies in order to find a hidden goal. The goal is solely defined by a fixed location of the reward. In particular, there were no visual cues signaling its position. The only visual cue is at a fixed position outside the environment. Navigation maps for both modules are presented in Fig. 6. In contrast to the previous case, the locale module is able to learn the correct action required to take the agent toward the invisible goal. As no cue is available at the location of the goal, the taxon module cannot solve the task.

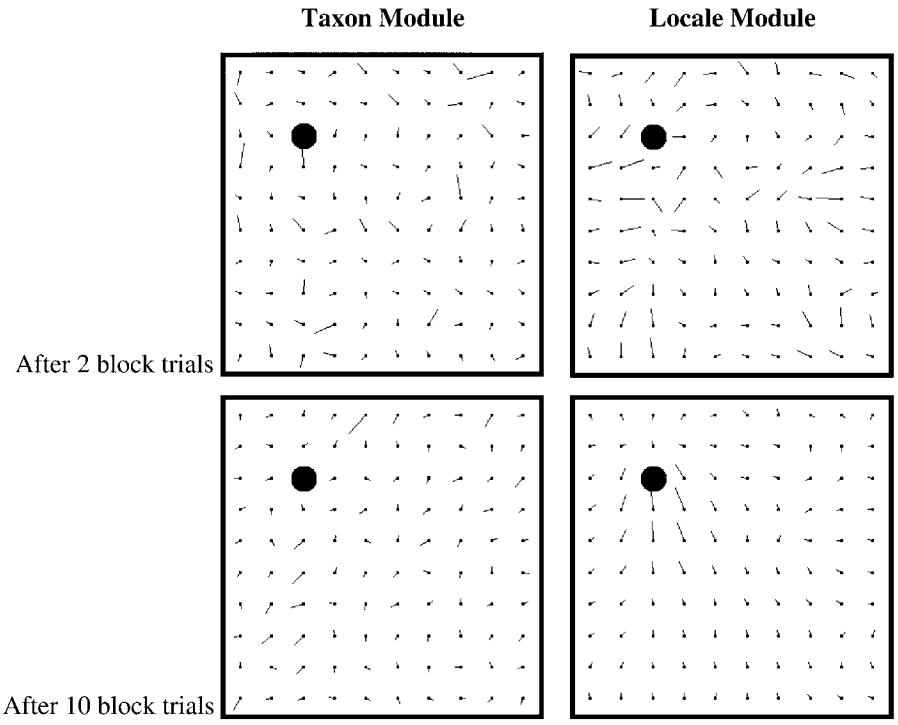

Fig. 6. Navigation maps after 2 (*top*) or 10 (*bottom*) blocks of training in the hidden water maze. *Left:* Taxon module. *Right:* Locale module.

Finally, we apply the experimental paradigm described in the section "Parallel Navigation Systems" (Devan and White, 1999), to test the simultaneous development of both locale and taxon strategies. During training in an environment with white walls, in blocks 1, 2, 4, 5, 7, and 8, the goal location was signaled by a dark visible cue, such that the task may be solved by either a locale or a taxon strategy. On blocks 3, 6, and 9, no visible cue was available, requiring the use of spatial information to solve the task.

Figure 7 shows the navigation maps for both the locale and the taxon modules. It can be seen that, after training, both modules are able to guide the agent toward the goal location (filled circle). It should be noticed that both modules have simultaneously learned to solve the task by using a different type of information. This becomes clear once the visible goal is moved to another location (Fig. 7, bottom) and the taxon strategy (cue response) guides the agent to the new location, whereas the locale module still points toward the location of the goal in the training phase.

The predicted reward (normalized to one) for both modules at different stages of training is shown in Fig. 8. It corresponds to the action value A_k of the action selected at different points in the environment (as shown in Fig. 7). At the end of block 8 (the last block with the visible platform), both modules successfully predict the location where the reward is delivered. As shown above, before the competition trial, spatial information still leads the agent toward the previous location of the platform. The taxon module, in contrast, successfully points toward the landmark (and goal) location. If training continues in the competition situation, the locale module will start learning the new location of the goal.

The average escape latency (timesteps required to reach the goal) over 10 simulations

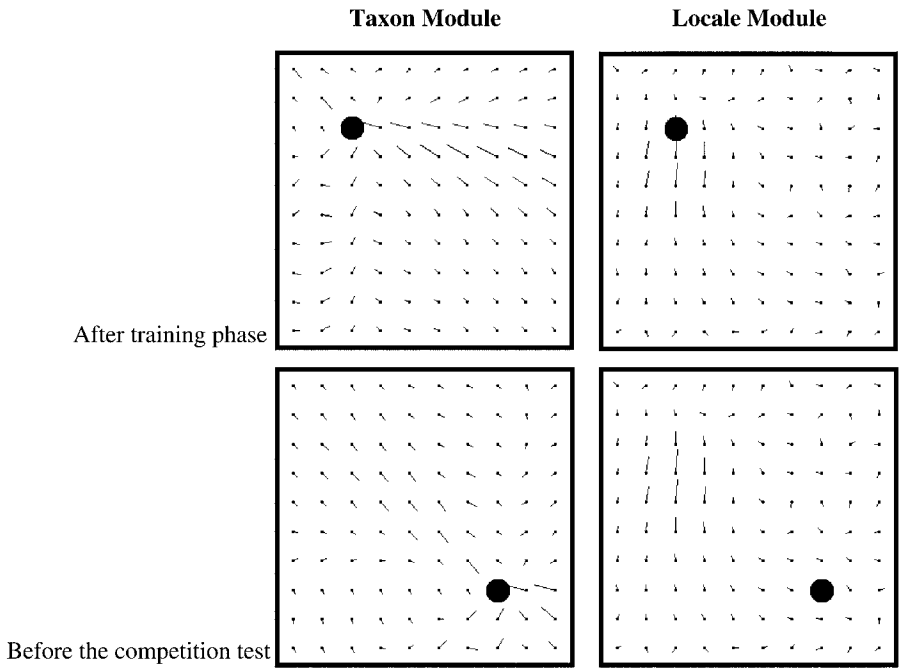

Fig. 7. Navigation map at the end of the training phase. *Top:* The filled circle marks the location of the platform in both the hidden and visible training trials. *Bottom:* Competition trial. The visible target has been moved to a new location. *Left:* Taxon strategy. *Right:* Locale strategy.

is showed in Fig. 9. The simultaneous development of taxon and locale strategies can be observed as the model improves its performance in both visible (blocks 1, 2, 4, 5, 7, and 8) and hidden (blocks 3, 6, and 9) trials. This result qualitatively reproduces those reported by Devan and White (1999).

The competition trial in the Devan and White experiment was designed to dissociate the response to the different kinds of information (i.e., the use of different navigational strategies). Representative trajectories in the competition trial for two runs of the simulation are shown in Fig. 10. As observed in intact animals (Fig. 1), in some cases the agent goes first to the place where the platform was during the training (place response), and then switches to a taxon strategy, whereas in other cases, the agent adopts a taxon strategy from the beginning and swims directly toward the visible goal.

Figure 11 presents the probability after training for every module to be selected in the experiments shown above (average over 10 simulations). When the agent is solving the visible maze, the gating system is more likely to select the taxon module than the locale module, as opposed to the hidden maze condition in which the locale strategy is preferred. In the case of the combined cue-place task, after training (block 9) both strategies can solve the task. Hence, at that point of training both modules have almost equal probabilities of being selected. Once the agent is exposed to the conflict condition (competition test), the place response is no longer suitable to solve the task. This causes a change of strategy in the middle of the task in those cases in which a locale strategy was chosen at the beginning of the trial (Fig. 10). Finally, if the agent is exposed during several trials (three extra training blocks in this case) to the new condition, and the locale strategy persistently fails in solving the task, the probability of selecting the taxon module will increase.

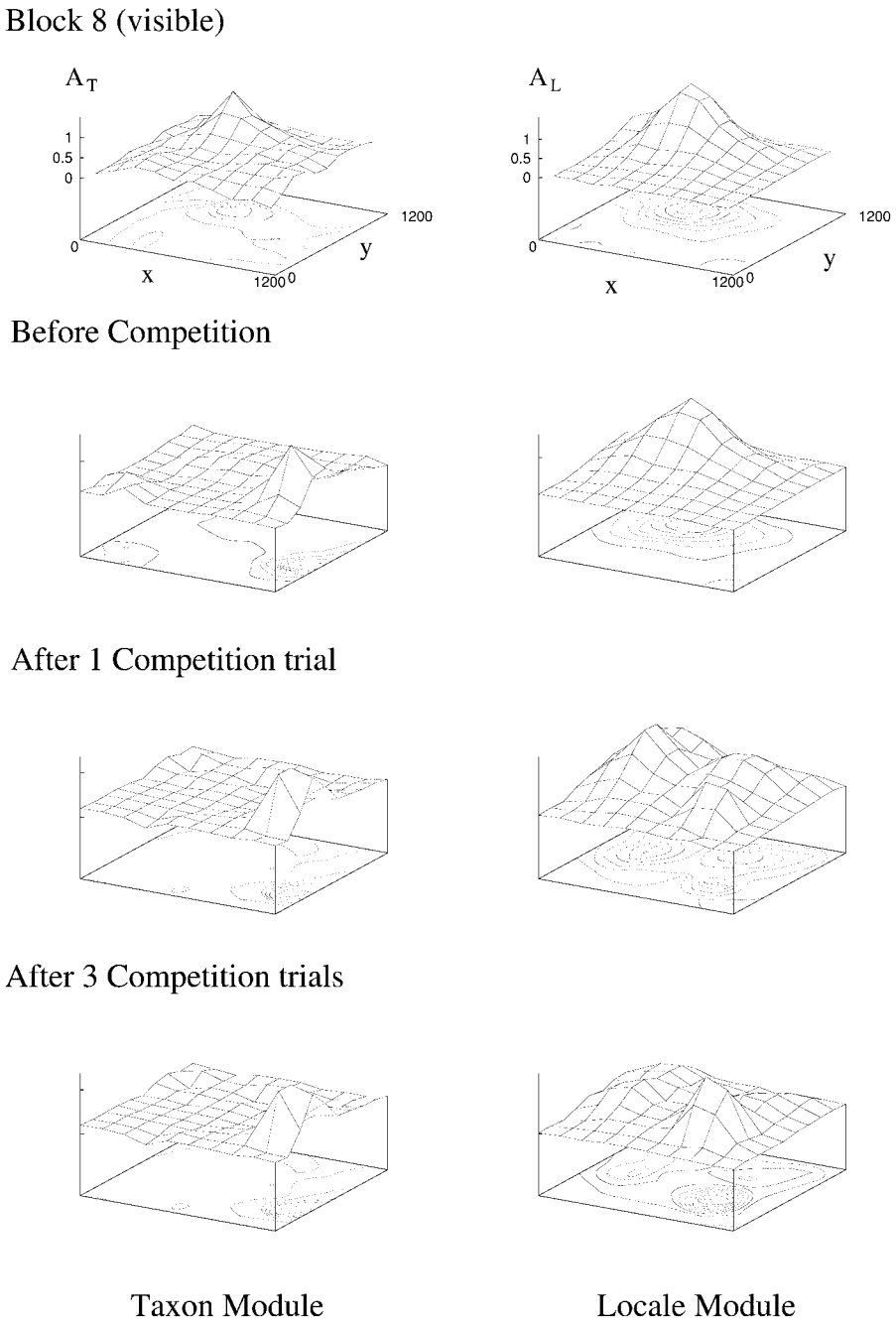

Fig. 8. Reward predicted by each module at different stages of training. *From top to bottom*: End of block 8, last training block with the visible goal. Both modules predict the maximum reward at the location of the goal. Before competition, the taxon module predicts the maximum reward at the new location of the landmark. The locale strategy still leads to the same location as before. After 1 and 3 competition trials, the locale strategy starts to learn the new location but still predicts a high reward for the former location of the platform.

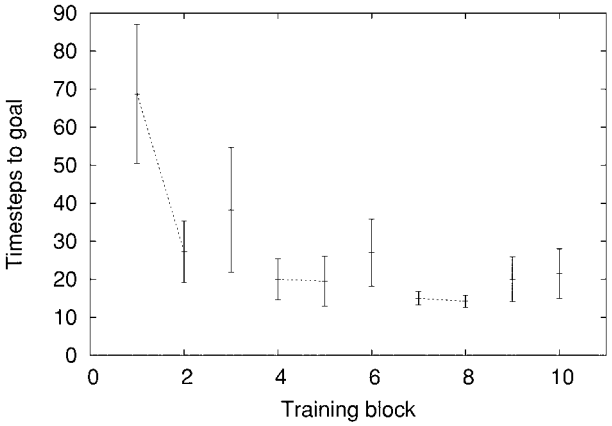

Fig. 9. Average escape latency in the combined cue-place learning task (Devan and White, 1999). Starting positions on every trial are selected such that it takes the agent at least 10 steps to reach the target location. Training blocks 3, 6, and 9 correspond to trials using a hidden goal. The competition trail is presented at block 10.

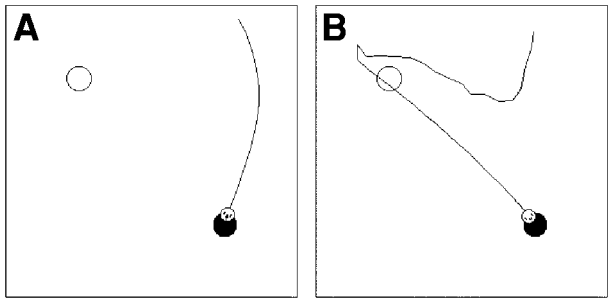

Fig. 10. Example of trajectories observed during the competition test. **(A)** The agent chooses a taxon strategy since the beginning of the trial (cue responder). **(B)** The agent follows a locale strategy before going toward the visible goal (place responder).

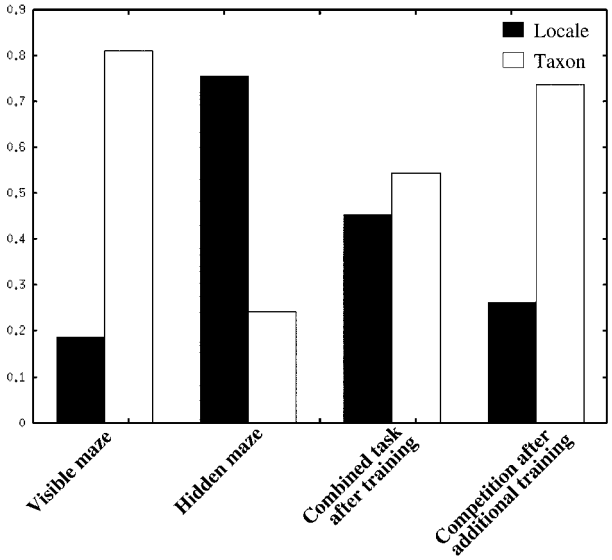

Fig. 11. Probabilities for the different modules of being selected. The taxon module has a higher probability of being selected when the agent solves the visible maze and, conversely, the locale module is preferred in the case of the hidden maze. After nine blocks of training in the combined cue-place task, both modules have rather equal probabilities of being selected, as both strategies are able to solve the task. The last two columns show the probability of selecting taxon or locale strategy after three blocks of trials with the competition setup; that is, the platform is visible but at a location different from the one encountered during the previous training phase. Since the learned place response does not solve the task, the probability of selecting the taxon strategy increases.

In another experiment, Pearce et al. (1998) train rats to search for a platform located at a fixed distance and direction from a single landmark. The training consists of 11 blocks of four trials. At the beginning of each block both the landmark and the platform were moved to a different location, and remained in the same place for the entire block session (Fig. 12A). In this experiment the landmark gives correct, although not precise information about the location of the goal; whereas place information will be disrupted every time a new block starts. Consistent with this, animals with hippocampal lesions perform better in the first trial of each block than control animals, presumably guided by the landmark location. At every new block, intact animals tend to explore the region of the pool where the platform was previously located before swimming to the new location, indicating the use of spatial information to solve the task (i.e., locale strategy). Lesioned animals, in contrast, take a more direct path to the platform (and landmark) location.

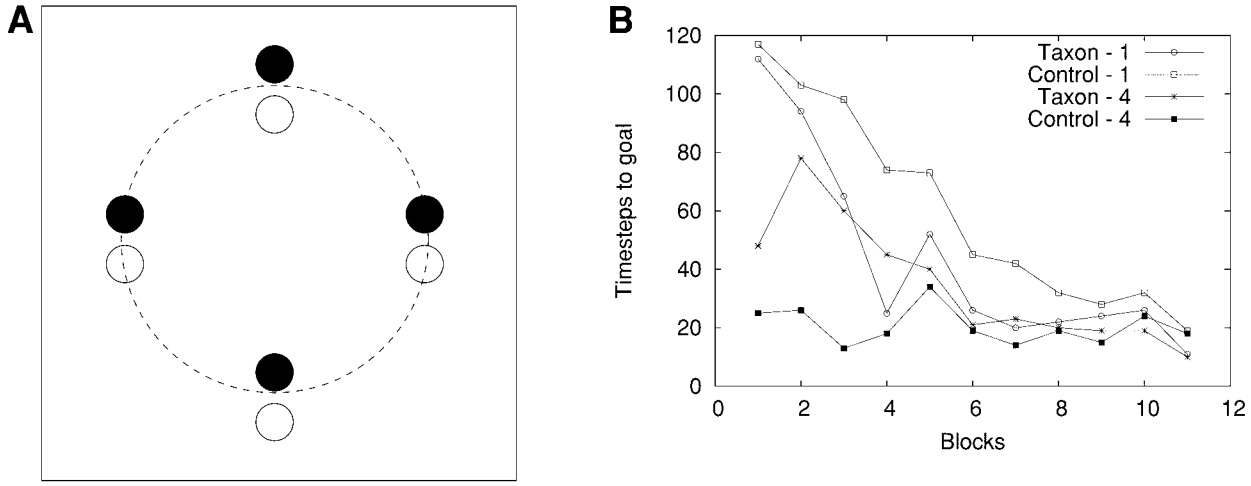

Fig. 12. Experimental paradigm analogous to the one used by Pearce et al. (1998) **(A)** The experiment takes place in a square arena (120 × 120 cm) with a single landmark. The landmark can be located in one out of four possible positions (filled circles). The goal (open circles) is always located at a fixed distance and direction from the landmark. **(B)** Mean latency in the first (*Control-1* and *Taxon-1*) and fourth (*Control-4* and *Taxon-4*) trials of each training block. Control: Actions are selected by the competition of locale and taxon strategies. Taxon: Actions are selected using a pure taxon strategy (as described in the subsection "Taxon Navigation Strategy").

Despite its poor performance in the first trial, control animals significantly improve their performance during each block of training whereas rats with damage to the hippocampus do not. This suggests that lesioned animals rely on the less precise information from the landmark location (a stimulus–response behavior) whereas control rats also use spatial information to solve the task. Every time a new block starts, as spatial information is no longer suitable to reach the platform, the rat changes its navigation strategy and uses the information provided by the landmark location, similar to the place responder behavior reported by Devan and White (1999).

Figure 12B shows the results of our model using this paradigm. We compare the performance of the competition model (*Control*) with the results of using a purely taxon strategy (*Taxon*) as described in the subsection "Taxon Navigation Strategy." The figure shows the average latency over ten simulations for the first and last trial of each block. The performance of the competition mechanism is consistent with the results reported by Pearce et al. (1998) for control animals; it yields longer latencies for the first trial of each block but latencies significantly decrease during the three subsequent trials. The performance of a pure taxon strategy is not negatively affected by the start of a new trial, but it does not improve during the training block.

Discussion

We propose a system-level model of navigation able to reproduce the behavior of animals in conflict situations, using a model of multiple experts with reinforcement learning. It relies on the theory of parallel navigating systems competing for selection of the most appropriate action. Based on the results of neurobehavioral experiments, we propose those systems to involve the hippocampus and the ventral striatum for locale strategies (Morris, 1981; Redish, 1999), and the dorsolateral striatum implementing taxon strategies (Packard and Knowlton, 2002). The strategy selection will depend on the characteristics of

the task to be solved and the training process (White and McDonald, 2002).[5] We attempt to focus on the computation underlying the strategy selection in navigational tasks. This selection is accomplished as a competition among different modules, and each module implements a different strategy. The likelihood of a module to be selected for action control depends on the future reward it predicts and the accuracy of this prediction.

One of the assumptions of the model is that predicted reward (A_k in Eq. 10) influences the decision-making process, such that actions predicting high future rewards can be selected, even if they are less likely to be successful (small g_k). If this is the case, the starting position can influence the strategy selection in the competition trial (i.e., at locations closer to the cued target the taxon module will predict higher reward than the locale module, favoring the preference for cue responses). In their experiment, Devan and White (1999) chose starting locations equidistant to the new target location and its former position (the one used during training). A similar experiment with systematic changes in the starting location on competition trials can give qualitative information about the relative importance of the predicted reward (A_k) and biasing mechanisms depending on the experience (g_k) in the selection of navigation strategies. According to our model, starting positions close to the location of the platform during training will favor the selection of locale strategies. Figure 13 shows the probability of selecting a locale strategy after one competition trial. At this stage, both strategies have almost the same probability of being chosen (average gating value $g_L = 0.45$), but near to the former location of the goal ($x = 300$, $y = 800$), the high reward

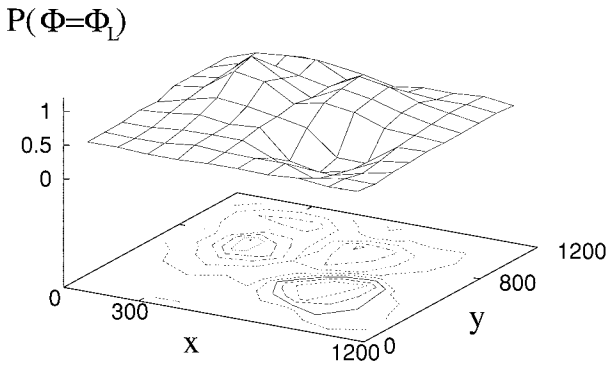

Fig. 13. Probability of selecting a locale strategy after one competition trial at different positions in the arena. At this point, owing to the training procedure, both strategies have approximately equal gating values and the strategy selection depends mainly on the reward predicted by each module. Therefore, the probability of selecting one strategy will be influenced by the starting position (the closer it is from the expected location of the goal, the greater the reward). In this case, the probability of selecting the locale strategy is maximal near the location of the goal during training.

predicted by the locale module results in a high probability [$p(\Phi = \Phi_L) > 0.80$, around the goal location] of selection for this strategy.

The model implements this biasing mechanism as a gating network. Gating values are learnt such that modules showing small reward prediction error will have more probability of being selected for action control (Eqs. 12 and 16). Furthermore, the gating value also modulates the learning rate of each module (Eq. 13), such that a more significant update will be done in modules with high gating values, and consistently small prediction errors.

Even though the biological mechanism for the competition has yet to be clarified, recent experiments suggest that cholinergic activity might convey information alike to the gating signals in our model. Several experiments provide evidence for the cholinergic role in modulating learning and memory (Ragozzino and Gold, 1995; Ragozzino et al., 1998; Hasselmo, 1999; Chang and Gold, 2003), and it has been

[5]Endogenous factors such as stress (Kim and Baxter, 2001), hormonal status (Marriott and Korol, 2003), or motivation (White and McDonald, 2002; Mizumori et al., 2004) also affect the selection of learning strategies.

proposed that the role of acetylcholine (ACh) is to balance the contribution of different neural systems in learning a given task (Gold, 2003, 2004). Following this hypothesis, Chang and Gold (2003) measured simultaneously ACh levels in the hippocampus and the striatum of rats solving a cross-maze task. When trained in this task, animals gradually shift from locale strategies in earlier trials to response strategies as training continues (Packard and McGaugh, 1996). Consistently, at the onset of the training procedure, ACh release in the hippocampus rapidly increased to its asymptotic value, coinciding to the use of locale strategies, and remains at this level for subsequent trials. In contrast, cholinergic release in the striatum increased at a much slower rate, in a pattern related to the gradual shift toward response strategies.[6] In another experiment using the same task, McIntyre et al. (2003) report that the relative levels of ACh in the hippocampus and striatum *prior* training reliably predict how long the locale strategy will be preferred. These results give further support to the hypothetic role of ACh as biasing the preference for a given strategy, which may be related to the gating values g_k in our model. Cholinergic levels have also been proposed, in a statistical framework, to act as a gate to learning in a specific neural system by reflecting the uncertainty in its predictions (Dayan et al., 2000; Yu and Dayan, 2002, 2003).

In addition, ACh has been proposed as regulating the speed of memory update (Hasselmo and Bower, 1993; Hasselmo, 1999), which in modeling terms may correspond to the learning rate (Doya, 2002). This dual role of ACh biasing the strategy selection and modulating the learning speed in the different systems implementing each strategy is consistent with the role the scale factor h_k (Eq. 12) plays in the learning rules for the gating system (Eq. 16 updates the gating network such that $g_k \rightarrow h_k$), and the separate navigation modules (Eq. 13).

In the cross-maze experiment described by Chang and Gold (2003), both locale and response strategies lead the animal to the goal. Similarly, ACh levels in both hippocampus and striatum increased during training and remained at high levels till the end of the experiment. In contrast, the protocol by Devan and White (1999) includes hidden trials, requiring the use of hippocampal-dependent strategies, at different points of the training phase. If the release of ACh is assumed to be correlated to preference for one strategy (and its engagement in solving the task), we predict that striatal cholinergic release will not increase during those trials (blocks 3, 6, and 9). On the other hand, during the competition trial, even if the locale strategy leads to the wrong location, ACh levels in the hippocampus will remain at high values, as this structure keeps engaged in the task learning the new location of the goal.

The question of what mechanism regulates the relative levels of ACh in the different structures involved in learning remains unsolved. One possibility is that an input system (to be identified) controls the ACh release in the forebrain. However, neuroanatomical differences of the cholinergic system in the hippocampus and striatum[7] suggest that this regulation is more likely owing to presynaptic mechanisms of release within each neural structure (Gold, 2004).

In our model, by having modules driven by different input spaces (place and cue information)

[6] After extensive training, animals consistently follow a stimulus–response strategy, despite high levels of ACh in both the hippocampus and the striatum. A systematic preference for the striatal system seems to control the strategy selection when both strategies are equally good to solve the task. This constitutes and additional way of interaction among systems involved in navigation, not included in our model.

[7] ACh in the striatum is derived from intrinsic cholinergic neurons, whereas hippocampal ACh is derived from projecting neurons from the basal forebrain cholinergic neurons.

we are able to solve tasks requiring different navigation strategies. Bioinspired robots can use the same approach by maintaining different mappings of the perception–action relations. Action control in autonomous robots can be performed by choosing among reactive behavior in egocentric coordinates (e.g., Braitenberg-like obstacle avoidance, approaching a beacon) or trajectory planning in an allocentric frame of reference. The selection can be based on a competition mechanism like the one we proposed in this model, which takes into account how well each representation has performed in the past and the predicted outcome of the proposed actions. Conversely, robots can be used to test the validity of models of bioinspired navigation. The use of realistic sensory signals with the inherent noise associated to this input constitutes a powerful testbed for the robustness of the model. Previously, the locale navigation strategy has been already tested using realistic two-dimensional visual input and odometry information from a Kephera robot (Arleo et al., 2004). Future plans for our research include tests of the competition model using a similar setup, and a more complete set of behavioral paradigms and conflict situations [e.g., the cross-maze task described by Packard and McGaugh (1996)].

Acknowledgment

This work was supported by the Swiss National Science Foundation under Grant No. 200020-100265/1.

References

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

Arleo, A. and Gerstner, W. (2001) Spatial orientation in navigating agents: Modeling head-direction cells. Neurocomputing 38–40, 1059–1065.

Arleo, A., Smeraldi, F., and Gerstner, W. (2004) Cognitive navigation based on nonuniform gabor space sampling, unsupervised growing networks, and reinforcement learning. IEEE Trans. Neural Netw. 15, 639–652.

Baldasarre, G. (2002) A modular neural-network model of the basal ganglia's role in learning and selecting motor behaviours. Cogn. Syst. Res. 3, 5–13.

Blair, H. and Sharp, P. (1995) Anticipatory head direction signals in anterior thalamus: Evidence for a thalamocortical circuit that integrates angular head motion to compute head direction. J. Neurosci. 15(9), 6260–6270.

Burgess, N., Jackson, A., Hartley, T., and O'Keefe, J. (2000) Predictions derived from modelling the hippocampal role in navigation. Biol. Cybern. 83, 301–312.

Chang, Q. and Gold, P. E. (2003) Switching memory systems during learning: changes in patterns of brain acetylcholine release in the hippocampus and striatum in rats. J. Neurosci. 23, 3001–3005.

Da Cunha, C., Wietzikoski, S., Wietzikoski, E. C., et al. (2003) Evidence for the substantia nigra pars compacta as an essential component of a memory system independent of the hippocampal memory system. Neurobiol. Learn. Mem. 79, 236–242.

Dayan, P., Kakade, S., and Montague, P. R. (2000) Learning and selective attention. Nat. Neurosci. 3, 1218–1223.

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

Doya, K. (2002) Metalearning and neuromodulation. Neural Netw. 15, 495–506.

Etienne, A. S. and Jeffery, K. J. (2004) Path integration in mammals. Hippocampus 14, 180–192.

Etienne, A. S., Maurer, R., Berlie, J., et al. (1998) Navigation through vector addition. Nature 396, 161–164.

Gold, P. E. (2003) Acetylcholine modulation of neural systems involved in learning and memory. Neurobiol. Learn. Mem. 80, 194–210.

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

Hasselmo, M. E. (1999) Neuromodulation: acetylcholine and memory consolidation. Trends Cogn. Sci. 3, 351–359.

Hasselmo, M. E. and Bower, J. M. (1993) Acetylcholine and memory. Trends Neurosci. 16(6), 218–222.

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

Marriott, L. K. and Korol, D. L. (2003) Short-term estrogen treatment in ovariectomized rats augments hippocampal acetylcholine release during place learning. Neurobiol. Learn. Mem. 80, 315–322.

McIntyre, C. K., Marriott, L. K., and Gold, P. E. (2003) Patterns of brain acetylcholine release predict individual differences in preferred learning strategies in rats. Neurobiol. Learn. Mem. 79, 177–183.

McNaughton, B. L., Barnes, C. A., Gerrard, J. L., et al. (1996) Deciphering the hippocampal polyglot: the hippocampus as a path integration system. J. Exp. Biol. 199, 173–185.

Mizumori, S., Yeshenko, O., Gill, K., and Davis, D. (2004) Parallel processing across neural systems: Implications for a multiple memory system hypothesis. Neurobiol. Learn. Mem. 82, 278–298.

Morris, R. G. M. (1981) Spatial localization does not require the presence of local cues. Learn. Motiv. 12, 239–260.

Muller, R. and Kubie, J. (1987) The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. J. Neurosci. 7, 1951–1968.

O'Keefe, J. and Burgess, N. (1996) Geometric determinants of the place fields of hippocampal neurons. Nature 381, 425–428.

O'Keefe, J. and Nadel, L. (1978) The Hippocampus as a Cognitive Map. Clarendon Press, Oxford.

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

Packard, M. G. and Knowlton, B. J. (2002) Learning and memory functions of the basal ganglia. Annu. Rev. Neurosci. 25, 563–593.

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

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

Ragozzino, M. E. and Gold, P. E. (1995) Glucose injections into the medial septum reverse the effects of intraseptal morphine infusions on hippocampal acetylcholine output and memory. Neuroscience 68, 981–988.

Ragozzino, M. E., Pal, S. N., Unick, K., Stefani, M. R., and Gold, P. E. (1998) Modulation of hippocampal acetylcholine release and spontaneous alternation scores by intrahippocampal glucose injections. J. Neurosci. 18, 1595–1601.

Redish, A. (1999) Beyond the Cognitive Map, From Place Cells to Episodic Memory. MIT Press-Bradford Books, London.

Sharp, P. E., Blair, H. T., Etkin, D., and Tzanetos, D. B. (1995) Inuences of vestibular and visual motion information on the spatial firing patterns of hippocampal place cells. J. Neurosci. 15, 173–189.

Sutton, R. and Barto, A. G. (1998) Reinforcement Learning—An Introduction. MIT Press-Bradford Books, London.

Taube, J. S., Muller, R. I., and Ranck, Jr., J. B. (1990) Head direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J. Neurosci. 10, 420–435.

Tolman, E. C. (1948) Cognitive maps in rats and men. Psychol. Rev. 55, 189–208.

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

Yu, A. J. and Dayan, P. (2002) Acetylcholine in cortical inference. Neural Netw. 15, 719–730.

Yu, A. J. and Dayan, P. (2003) Expected and unexpected uncertainty: ACh & Ne in the neocortex. In: Advances in Neural Information Processing Systems 15. Thrun, S., Becker, S., and Obermayer, K., (eds.). MIT Press, Cambridge, MA: pp. 157–164.