

available at [www.sciencedirect.com](http://www.sciencedirect.com)[www.elsevier.com/locate/brainres](http://www.elsevier.com/locate/brainres)
**BRAIN  
RESEARCH**

## Research Report

# A reinforcement learning approach to model interactions between landmarks and geometric cues during spatial learning

Denis Sheynikhovich\*, Angelo Arleo

UPMC-Paris 6, CNRS-UMR7102, 9 quai St. Bernard, F-75005 Paris, France

### ARTICLE INFO

#### Article history:

Accepted 26 September 2010

#### Keywords:

Spatial learning

Landmarks

Geometry

Navigational strategies

Reinforcement learning

Computational model

### ABSTRACT

In contrast to predictions derived from the associative learning theory, a number of behavioral studies suggested the absence of competition between geometric cues and landmarks in some experimental paradigms. In parallel to these studies, neurobiological experiments suggested the existence of separate independent memory systems which may not always interact according to classic associative principles. In this paper we attempt to combine these two lines of research by proposing a model of spatial learning that is based on the theory of multiple memory systems. In our model, a place-based locale strategy uses activities of modeled hippocampal place cells to drive navigation to a hidden goal, while a stimulus–response taxon strategy, presumably mediated by the dorso-lateral striatum, learns landmark-approaching behavior. A strategy selection network, proposed to reside in the prefrontal cortex, implements a simple reinforcement learning rule to switch behavioral strategies. The model is used to reproduce the results of a behavioral experiment in which an interaction between a landmark and geometric cues was studied. We show that this model, built on the basis of neurobiological data, can explain the lack of competition between the landmark and geometry, potentiation of geometry learning by the landmark, and blocking. Namely, we propose that the geometry potentiation is a consequence of cooperation between memory systems during learning, while blocking is due to competition between the memory systems during action selection.

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

A long-standing question in behavioral neuroscience is whether spatial learning can be accounted for by the traditional associative theory of learning. The associative theory posits that all available cues compete for control over behavior (Pavlov, 1927). In contrast, cognitive theorists since Tolman propose that some spatial tasks require integration of

multiple cues in order to form a map-like representation of the environment, which is then used to navigate (Tolman, 1948). The issue of an interaction between geometric cues and other spatial cues during goal navigation has gained a considerable interest during the last few decades partly because of its relation to this general question (Cheng and Newcombe, 2005; Pearce et al., 2001). Geometric cues are defined by the shape of the testing arena, such that e.g. a food hidden in one corner of

\* Corresponding author. Fax: +33 144272669.

E-mail addresses: [denis.sheynikhovich@upmc.fr](mailto:denis.sheynikhovich@upmc.fr) (D. Sheynikhovich), [angelo.arleo@upmc.fr](mailto:angelo.arleo@upmc.fr) (A. Arleo).

URL: <http://www.anc.upmc.fr/dsheynikhovich.html> (D. Sheynikhovich).

a rectangular room can be found with 50% probability based only on the geometric cues (which in this case can be represented by a corner with, e.g. long wall on the right and short wall on the left). Spatial cues of other types, e.g. a beacon attached to the goal or a cue-card on the wall, can also signal the goal location. The associative theory predicts that geometric cues should compete with these other cues. The cue competition manifests itself with such phenomena as blocking (Kamin, 1969) and overshadowing (Pavlov, 1927). In blocking, if a cue alone is learned to reliably predict the reward, it will prevent (or block) learning about subsequently added cues that predict the reward as well. In overshadowing, if the two cues are present from the outset of training, the association between one of the cues and reward will be learned less well than if the training were performed with that cue alone (the cue is said to be overshadowed by the second cue during training). In the spatial domain, blocking would occur if learning that a landmark predicts well the location of a goal would prevent the animal from learning the position of the goal with respect to the shape of the testing arena (Wall et al., 2004). Quite a few behavioral studies have been performed to test whether blocking and overshadowing actually occur between geometric and non-geometric cues, but the results are controversial (for reviews see Cheng and Newcombe, 2005; Miller and Shettleworth, 2007). In particular, most studies report that geometric cues neither block nor overshadow other cues (Wall et al., 2004; Hayward et al., 2003; McGregor et al., 2009; Pearce et al., 2001). Moreover, in even a sharper contrast to what is predicted from cue competition, the presence of non-geometric cues signaling goal location seems to *potentiate* learning of geometric cues, rather than to diminish it in some studies (Pearce et al., 2001; Graham et al., 2006). On the contrary, a few studies report that blocking and overshadowing phenomena do occur in some experimental paradigms (Pearce et al., 2006; Pearce, 2009).

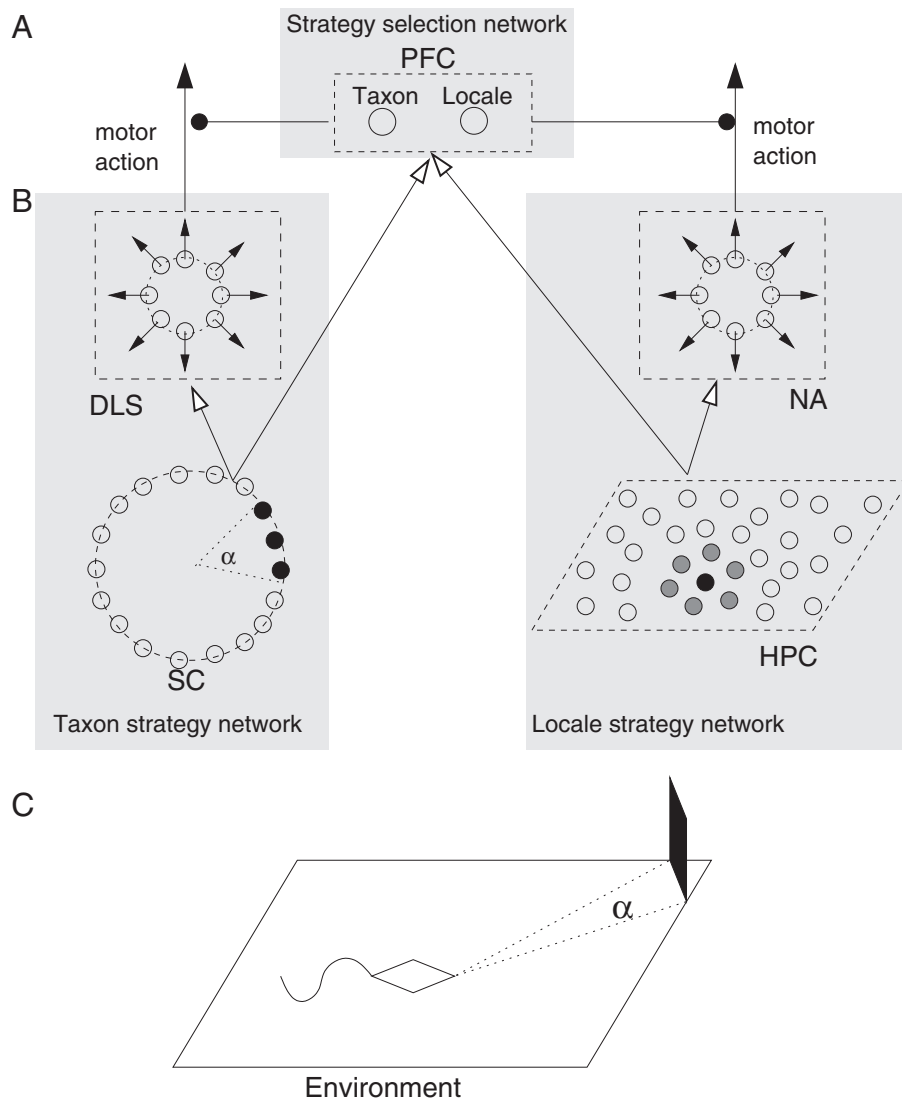
Recently, Miller and Shettleworth (2007) proposed a purely associative model that provides the first explanation of independence, enhancement, or blocking between geometric and non-geometric cues during learning. In this model, based on classical conditioning principles (Rescorla and Wagner, 1972), all sensory cues that might potentially influence goal learning are assigned corresponding associative strengths, which measure how well a reward is predicted by the cue. During learning, the associative strength of a cue is updated depending on whether the reward was obtained or not and on how frequently the sensory cue is observed. Miller and Shettleworth (2007, 2008) have shown that the model is able to reproduce a large array of behavioral data due to a property termed *feature enhancement*. Feature enhancement occurs early in training, when the associative strength of geometry is not well learned, and is responsible for the absence of blocking and overshadowing as well as potentiation of geometry learning. Late in training, the learning of the feature overshadows geometry learning.

In our previous work (Sheynikhovich et al., 2009) we proposed a model of spatial learning that attempted to explain rat behavior in navigation tasks using available knowledge about anatomical connections between neural structures implicated in navigation and neural activity within these structures. In particular, we have shown that the model was

able to reproduce the key results by Cheng (1986), who was the first to show experimentally that the relationship between geometry and feature learning might not be consistent with the general associative theory of learning (see Pearce et al., 2001; Shettleworth, 1998). This model is based on the assumption that separate independent memory systems in the rat's brain govern different navigational strategies (Fig. 1B). The first, *taxon* strategy is assumed to be mediated by the dorso-lateral striatum and learns the association between visual cues and rewarded motor responses using a simple reinforcement learning rule. This strategy is responsible for what is usually termed 'feature'- or 'landmark'- or 'beacon'-learning in the geometry-related literature (Cheng, 1986; Wall et al., 2004; Pearce et al., 2001). The second, *locale* strategy uses spatial representation built on the basis of simulated hippocampal place cells and entorhinal grid cells in order to find the goal location. We proposed that environmental geometry is implicitly encoded in the activities of place cells and so the behavioral decisions usually attributed to the influence of geometric cues might be explained by the sensory information processing during locale strategy. In this earlier model, the selection between different strategies during goal navigation was implemented using a simple scheme in which the strategy that was more successful during a fixed number of preceding trials was chosen for the next trial. Such a simple strategy selection scheme did not allow us to address questions related to the issue of competition between environmental geometry and other spatial cues, although a mere presence of independent strategies already suggested that their learning was not always subject to blocking and overshadowing phenomena (Sheynikhovich et al., 2009).

In the present paper we propose that a similar model, but augmented with a simple strategy selection rule, can be used to address questions of blocking and overshadowing between geometric cues and landmarks. In this new model, a strategy selection network (Fig. 1A), presumably located in the prefrontal cortex, governs the choice between memory systems using a reinforcement learning rule similar to that of Chavarriga et al. (2005a,b) (see also Dollé et al., 2010). We do not attempt here to apply this model to reproduce an extensive set of available experimental data on this subject. Rather, we show that the proposed model possesses properties that are similar to feature enhancement in the associative model by Miller and Shettleworth (2007), and hence may explain a similar array of data. The advantage of our model compared to other similar models (Miller and Shettleworth, 2007; Miller and Shettleworth, 2008; Dawson et al., 2010) is that geometric cues in the model are encoded implicitly by the locale strategy and so the modeler is not required to explicitly insert into the model such parameters as 'background cues', 'correct/incorrect geometry', wall lengths, etc. Moreover, the model architecture can be mapped on the biological network implicated in behavior, works in realistic time scale and generates trajectories of simulated animals during learning instead of providing rather abstract predictions of behavior in terms of choice probabilities.

In the following section we briefly introduce, following Miller and Shettleworth (2007), the property of feature enhancement, using as an example their simulation of the behavioral task described in Wall et al. (2004). In the Results



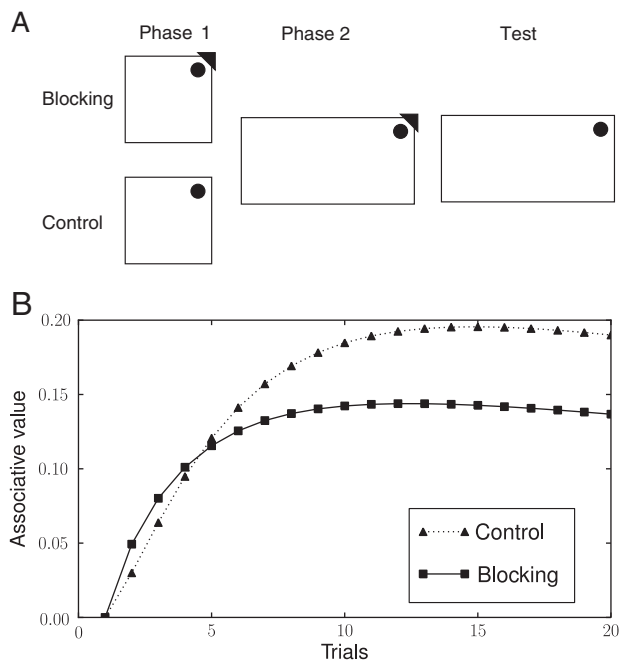
**Fig. 1 – The architecture of the navigation model.** Open (filled) circles denote inactive (active) cells; the large open arrows denote all to all feed-forward projections between corresponding neural structures; the large filled arrows denote motor output of the model; the small filled arrows in the DLS and NA denote preferred direction of the motor (action) cells in these structures; ball-arrows denote inhibition. SC — sensory cortex, DLS — dorso-lateral striatum, HPC — hippocampus, NA — nucleus accumbens, PFC — prefrontal cortex. **A.** The strategy selection network consisting of two units that correspond to the two strategies shown in **B**. **B.** Taxon and locale strategy networks that correspond to the striatal and hippocampal memory systems, respectively. **C.** An illustration of a general experimental setup. The rhombus denotes the rat, the black parallelogram denotes the landmark. The visual angle  $\alpha$  that the landmark takes at the position occupied by the rat corresponds to the number of active sensory cells in **B**. See [Experimental procedures](#) for a detailed description.

section we show that our model can be used to reproduce rat behavior in the same task and that feature enhancement occurs in our model as well.

### 1.1. Experiment of Wall et al. (2004) and feature enhancement

Feature enhancement can be understood by considering the application of the model of Miller and Shettleworth (2007) to simulate rat behavior in the experiment of Wall et al. (2004). In phase 1 of this experiment, two groups of rats were trained to find food in a corner of a square enclosure (Fig. 2A, Phase 1).

For the first (blocking) group, the corner associated with the food was marked with a landmark. For the second (control) group, no landmark was present to distinguish the rewarded corner from the other corners. The starting positions varied from trial to trial. After training to criterion, the blocking group learned that the landmark was a good predictor of the food location, while choices of the control group were random, ensuring that neither geometric nor other cues could have been used to remember the goal location for this group. In phase 2 of the experiment, both groups were retrained in a rectangular enclosure with the same landmark indicating the correct corner (Fig. 2A, Phase 2). The associative theory would



**Fig. 2 – A.** The experimental design in the study of Wall et al. (2004). The rectangles denote environmental enclosures, the black circle denotes goal location and the black triangle denotes the placement of a landmark in this experiment (see text for details of the experiment). **B.** The evolution of the associative value of the correct geometry in the model of Miller and Shettleworth (2007) during phase 2 of the simulated experiment outlined in A.

predict that if the landmark was in competition with the geometric cues, rats in the blocking group would not be able to learn that the geometry of the environment could be used to reach the goal with 50% probability. This is because the landmark was already known to be a good predictor of reward from phase 1. In contrast to this prediction, the test performed in the rectangular environment in the absence of the landmark (Fig. 2A, Test) resulted in both groups going equally often to the correct corner and its diagonally opposite corner, and almost never to the other two corners. The preference for the two geometrically correct corners shows that the training with the landmark did not block geometry learning in this case.

In order to model rats choices in this experiment, (Miller and Shettleworth, 2007) assumed that four types of cues were important during this phase: (i) the feature; (ii) the ‘correct geometry’, i.e. some aspects of geometry that are common for the correct and its diagonally opposite corner; (iii) the ‘incorrect geometry’ of the other two corners; and (iv) background cues present in all the corners (the presence of background cues ensures that all four corners are chosen at random in the absence of any specific information). Feature enhancement can be observed if we consider the associative value of the correct geometry in this model (Fig. 2B, see Miller and Shettleworth, 2007). The presence of the feature early in training results in a higher associative value of the correct geometry in the blocking group relative to the control group. Later in training, an overshadowing occurs so that the

associative value of geometry in the control group becomes smaller than that in the blocking group. In a number of simulations the authors have demonstrated that the same feature enhancement effect can be used to explain a large variety of related experimental data (Cheng, 1986; Pearce et al., 2001; Graham et al., 2006; Pearce et al., 2006).

## 2. Results

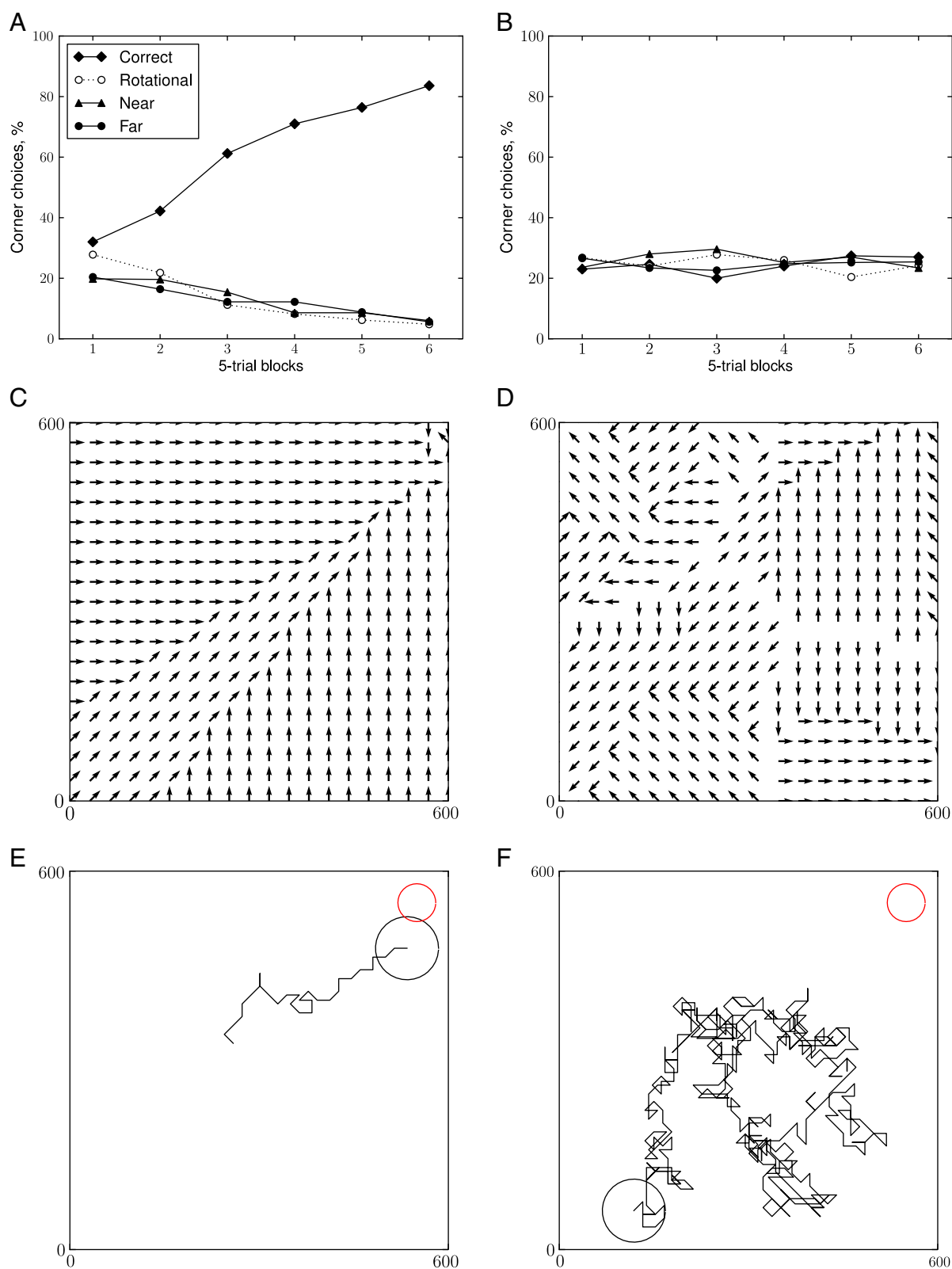
### 2.1. Simulation 1: modeling the experiment of Wall et al. (2004)

We first tested whether our model can reproduce the principal results of Wall et al. (2004) (i.e. of their Experiment 3). We used two identical groups (the *blocking group* and the *control group*) of 50 simulated rats each. In phase 1 of the simulated experiment, the blocking group was trained in a square environment with a landmark in the North–East corner, while the control group was trained in the same environment but without the landmark (see [Experimental procedures](#) and Fig. 2A). The performance was measured by counting the first visits to different corners of the training environment across blocks of training trials. A trial outcome was considered ‘correct’ when the first visited corner was the rewarded (i.e. North–East) corner and ‘rotational error’ when the first corner was the one diagonally opposite to the rewarded corner. ‘Near error’ and ‘Far error’ corresponded to trials in which the South–East and North–West corners, respectively, were visited first.

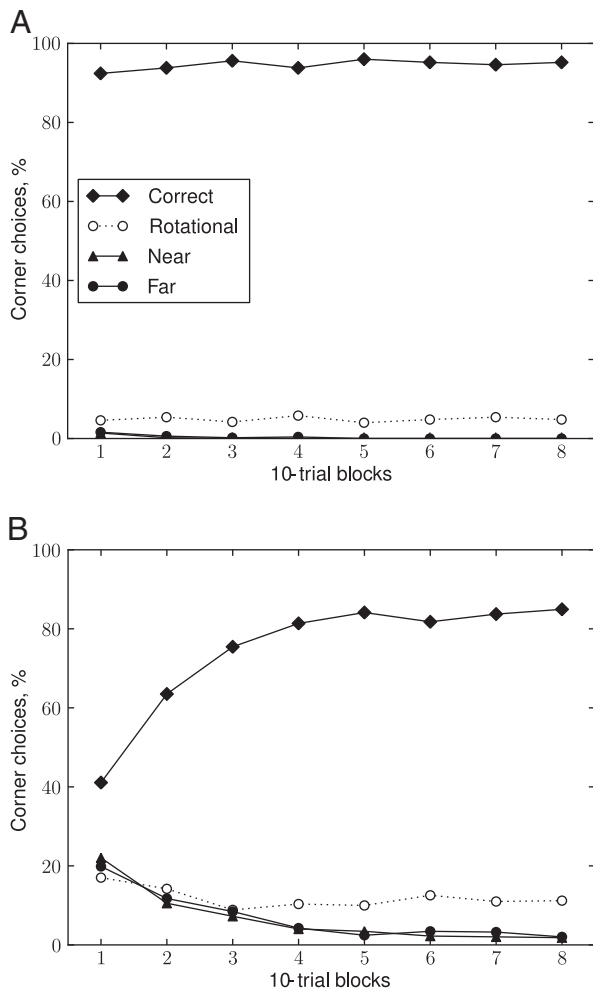
The average percentages of different outcomes for each group are shown in Figs. 3A,B. As expected, only the blocking group was able to increase the number of correct outcomes with training, while the outcomes of the control group remained at chance level for all trials. Learning in the blocking group is due to the taxon strategy which learned to turn towards the landmark. This is illustrated in Figs. 3C–F by navigation maps and sample trajectories generated in our model by the end of phase 1. The navigation maps show for each sample location of the environment the optimal action proposed by a particular strategy.

In order to test the performance of the two groups during phase 2 of the experiment, the simulated animals from phase 1 were retrained in a rectangular environment with the landmark in the North–East corner for both groups (see Fig. 2A). The mean outcomes for each group are shown in Fig. 4. Similarly to the rats in the experiment of Wall et al. (2004), simulated animals from the blocking group were choosing the correct corner most of the time starting from the beginning of training. This is because the taxon strategy transferred landmark-related information from the previous phase. In contrast, the control group was at chance level at the beginning of training, but quickly increased the rate of correct responses due to the presence of unambiguous information about the goal location.

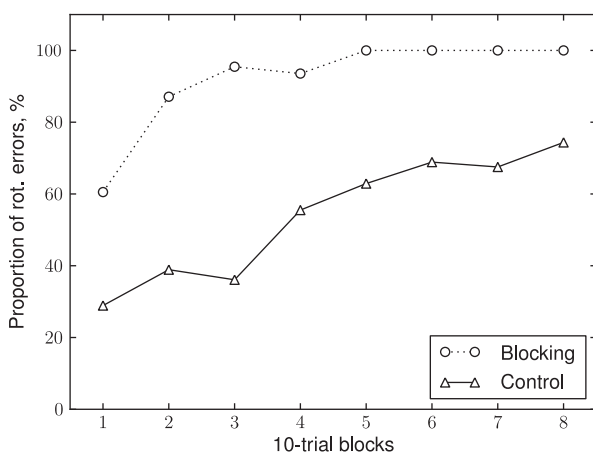
In addition to reproducing the rates of correct choices in the two experimental groups, the model also reproduces the elevated number of rotational errors relative to other errors during phase 2 (Wall et al., 2004). Indeed, in both simulated groups the proportion of rotational errors in the total number



**Fig. 3 – A-B.** Percentages of different trial outcomes across training of simulated rats in the blocking group (A) and control group (B) during phase 1 of simulation 1. C-D. Navigation maps for the taxon strategy (C) and the locale strategy (D) generated by two simulated animals during phase 1. E-F. Example of the trajectory of a simulated animal from the blocking group (E) and from the control group (F) at the end of training in phase 1 of simulation 1. The large black circle denotes the rat; small red circle denotes the goal.



**Fig. 4 – Percentages of different trial outcomes for the blocking group (A) and control group (B) during phase 2 of simulation 1.**



**Fig. 5 – The proportion of rotational errors in the total number of errors for the blocking (circles) and control (triangles) groups across blocks of training trials in phase 2 of simulation 1.**

of errors was high (Fig. 5). Moreover, this proportion was higher in the blocking group than in the control group, consistent with the experimental data. Average ratios of the number of rotational errors over the total number of errors were 90% for the blocking group, versus 50% for the control group in our simulations. These ratios in the experiments were 61% versus 47% for the blocking and control groups respectively (Wall et al., 2004). This result is important because it contradicts the prediction derived from the associative theory: since the landmark already predicts well the goal position for the blocking group in phase 2, learning of geometry should be blocked and no more rotational errors than other errors should be made by this group, which is not the case in the experiment as well as in our simulation.

The final probe trial was performed after training in phase 2. Simulated rats from both groups were run in the same rectangle as was used in phase 2, but in the absence of the landmark. In our simulation, the proportion of choices of the two geometrically correct corners (i.e. the rewarded corner and its diagonally opposite) was 94% for the blocking group and 90% for the control group, suggesting that the goal information from the training phase was transferred to the testing environment by both groups. This is comparable with the numbers reported by Wall et al. (2004) (83% and 75% for the blocking and control groups respectively). Similarly to the results of phase 2, these results contradict the prediction of the associative theory.

In modeling this experiment, Miller and Shettleworth (2007) observed feature enhancement, i.e. a speed-up of geometry learning in the presence of a landmark. A similar speed-up is observed in our model, as shown by the higher percentage of rotational errors in the blocking group relative to controls (Fig. 5), suggesting that geometry was learned faster by the rats in this group. This effect is due to the influence of training in phase 1, and the mechanism of this effect in our model is illustrated by the next simulation.

## 2.2. Simulation 2: cooperation between strategies

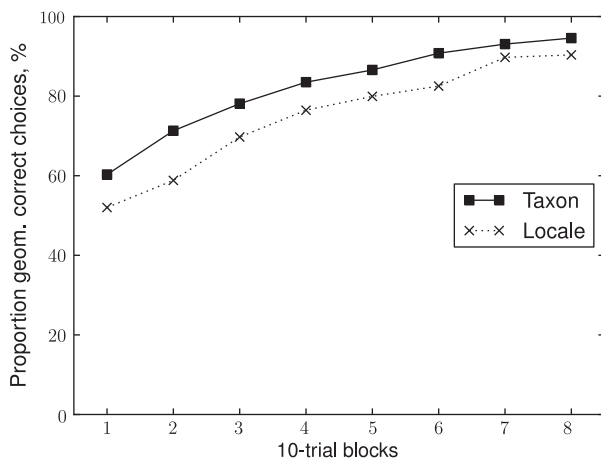
In addition to competition between the strategies during selection, a cooperation between strategies during learning takes place in our model. Namely, when the simulated animal performs a movement generated by, e.g., the taxon strategy network, this action is considered to be an 'exploratory' action by the locale strategy network (see Experimental procedures). If the behavior generated by the taxon strategy is nearly optimal, then these 'exploratory' actions result in faster learning within the locale strategy, because exploration is done always in the correct direction. In phase 2 of simulation 1 the simulated rats from the blocking group re-used information about the goal from phase 1 and had a high success rate from the beginning of phase 2 (see Fig. 4A). In these simulated animals, the faster learning of the locale strategy (expressed by the high number of rotational errors, Fig. 5) is due to the cooperative interactions between the two strategies. In contrast, in the control group both taxon and locale strategies learned simultaneously during phase 2 and hence no cooperation effect was observed.

To demonstrate the strategy cooperation effect even more explicitly, we performed the following computer simulation.

In phase 1 we trained two groups of simulated rats (referred to as *taxon group* and *locale group*) identically to phase 1 in simulation 1. In phase 2, we switched off strategy competition for both groups such that the taxon group always selected the taxon strategy, whereas the locale group always selected the locale strategy. The experimental environment was identical to that used in phase 2 of simulation 1. We tested how much was learned about geometry in the two groups by performing a test trial (identical to the probe trial in the previous simulation) after *each* training trial during phase 2. Since no landmark was present during the test trials, the amount of knowledge about geometry was reflected in the proportion of geometrically correct choices (i.e. how often the correct and its diagonally opposite corners were selected). Fig. 6 shows the proportion of geometrically correct choices for the taxon and locale groups as a function of the number of testing trials. It can be observed that even though the locale strategy was *never* chosen by the taxon group during training, the geometrically correct corners were chosen by that group more often than by the locale group during the test trials. This result clearly demonstrates the effect of strategy cooperation in our model: the presence of the predictive landmark during phase 2 sped up learning of geometry. Thus, cooperation between navigational strategies in our model results in the effect similar to feature enhancement in the model of Miller and Shettleworth (2007).

### 2.3. Simulation 3: blocking within the taxon network

The final simulation demonstrates that learning in the taxon strategy network in our model occurs in agreement with the classical associative learning theory and is subject to blocking. It is important, since in our model the taxon strategy network represents stimulus–response learning in the dorsal striatum. Experimental results from a number of studies (Doeller and Burgess, 2008; Balleine et al., 2007; Williams and Eskandar, 2006) support the notion that learning in the dorsal striatum



**Fig. 6 – The proportion of choices of geometrically correct corners (i.e. correct and its diagonally opposite) for taxon (squares) and locale (crosses) groups across blocks of testing trials in phase 2 of simulation 2. The difference is statistically significant (one-way ANOVA,  $p < 0.05$ ) in all blocks except block 7.**

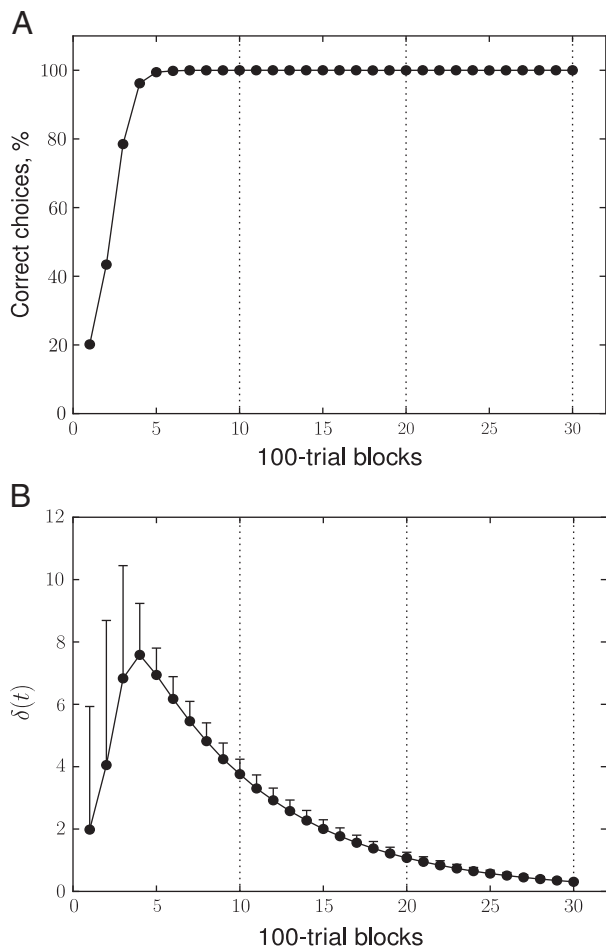
obeys associative reinforcement with a single error-prediction signal (Rescorla and Wagner, 1972).

In this simulation, we used the square environment as before with either one or two landmarks present at the goal corner of the environment. We used a standard blocking protocol (see e.g. Waelti et al., 2001) in which a group of simulated animals (the *blocking group*) was trained to turn towards a landmark from the center of the environment during the first (pre-training) phase. During the second (compound) phase, the pre-trained landmark was presented together with a second, target, landmark at the same corner. Since the pre-trained landmark is a good predictor of reward at the outset of the compound phase, it should block learning to turn towards the target landmark. The second, *control group*, was presented with both landmarks from the beginning of training, in which case no blocking should occur. The test was performed by presenting the target landmark alone to both groups. According to the associative learning theory, the performance of the blocking group during the test phase should be worse than that of the control group as a consequence of blocking. In this task, the position of the simulated rat was always fixed at the center of the environment and the reward was given when the simulated rat turned towards the hidden goal position (at the North-East corner, see *Experimental procedures*).

According to the Rescorla–Wagner model of classical conditioning, blocking will occur if the associative strength of the reward-predicting stimulus will be close to its asymptotic value as a result of pre-training (Rescorla and Wagner, 1972). The difference between the two corresponds in our model to the difference  $\delta(t)$  (Eq. (4)) between the expected and actual reward linked to the reward-predicting cue (landmark in our case). Therefore, in order to ensure that blocking will happen, we trained simulated animals in the pre-training phase until  $\delta(t)$  reached a small value. The learning curve of the blocking group with the corresponding evolution of  $\delta(t)$  are shown in Fig. 7. After training block 30 (i.e. when the value of  $\delta(t)$  was close to 0), the pre-training phase ended and the next, compound phase, started.

During the compound phase, a second landmark was added to the environment and training continued for another 5 blocks of trials. In animal studies, a test trial is usually performed by presenting the added landmark alone after a number of compound training trials. We took the advantage of a computer experiment by performing the test trial after each compound training trial (i.e. the training and test trials were interlaced). This way we could measure the rate of learning of the blocked landmark. In Fig. 8 we plot the performance of the simulated rats from the blocking group during blocks of test trials (dashed line, Blocking 1), that is compared to the performance of the control group (solid line, Control) which was trained with both landmarks from the beginning. The strong blocking effect is expressed in the absence of learning of the second landmark in the blocking group compared to the fast learning in the control group.

To show that the blocking effect can be diminished by insufficient pre-training, we performed two additional experiments which were identical to the one just described, but in which pre-training was stopped after trial block 20 and trial block 10 (marked by dotted lines in Fig. 7). As shown in Fig. 8, a

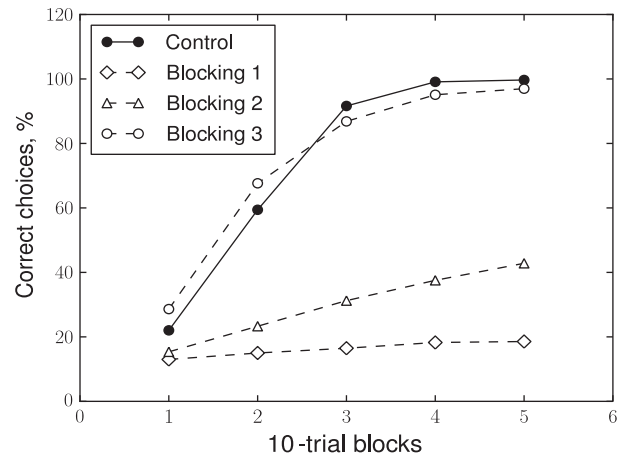


**Fig. 7 – Learning in the blocking group during the pre-training phase of simulation 3. A. The percentage of correct choices (i.e. when the simulated rat turned towards the landmark) across blocks of training trials. B. Reward-prediction error  $\delta(t)$  ( $\pm$ STD) as a function of training block number. Dotted lines mark training blocks after which pre-training stopped for different simulations (see text).**

smaller number of pre-training trials results in the weakening (dashed line, Blocking 2) or disappearance (dashed line, Blocking 3) of the blocking effect.

### 3. Discussion

In this paper we propose a new approach to the issue of how geometric cues interact with landmarks during spatial learning. This approach uses available knowledge about anatomy and neurophysiology of brain networks implicated in navigation in order to explain behavioral data traditionally analyzed on the basis of associative theory of learning (Jeffery, 2010; Pearce, 2009). The architecture of the model presented in this paper is based on the theory of separate memory systems in the brain of the rat (White and McDonald, 2002) which is in turn built upon the body of neurophysiological and lesion studies. In this theory, the hippocampus (HPC) is responsible for stimulus-stimulus associations and multi-sensory inte-



**Fig. 8 – Blocking during the compound phase of simulation 3. Percentages of correct choices for the control group (solid line) and blocking group (dashed lines) across blocks of training trials. For the blocking group, pre-training phase was stopped at trial 30 (Blocking 1), trial 20 (Blocking 2) or trial 10 (Blocking 3). Blocking effect disappears if the pre-training phase is not sufficiently long.**

gration, while the dorso-lateral striatum (DLS) is responsible for the formation of stimulus-response associations and reactive behavior. The behavioral role of the locale and taxon networks in the present model corresponds to the role of the HPC and DLS for rat behavior. In addition, the model includes a strategy selection network that is responsible for switching behavioral strategies. A number of studies suggest that the prefrontal cortex (PFC) can support strategy switching during goal-oriented behavior (Rich and Shapiro, 2007, 2009; Ragazzino et al., 1999).

The advantage of this approach is that the behavior of the model in a particular task can be related with the function of and neural activity within the existing neural structures. In particular, in our previous work we have suggested that rat behavior in the experiment of Cheng (1986), who was the first to test non-competitive interaction between landmarks and geometry in a behavioral paradigm, can be explained by an interaction between behavioral strategies mediated by different memory systems. On the basis of our results we proposed that symmetric errors often committed by rats and other animals in environments with ambiguous shape may result specifically from the use of the HPC-based memory system, mediating locale strategy. This suggestion permitted us to relate results of geometry-related experiments to neurophysiological data about activity of place cells and grid cells in environments with variable shape (Sheynikhovich et al., 2009). In the model presented here we used a simplified implementation of locale strategy in which place cells were modeled by Gaussian functions. Nevertheless, this simple model includes the ability to generate symmetric errors in a way similar to that proposed in Sheynikhovich et al. (2009). These errors result from mis-estimation of an allocentric head direction during self-localization (see Experimental procedures).

One of the purposes of the present paper was to compare our model with the associative model of Miller and Shettleworth (2007). Several conclusions can be made in relation to this issue.



First, learning to approach the landmark did not block subsequent learning of geometric cues in our simulations. Moreover, an elevated proportion of rotational errors in the blocking group during phase 2 of simulation 1 (Fig. 5) suggests that presence of the landmark increased the speed with which geometry was learned. These results are consistent with experimental data showing absence of blocking and overshadowing between landmarks and environmental geometry as well as potentiation of geometry learning by landmarks (Wall et al., 2004; Hayward et al., 2003, 2004; Cheng and Newcombe, 2005; Pearce et al., 2001). In our model, the absence of overshadowing and blocking is due to parallel learning in the HPC- and DLS-mediated memory systems, and corresponding navigation strategies. Indeed, learning in both systems occurs when movements performed by one of the systems lead to reward. The only condition for learning is that there should exist some stable relationship between sensory inputs, movements and reward (obviously, if no landmark is present in the environment, the taxon strategy will not be able to learn). The effect of potentiation of geometry learning by landmarks in our model is due to strategy cooperation, as demonstrated by simulation 2.

Miller and Shettleworth (2007, 2008) provided a similar but slightly different explanation for the absence of blocking and overshadowing. In their model, associative strengths of sensory cues that co-occur with reward will increase with training (provided zero initial conditions). This increase occurs in parallel for all cues, including geometric and non-geometric cues. This parallel learning property is somewhat similar to the parallel learning of strategies in our model. However, in our model the different types of cues are separated into two groups according to the memory systems that work with these types of cues. By construction of the model, all geometric cues are encoded implicitly by the place cell activities and so will be dealt with by the HPC-mediated memory system and locale strategy. Landmarks, beacons and other cues that can be used to drive stimulus–response associations will be processed by the DLS-mediated system and (possibly multiple) taxon strategies.

Second, since the HPC- and DLS-mediated systems learn in parallel, one of these systems cannot prevent the other from learning. However, effects similar to blocking can be induced on the level of strategy selection. Consider the probe trial in our simulation 1. At the time of the probe trial, the value of the taxon strategy in the blocking group is higher than that in the control group due to longer training with the landmark. Recall that in our model, when no landmark is present, taxon strategy proposes movements in random directions and so simulated rats in the blocking group should make more random movements than those in the control group during the test. However this difference can hardly be seen, because the stochasticity in strategy choice (Eq. (2)) leads to selection of the locale strategy on some time steps inducing a bias towards geometrically correct corners. Suppose that we modify the selection rule such that after some point in training (e.g. when reward prediction error  $\delta$  in Eq. (3) will become smaller than some threshold within certain time window) strategy selection will not be stochastic any more so that the best (i.e. taxon) strategy is exclusively selected in the blocking group. Suppose that by the time when this happens, the control group has not

yet reached the threshold. In this case, the behavior of simulated animals from the blocking group during the probe trial will be random, corresponding to blocking, while the controls will be biased towards the geometrically correct corners, as before. Note that increasing parameter  $\beta$  in Eq. (2) will essentially lead to exclusive selection of the winner strategy and so blocking can be introduced in the model by making  $\beta$  depending on the inverse of  $\delta$ . On the basis of present results we cannot say whether such a model of blocking is capable of reproducing related experimental data (Pearce, 2009).

The last issue that we would like to mention in relation to the comparison between our model and that of Miller and Shettleworth concerns different predictions derived from these models for a landmark transfer test described in Miller and Shettleworth (2008). In our model, taxon strategy learns egocentric rotations relative to the representation of a landmark in the view field. The behavior generated by this strategy can be approaching the landmark as well as going in the direction opposite to the landmark, and this behavior can be transferred between different environments. Consider the experiment in which animals are trained to approach the corner opposite to a corner with the landmark in a rectangular environment. Our model predicts that if the animals were tested in a square environment with the landmark in one of the corners, they would go more often to the corner opposite to the landmark. In contrast, the model of Miller and Shettleworth (2008) predicts equal choice probability for the three corners without the landmark.

An important aspect of the model is that learning in the three different subnetworks is governed by the same TD-learning algorithm. However, due to differences in the inputs to these networks, their properties differ. Namely: (i) Learning in the taxon strategy network is consistent with the classical associative theory as shown by the results of simulation 3; (ii) In the locale strategy network, inputs are given by the Gaussian-like activities of simulated place cells. As in many models of place cells, we assume here that their place-sensitive activity results from integration of multi-sensory information at a particular location, and hence is not based on cue competition. Integration of multi-sensory input can be implemented by unsupervised Hebbian-like learning rules during environment exploration (see Redish, 1999, for review). The reward-based learning is required to generate goal-directed behavior on the basis of the learned place code. (iii) Behavior of the full model results from competitive and cooperative interactions between the two strategies and in general cannot be described by the associative theory (simulations 1 and 2). These properties are consistent with the view that striatal learning is controlled by single error signal (reward-prediction error), while learning in the hippocampus is based (at least partially) on reward-independent associative learning (Doeller and Burgess, 2008; Doeller et al., 2008).

One observation about strategy cooperation in our model may be interesting in light of the results by Doeller and Burgess (2008) concerning an asymmetry in blocking between environmental boundaries and landmarks in human virtual reality study. They found that learning a hidden goal location with respect to boundaries blocked subsequent landmark learning, but no blocking was observed when the goal was first

paired with the landmark. From our simulation 2 it is clear that learning with the landmark in phase 1 does not block but helps geometry learning in phase 2, which is consistent with these results. What happens if the learning with respect to geometry occurs in phase 1 and the landmark is introduced in phase 2? In the study of Doeller and Burgess (2008), the landmark was not always in the field of view during phase 2. Our model suggests that in this case movements according to the locale strategy cannot be mapped to egocentric turns required by the taxon network to learn and so the taxon network will not learn from movements generated by the locale strategy during these periods of time. It is an open question whether this observation can be used to explain some of the results reported by Doeller and Burgess (2008).

Finally, we note here that our model cannot provide clear explanation for the experimental data suggesting that locale strategies dominate in the beginning of learning, while locale strategies take over after prolonged training (Packard and McGaugh, 1996, Sheynikhovich et al., 2010). In our model, the strategy, that better predicts reward, will dominate. This leads to a speculation that, at least in the plus maze, the place code is more accurate than sensory cues in the beginning of training, with sensory cues gaining more accuracy with prolonged training. Testing this suggestion is the matter of a future study.

#### 4. Experimental procedures

Our neural network model of navigation consists of three interconnected (sub-)networks that represent neural structures participating in navigation behavior (Figs. 1A,B). The first, taxon network, is responsible for stimulus-response behavior, such as approaching a visible landmark. It models the anatomical connections between sensory cortices (SC), that encode a visual representation of the landmark, and motor-related activity of the dorso-lateral striatum (DLS) of the basal ganglia. The second, locale network, is responsible for map-based navigation, usually attributed to the hippocampus (HPC). The input of this network is represented by the activities of simulated place cells, and the motor actions are assumed to be generated by the nucleus accumbens (NA) of the basal ganglia, which receives direct projection from the HPC (see Redish, 1999, for review). The last, strategy selection network, is responsible for the choice of a strategy (i.e., taxon or locale) that will take control over behavior. This function is usually attributed to the prefrontal cortex (PFC), which is required for strategy switching and where different subpopulations of neurons were shown to code for different navigation strategies (Rich and Shapiro, 2007, 2009). The input to the strategy selection network is represented by the combined activities of place cells and landmark-encoding cells and corresponds to the connections from the HPC and DLS to the PFC.

Each of the three networks has the same simple architecture of a single-layer perceptron. The input layer encodes sensory input to the network, while the output layer represents actions that this network generates. Learning of the weights in each network is described by a standard temporal-difference learning rule known as Q-learning,

while action selection is performed according to ‘softmax’ criterion (see Learning equations below). The three networks are hierarchically organized in two levels: the two networks of the lower level (taxon and locale) learn the corresponding strategies, while a single network of the upper level (strategy selection network) learns to select which of the two networks of the lower level will take control over behavior. The selection is assumed to occur by inhibiting the motor output of the non-winner strategy. Below we describe the inputs and outputs of the three networks, learning equations and experimental setup.

##### 4.1. Taxon strategy

Inputs of the taxon strategy network encode a visual representation of  $N_{lm}$  landmarks ( $N_{lm}=1$  in Simulations 1 and 2, and  $N_{lm}=2$  in Simulation 3) modeled by  $N_{lm}$  populations consisting of  $N_{sc}$  sensory cells. In each population, the activity of  $j$ -th sensory cell signals either presence ( $r_j^{sc}=0$ ) or absence ( $r_j^{sc}=0$ ) of the corresponding landmark in the egocentric direction  $\phi_j=2\pi j/N_{sc}$  (in the egocentric reference frame, the zero direction coincides with the head direction of the simulated animal). The number of active units in each population corresponds to the visual angle that the corresponding landmark takes at the position occupied by the simulated rat (see Fig. 1C).

The sensory units project to  $N_{tax}$  taxon-action cells with weights  $w_{ij}^{tax}$ . Each taxon-action cell encodes an egocentric rotation by angle  $\xi_i^{tax}=2\pi i/N_{tax}$ . The activities of the taxon-action cells are calculated according to Eq. (1). The actual rotation to be performed on the next step is chosen based on these activities according to Eq. (2). During training, weights  $w_{ij}^{tax}$  are updated using Eq. (3) at each time step.

Note that the absence of landmarks in the environment corresponds to the zero activation of the sensory cells. This leads to a pseudo-random action choice in the taxon network (see Eq. (2)) which corresponds to movements in pseudo-random directions. Such a behavioral response is in agreement with studies showing that animal performance drops to chance when the reward-related stimulus is removed (see, e.g., Roberts and Pearce, 1999).

##### 4.2. Locale strategy

Inputs of the locale strategy network represent the position of the simulated rat in the environment and are modeled by the activities of  $N_{pc}$  simulated place cells with Gaussian spatial receptive fields, centers of which are distributed uniformly over the environment. More precisely, the activity of place cell  $j$  is calculated as  $r_j^{pc}=\exp(-d_j^2/2\sigma^2)$ , where  $\sigma$  is the width of receptive field and  $d_j$  is the distance between the position of the simulated animal in the maze and the center of the  $j$ -th receptive field.

Place cells project to  $N_{loc}$  place-action cells with weights  $w_{ij}^{loc}$ . Each place-action cell encodes the allocentric direction of movement  $\xi_i^{loc}=2\pi i/N_{loc}$ . The activities of the place-action cells are calculated according to Eq. (1). The direction of movement to be performed is chosen based on these activities according to Eq. (2). During learning, weights  $w_{ij}^{loc}$  are updated using Eq. (3) at each time step. In

all the simulations in this paper we assumed that place cells were created before the training trials started.

In our earlier work (Sheynikhovich et al., 2009) we proposed that geometric cues present in the environment can be implicitly encoded in the activities of place cells. If the shape of the environment is non-symmetric then population activity of place cells can be used to decode the true position of the animal. If, in contrast, the environmental shape is ambiguous, then the location signaled by the place cells is subject to symmetrical errors. In this model, symmetric errors in the place cell activities result from the error in the estimation of allocentric head direction. The results from this earlier work can be summarized in the following simple assumptions used to generate place cell activity in the current model: (i) The simulated animal performs reorientation at the start of each trial. This process results in an estimation of the allocentric orientation  $\Phi$  up to the symmetric error (in the present model we discard noise in direction estimation coming from other sources). For example, if the true initial orientation in a (twofold symmetric) rectangular environment is  $\pi/2$ , then the estimated orientation  $\Phi$  can be either  $\pi/2$  or  $3\pi/2$ , with equal probability (irrespective of whether landmarks are present or not). (ii) The activities of place cells after reorientation correspond to the estimated orientation  $\Phi$ . In our example, assume that the true position of the simulated animal in the rectangle was the center of the southern wall looking towards the center. If reorientation results in a symmetric error (i.e.  $\Phi=3\pi/2$ ) then the place cell activities would not correspond to the true position, but to its rotationally opposite (i.e. the center of the northern wall). In the present simulations this is done algorithmically by resetting the centers of the receptive fields in accord with the estimation  $\Phi$ . (iii) Once the reorientation is performed, estimation  $\Phi$  is taken as if it were the ‘true’ allocentric direction and subsequent movement directions  $\xi_i^{\text{loc}}$  dictated by the locale strategy are calculated relative to this estimated head direction. In our example, if the action generated by the locale strategy corresponds to the allocentric movement in the direction of  $\pi/4$ , the simulated animal will actually move in the direction of  $5\pi/4$  due to the symmetric error committed during reorientation. After each time step during a trial, the estimation of the allocentric orientation is updated according to

$$\Phi(t) = \Phi(t-1) + \Delta\Phi,$$

where  $\Delta\Phi$  is the rotation angle since the previous time step. Estimation  $\Phi$  is a simple model of the internal direction estimation thought to be performed by the head direction cell network in the rat (Arleo and Gerstner, 2001; Wiener and Taube, 2005).

Under the assumptions described above, the simulated animal will be subject to symmetric errors during goal navigation in environments with ambiguous shape (Cheng, 1986). Importantly, these errors cannot be decreased by longer reward-based training, since they are committed during reorientation at the level of place cells, i.e. upstream of the place-action learning network that is updated during training (Sheynikhovich et al., 2009).

#### 4.3. Selection between strategies

Inputs to the strategy selection network are represented by the combined activities of  $N_{\text{pc}} + N_{\text{sc}}$  input cells from taxon and locale strategy networks. An intuition for this particular choice of input is the following: given all available information at the current location, the selection network decides which strategy should be chosen to take control over behavior.

The input cells project to two strategy-selective (action) cells with weights  $w_{ij}^{\text{sel}}$ . The two strategy-selective cells code for taxon and locale strategies, respectively. Activity of a strategy-selective cell is calculated according to Eq. (1). A winner strategy is selected on each time step according to the action selection scheme Eq. (2). During training, weights  $w_{ij}^{\text{sel}}$  are updated using Eq. (3) at each time step.

An important property of the model is that motor actions, performed by the winner strategy, are considered ‘exploratory’ by the other strategy. In this way, both strategies learn after a motor action is performed. We assume here that information about the motor action to be performed (that may be present in the brain in the form of, e.g., an efference copy, see Andersen et al. (1997)) is shared between the taxon and locale subnetworks. However, since actions in these subnetworks are defined in different spatial reference frames (i.e. an egocentric frame for the taxon strategy and an allocentric one for the locale strategy), a conversion between these reference frames must be performed before learning (Andersen et al., 1997). This conversion depends on the current estimation of the allocentric orientation  $\Phi$  (see the previous section) and is implemented as follows:

- Suppose that at time  $t$  the taxon strategy is the winner and an egocentric turn by  $\xi_i^{\text{tax}}$  is to be performed. In the local strategy subnetwork, this action is converted to  $\xi_i^{\text{loc}} = \Phi + \xi_i^{\text{tax}}$  and considered as ‘exploratory’, similarly to internally generated (i.e., according to Eq. (2)) exploratory actions.
- If the locale strategy is the winner and a movement in the allocentric direction  $\xi_i^{\text{loc}}$  is to be performed, then this action is converted to  $\xi_i^{\text{tax}} = \xi_i^{\text{loc}} - \Phi$  in the taxon strategy network and is considered as ‘exploratory’ by that network.

#### 4.4. Learning equations

Basic learning equations in our model implement standard Q-learning algorithm (Sutton and Barto, 1998) where states and actions are encoded in the firing rates of activities of input and output neurons, respectively (Arleo and Gerstner, 2000; Sheynikhovich et al., 2009). More precisely, a state  $s_t$  at time  $t$  is represented by the activities of input units  $r_j^{\text{state}(t)}$ . These project via connections with weights  $w_{ij}$  to action units with activities  $r_i^{\text{action}}$ :

$$r_i^{\text{action}}(t) = \sum_j w_{ij} r_j^{\text{state}(t)}. \quad (1)$$

Each action unit  $i$  represents an action  $a_i$  available in the state  $s_t$ . Activity of an action cell (Eq. (1)) is interpreted as the Q-value  $Q(s_t, a_i)$  of performing action  $a_i$  at state  $s_t$ . The action  $a_t$ , that will actually be performed at time  $t$ , is chosen

stochastically from all available actions on the basis of the Q-values according to the softmax criterion (Sutton and Barto, 1998). More precisely, the probability that action  $a_i$  will be chosen at time  $t$  is given by:

$$p(a_t = a_i) = \frac{\exp(\beta Q(s_t, a_i))}{\sum_j \exp(\beta Q(s_t, a_j))} \quad (2)$$

with  $\beta$  as a constant parameter (see Table 1 for parameter values). The stochastic action choice (Eq. (2)) ensures that different actions are explored during learning.

During learning, the weights  $w_{ij}$  are adjusted on each time step according to the standard temporal-difference learning rule:

$$\Delta w_{ij}(t) = \eta \delta(t) e_{ij}(t) \quad (3)$$

where  $\eta$  is the learning rate,

$$\delta(t) = R_t + \gamma Q(s_t, a_t^*) - Q(s_{t-1}, a_{t-1}) \quad (4)$$

is the reward prediction error and  $e_{ij}(t)$  is the eligibility trace which memorized state-action pairs activated in the past. The eligibility trace of a synapse (Sutton and Barto, 1998) is increased each time the synapse participates in generating a movement, i.e.  $e_{ij}(t+1) = e_{ij}(t) + r_j^{\text{state}(t)}$  if action  $a_i$  is performed at time  $t$ . All eligibility traces decay with time according to  $e_{ij}(t+1) = \gamma \lambda e_{ij}(t)$ , where  $0 < \gamma, \lambda < 1$ . In Eq. (4),  $Q(s_t, a_t^*)$  is the Q-value of the optimal action at time  $t$ , i.e.  $a_t^* = \text{argmax}_i Q(s_t, a_i)$ , while  $Q(s_{t-1}, a_{t-1})$  is the Q-value of the action actually chosen at time  $t-1$  (due to the stochastic action choice, Eq. (2), the optimal action is not always chosen).

#### 4.5. Experimental setup and computer simulations

Parameters of the simulation environments and general experimental procedure were similar to those reported in Wall et al. (2004). In particular, two simulated environments were used throughout the simulations, a square environment  $60 \times 60$  cm and a rectangular environment  $120 \times 60$  cm. In some simulations a landmark was added to the environment. The landmark was always put in the North–East corner of the environment (coordinates of the center of the landmark were (55, 55) cm in the square environment and (115, 55) cm in the rectangular environment relative to the [0, 0] origin at the

South–West corner). Orientation of the landmark was always  $135^\circ$  relative to zero direction defined to lie along the southern wall of the environment and its length spanned the right-angled corner (see Fig. 1C). A circular goal zones of radius  $r_{\text{goal}} = 3$  cm were located in the four corners of the environments (at the distance of 7 cm from the corresponding corner).

In simulations 1 and 2, the simulated rat (represented by a circle of radius  $r_{\text{rat}} = 5$  cm) was put at one of eight starting positions in pseudo-random order at the start of each training trial. Similarly to starting positions used in the experiment of Wall et al. (2004), the eight positions were the middles of the four walls (with the orientation towards the center of the environment), or the center of the environment (with the orientation towards the middles of the four walls). In simulation 3, the simulated animal stayed at the center of the square environment for the whole experiment, with initial orientations at the start of a trial chosen randomly from  $0, \pi/4, \dots, 7\pi/4$ .

During a learning trial, the movements of the simulated animals were controlled by different strategies as described above. Thus, on each time step (i) a strategy which will guide behavior was chosen (see Selection Between Strategies); (ii) if the taxon strategy was chosen, the simulated animal turned by the angle  $\xi_i^{\text{tax}}$ , corresponding to the action  $a_t^{\text{tax}}$  (see Taxon Strategy); if the locale strategy was chosen, the simulated animal turned in the allocentric direction  $\xi_i^{\text{loc}}$  that corresponded to action  $a_t^{\text{loc}}$  (see Locale strategy). In simulations 1 and 2, once the movement direction was chosen, the simulated animal performed a forward movement by 2 cm which corresponded to a rat moving with constant speed of 16 cm per second with time steps of 0.125 s. Steps (i) and (ii) were performed until one of the four corners was reached, or after 1000 time steps (corresponding to  $\approx 2$  min maximal trial time). Reaching a corner was detected when the distance between the rat position and the center of the goal zone was less than  $r_{\text{goal}} + r_{\text{rat}}$ . Reward  $R = 10$  was given upon the reaching the North–East corner, while  $R = 0$  for all other outcomes. In simulation 3, the positive reward was given when the simulated rat turned towards the landmark, while zero reward was given otherwise (i.e. a trial consisted of one time step). Average over 50 simulated animals is shown in all plots.

Table 1

Size of input populations $N_{\text{pc}}, N_{\text{sc}}$	400
Size of output populations $N_{\text{tax}}, N_{\text{loc}}$	8
Q-learning parameters $\gamma, \lambda$	0.9
Reward for reaching the goal $R$	10.0
Learning rate $\eta$	0.01
Softmax parameter $\beta$ (selection network/strategy networks)	1.0/5.0
Place field width $\sigma$ , (mm)	50

Model parameters. Exact population sizes  $N_{\text{pc}}, N_{\text{sc}}, N_{\text{tax}}$  and  $N_{\text{loc}}$  are not essential and can be increased without qualitatively changing the results. Learning parameters  $\gamma, \lambda, \eta$  and  $\beta$  were chosen by hand to maximize learning speed in Simulation 1. Place field width  $\sigma$  was chosen to approximate real place field width (with the size of the total field  $6\sigma \approx 30$  cm, (O'Keefe and Burgess, 1996).

#### Acknowledgments

This work is granted by French National Research Agency, projects EVO-NEURO ANR-09-EMER-005. We would also like to thank Agnès Guillot, Laurent Dollé and Benoît Girard for useful discussions.

#### REFERENCES

Andersen, R.A., Snyder, L.H., Bradley, D.C., Xing, J., 1997. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* 20, 303–330  
URL <http://www.ncbi.nlm.nih.gov/pubmed/9056716>.

- Arleo, A., Gerstner, W., 2000. Spatial cognition and neuro-mimetic navigation: a model of hippocampal place cell activity. *Biol. Cybern.* 83, 287–299.
- Arleo, A., Gerstner, W., 2001. Spatial orientation in navigating agents: modeling head-direction cells. *Neurocomputing* 38–40, 1059–1065.
- Balleine, B.W., Delgado, M.R., Hikosaka, O., 2007. The role of the dorsal striatum in reward and decision-making. *J. Neurosci.* 27 (31), 8161–8165 URL <http://www.jneurosci.org/cgi/content/abstract/27/31/8161>.
- Chavarriga, R., Strösslin, T., Sheynikhovich, D., Gerstner, W., 2005a. Competition between cue response and place response: a model of rat navigation behaviour. *Connect. Sci.* 17 (1–2), 167–183.
- Chavarriga, R., Strösslin, T., Sheynikhovich, D., Gerstner, W., 2005b. A computational model of parallel navigation systems in rodents. *Neuroinformatics* 3 (3), 223–242.
- Cheng, K., 1986. A purely geometric module in the rat's spatial representation. *Cognition* 23 (2), 149–178.
- Cheng, K., Newcombe, N.S., 2005. Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychon. Bull. Rev.* 12 (1), 1–23.
- Dawson, M.R.W., Kelly, D.M., Spetch, M.L., Dupuis, B., 2010. Using perceptrons to explore the reorientation task. *Cognition* 114 (2), 207–226 PMID: 19796761. URL <http://www.ncbi.nlm.nih.gov/pubmed/19796761>.
- Doeller, C.F., 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, C.F., King, J.A., Burgess, N., 2008. Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proc. Natl Acad. Sci. USA* 105 (15), 5915–5920.
- Dollé, L., Sheynikhovich, D., Girard, B., Chavarriga, R., Guillot, A., 2010. Path planning versus cue responding: a bioinspired model of switching between navigation strategies. *Biol. Cybern.* 103 (4), 299–317.
- Graham, M., Good, M.A., McGregor, A., Pearce, J.M., 2006. Spatial learning based on the shape of the environment is influenced by properties of the objects forming the shape. *J. Exp. Psychol. Anim. Behav. Process.* 32 (1), 44–59 URL <http://www.ncbi.nlm.nih.gov/pubmed/16435964>.
- Hayward, A., McGregor, A., Good, M.A., Pearce, J.M., 2003. Absence of overshadowing and blocking between landmarks and the geometric cues provided by the shape of a test arena. *Q. J. Exp. Psychol. B* 56, 114–126.
- Hayward, A., Good, M.A., Pearce, J.M., 2004. Failure of a landmark to restrict spatial learning based on the shape of the environment. *Q. J. Exp. Psychol. B* 57 (4), 289–314.
- Jeffery, K.J., 2010. Theoretical accounts of spatial learning: a neurobiological view (Commentary on Pearce, 2009). *Q. J. Exp. Psychol.* URL <http://www.informaworld.com/10.1080/17470210903540771>.
- Kamin, L.J., 1969. Predictability, surprise, attention, and conditioning. In: Campbell, B.A., Church, R.M. (Eds.), *Punishment and Aversive Behavior*. Appleton Century Crofts, New York, pp. 279–296.
- McGregor, A., Horne, M.R., Esber, G.R., Pearce, J.M., 2009. Absence of overshadowing between a landmark and geometric cues in a distinctively shaped environment: a test of miller and shettleworth (2007). *J. Exp. Psychol. Anim. Behav. Process.* 35 (3), 357–370 URL <http://www.ncbi.nlm.nih.gov/pubmed/19594281>.
- Miller, N.Y., Shettleworth, S.J., 2007. Learning about environmental geometry: an associative model. *J. Exp. Psychol. Anim. Behav. Process.* 33 (3), 191–212 URL <http://www.ncbi.nlm.nih.gov/pubmed/17620021>.
- Miller, N.Y., Shettleworth, S.J., 2008. An associative model of geometry learning: a modified choice rule. *J. Exp. Psychol. Anim. Behav. Process.* 34 (3), 419–422 URL <http://www.ncbi.nlm.nih.gov/pubmed/18665724>.
- O'Keefe, J., Burgess, N., 1996. Geometric determinants of the place fields of hippocampal neurons. *Nature* 381, 425–428.
- Packard, M.G., McGaugh, J.L., 1996. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol. Learn. Mem.* 65 (1), 65–72.
- Pavlov, I.P., 1927. *Conditioned reflexes*. Oxford University Press, Oxford.
- Pearce, J.M., 2009. The 36th sir frederick bartlett lecture: an associative analysis of spatial learning. *Q. J. Exp. Psychol.* 62 (9), 1665 URL <http://www.informaworld.com/10.1080/17470210902805589>.
- Pearce, J.M., Ward-Robinson, J., Good, M., Fussell, C., Aydin, A., 2001. Influence of a beacon on spatial learning based on the shape of the test environment. *J. Exp. Psychol. Anim. Behav. Process.* 27 (4), 329–344 URL <http://www.ncbi.nlm.nih.gov/pubmed/11676084>.
- Pearce, J.M., Graham, M., Good, M.A., Jones, P.M., McGregor, A., 2006. Potentiation, overshadowing, and blocking of spatial learning based on the shape of the environment. *J. Exp. Psychol. Anim. Behav. Process.* 32 (3), 201–214 URL <http://www.ncbi.nlm.nih.gov/pubmed/16834489>.
- Ragozzino, M.E., Detrick, S., Kesner, R.P., 1999. Involvement of the Prelimbic–Infralimbic areas of the rodent prefrontal cortex in behavioral flexibility for place and response learning. *J. Neurosci.* 19 (11), 4585–4594 URL <http://www.jneurosci.org/cgi/content/abstract/19/11/4585>.
- Redish, A.D., 1999. *Beyond the cognitive map. From place cells to episodic memory*. MIT Press-Bradford Books, London.
- Rescorla, R.A., Wagner, A.R., 1972. A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A.H., Prokasy, W.F. (Eds.), *Classical Conditioning II: Current Research and Theory*. Appleton Century Crofts, New York, pp. 64–99.
- Rich, E.L., Shapiro, M.L., 2007. Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. *J. Neurosci.* 27 (17), 4747–4755.
- Rich, E.L., Shapiro, M., 2009. Rat prefrontal cortical neurons selectively code strategy switches. *J. Neurosci.* 29 (22), 7208–7219.
- Roberts, A.D., Pearce, J.M., 1999. Blocking in the morris swimming pool. *J. Exp. Psychol. Anim. Behav. Process.* 25 (2), 225–235 URL <http://www.ncbi.nlm.nih.gov/pubmed/10331921>.
- Shettleworth, S.J., 1998. *Cognition, Evolution, and Behavior*. Oxford University Press, New York.
- Sheynikhovich, D., Chavarriga, 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.
- Sheynikhovich, D., Dollé, L., Chavarriga, R., Arleo, A., 2010. In: Doncieux, S., Girard, B., Guillot, A., Hallam, J., Meyer, J.-A., Mouret, J.-B. (Eds.), *Minimal model of strategy switching in the plus-maze navigation task. : From Animals to Animats 11*, LNAI, vol. 6226. Springer, Heidelberg, pp. 390–401.
- Sutton, R., Barto, A.G., 1998. *Reinforcement Learning — An Introduction*. MIT Press, Cambridge, MA.
- Tolman, E.C., 1948. Cognitive maps in rats and men. *Psychol. Rev.* 55, 189–208.
- Waelti, P., Dickinson, A., Schultz, W., 2001. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412 (6842), 43–48.
- Wall, P.L., Botly, L.C.P., Black, C.K., Shettleworth, S.J., 2004. The geometric module in the rat: independence of shape and feature learning in a food finding task. *Learn. Behav.* 32 (3), 289–298 URL <http://www.ncbi.nlm.nih.gov/pubmed/15672824>.
- White, N.M., McDonald, R.J., 2002. Multiple parallel memory systems in the brain of the rat. *Neurobiol. Learn. Mem.* 77 (2), 125–184.
- Wiener, S.I., Taube, J.S., 2005. *Head Direction Cells and the Neural Mechanisms of Spatial Orientation*. MIT Press.
- Williams, Z.M., Eskandar, E.N., 2006. Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nat. Neurosci.* 9 (4), 562–568 URL <http://dx.doi.org/10.1038/nn1662>.