

Affordances, Motivation, and the World Graph Theory

Alex Guazzelli¹, Fernando J. Corbacho², Mihail Bota³, and Michael. A. Arbib⁴

USC Brain Project - University of Southern California

Los Angeles, CA 90089-2520

E-mail addresses: (1) aguazzel@rana.usc.edu (2) corbacho@rana.usc.edu

(3) mbota@rana.usc.edu and (4) arbib@pollux.usc.edu

Abstract

O’Keefe and Nadel distinguish two paradigms for navigation, the “locale system” for map-based navigation and the “taxon (behavioral orientation) system” for route navigation. This paper models the taxon system, the map-based system, and their interaction; and argues that the map-based system involves the interaction of hippocampus with other systems.

We relate taxes to the notion of an affordance. Just as a rat may have basic taxes for approaching food or avoiding a bright light, so does it have a wider repertoire of affordances for possible actions associated with immediate sensing of its environment. We propose that affordances are extracted by the rat posterior parietal cortex, which guides action selection by the premotor cortex and is also influenced by hypothalamic drive information.

The Taxon-Affordances Model (TAM) for taxon-based determination of movement direction is based on models of frog detour behavior, with expectations of future reward implemented using reinforcement learning. The specification of the direction of movement is refined by current affordances and motivational information to yield an appropriate course of action.

The World Graph (WG) theory expands the idea of a map by developing the hypothesis that cognitive and motivational states interact. This paper describes an implementation of this theory, the WG model. The integrated TAM-WG model then allows us to explain data on the behavior of rats with and without fornix lesions which disconnect the hippocampus from other neural systems.

Keywords: Affordance, navigation, motivation, hippocampus, parietal cortex, reinforcement learning.

1 Introduction

Whereas local spatial knowledge encodes the location of places to be approached and avoided in a particular limited area, based primarily on current perception and working memory, global spatial knowledge is concerned with transitions from one region to another (Schmajuk & Blair, 1993). Many data suggest that rats (i) have associative memory for complex stimulus configurations, (ii) can encode the spatial effect of their own movements, and (iii) are able to form sequences of actions to go from a starting location to a goal. In other words, rats are assumed to have a *cognitive map* (Tolman, 1948). To understand this notion, we must distinguish *egocentric* representations from *allocentric* representations. The former, based on the organism’s current view of the world, are appropriate for looking, reaching, grasping, and locomotion with respect to directly perceptible features of the landscape. The latter can be understood in terms of a subway map of a city. Such a map is not drawn from the viewpoint of any rider, but nonetheless sets out the spatial relations between different stations, lines that link them, and the junctions between these lines.

O’Keefe and Nadel (1978) distinguished two paradigms for navigation, one based on maps and the other based on routes, and proposed that independent neural systems exist in the brain to support these two types of navigation. They called these systems the *locale* system for map-based navigation and the *taxon* (behavioral orientation) system for route navigation. Our concern is not only to model these systems, but also to analyze their interaction, and the

role of motivation in their execution. The taxon system is based on egocentric spatial information which can be viewed as sets of guidance and orientation cues.

O'Keefe (1983) found that both normal and fornix-lesioned rats can learn to solve a simple T-maze in the absence of any consistent environmental cues other than the T-shape of the maze. If anything, lesioned animals learn this problem faster than normals. All this suggests that without hippocampus the rat uses a pure orientation cue; whereas with hippocampus, there is competition and cooperation between the orientation cue and the cue provided by the map-based system.

After criterion was reached on turns to a specific arm of the T-maze, probe trials with an 8-arm radial maze were interspersed with the usual T-trials. Control and fornix-lesioned animals consistently chose the side to which they were trained on the T-maze, e.g., to the left if food was located in the left arm of the T-maze. However, when in the 8-arm radial maze, many animals did not choose the 90° arm but preferred either the 45° or the 135° arm, suggesting that the rats solved the T-maze by learning to rotate within an egocentric orientation system at the choice point (the maze junction) through approximately 90° (O'Keefe, 1983). This led to the hypothesis of an orientation vector being stored in the animal's brain, with the actual turn depending on environmental cues to determine the appropriate movement most consonant with the encoded orientation.

With the short intertrial intervals used in this study (approximately 1 minute), well-trained lesioned rats readily learn to reverse the T-maze habit, e.g., if food had been located in the left arm, it is then placed in the right arm. During reversal, every third trial was a probe with the eight-arm radial maze to assess the position of the orientation vector at that stage of learning. Fornix-lesioned animals show a steady, incremental shift in the direction of the orientation vector from the original quadrant (left, if trained first with food in the left arm) through straight ahead and into the new reversal quadrant (right, in our example). If observed only in the T-maze, the shift from the incorrect to the correct arm of the T-maze is relatively abrupt. However, as mentioned above, the orientation bias gradually swings from the direction of the originally rewarded turn through straight ahead and over the reversed turn when behavior is observed in the eight-arm radial maze (Figure 7a).

The behavior of control rats during reversal training was different. Although performance on the T-maze proceeded in roughly the same pattern as in the lesioned rats, the underlying orientation vector of the group did not shift in a smooth fashion but jumped around. Examination of the individual curves also failed to reveal any systematic shifts (Figure 13a).

As described by Thinus-Blanc (1996), lesions of the hippocampal formation induce permanent deficits in spatial orientation in a wide variety of situations. Spatial deficits following hippocampal or fornix damage have been observed in the spatial version of the Morris water maze navigation task, with the hidden platform, but not in its nonspatial version in which the platform is visible or signaled by a conspicuous beacon (Morris et al., 1982, 1986; Schenk and Morris, 1985; Sutherland and Rodriguez, 1989; Packard and McGaugh, 1992). Whereas lesions of the hippocampal formation in the rat disrupt performance in various tasks requiring the localization of an inconspicuous goal on the basis of extra-maze landmarks, such lesions do not affect visual guidance toward a goal in close spatial relationship with a cluster of landmarks in the cross maze (O'Keefe and Conway, 1980). Similarly, rats with fornix transections show a selective impairment in the retention of a location discrimination but not in the retention of an object discrimination (Olton et al., 1980).

"Anyone who has used a subway map knows that it is neither Euclidean nor complete, though it must contain sufficient information to plan a reasonably short route from one station to another. Again, a map need not be homogeneous - we may know a route very well from one side of town to another, yet know little of the placement of individual houses save in our immediate neighborhood. Thus the inclusion of landmarks is based on some combination of need and familiarity, rather than some homogeneous criterion for representation." (Arbib, 1997). We suggest that this is true for the rat's cognitive map, too.

To go from a subway map to a cognitive map we must augment the map with the processes needed to use it. To use a subway map, for example, we must be able to locate the representations of where we are and where we want to go on the map, and then find a path on the map (lines and stations to change at) which we can use as we navigate towards our goal. In sum, the ingredients of a cognitive map are (i) an allocentric spatial framework; (ii) a

mechanism for the animal to locate itself and places of interest (e.g., goals) in that framework; and (iii) the means to combine current location and intended movement to infer new location, or to combine current location and desired location to infer movement.

Single-unit recordings from CA3 and CA1 pyramidal cells in the hippocampus show a high correlation between spike rate and the location of the animal in the environment, hence the cells are called *place cells*. These cells were found to fire at an elevated rate over a continuous, compact area, called the place field of the cell (O'Keefe & Dostrovsky, 1971). However, place cells show direction-dependent firing fields in some specific types of tasks, particularly those with restricted movement paths as in a maze (McNaughton et al., 1983; Markus et al., 1995). Nevertheless, there is still no evidence that the hippocampus proper (Dentate gyrus and Ammon's horn) can simultaneously encode the rat's current location and the goal of current navigation. In other words, it may provide the "you are here" function of a map, but not the "this is where you are going" and "how to get there" functions, which thus must depend on a larger system of which the hippocampus proper is but one part.

Arbib and Lieblisch (1977, see also Lieblisch and Arbib, 1982) represented the cognitive map as a graph, i.e., a set of nodes some of which are connected by edges, with nodes corresponding to a recognizable situation in the animal's world (cf. the stations on a subway map), and with each edge representing a path from a recognizable situation to the next. This "world graph" (WG) is constructed so that the organism has the ability to move from one point to another in a fashion determined by both cognitive knowledge and motivational states. The WG theory can then be used to analyze how the brain can encode sequences of actions required to pass from one spatial representation to another and, likewise, how an animal builds up a model of the world around it in such a way that, when under the influence of a particular drive, it can find its way to a location where it has learned that the drive will be reduced. A distinctive feature of the WG model is that "... a single place in the world may be represented by more than one node in the graph - as a result, for example, of the animal coming upon a place for the second time but not recognizing that it has been there before, perhaps because it encounters the place in a very different situation or motivational state."

In the next section, on the basis of available neurophysiological data in monkey and rat, visual affordances for movement are described and considered to be implemented by the parietal cortex. (Affordances also involve subcortical mechanisms, but these are beyond the scope of the present paper.) Section 3 deals with motivation and describes the way this is implemented and how it provides relevant information for goal-oriented navigation. A model of frog detour behavior provides the basis for the Taxon-Affordances Model (TAM) proposed in section 4, which also describes the behavioral experiment used for validation of TAM. The implementation of the WG theory is described in section 5. The integrated TAM-WG model is presented in section 6. Section 7 expands the results shown in section 5 and describes further experiments and their neural simulations with the use of the integrated model. Section 8 contains the conclusion.

2 Visual Affordances for Movement

Gibson (1966) observed that the pattern of optic flow contains valuable information that can be used to guide navigation through the environment without prior recognition of objects. In this way, Gibson made clear that there was a great deal of information which could be "picked up" for use by an animal even without the invocation of object recognition. We thus adopt Gibson's term *affordances* for parameters for motor interactions signaled by sensory cues without the necessary intervention of object recognition. Unlike Gibson, however, we view this "pick up" not as "direct" but rather as involving neural mechanisms, at least some of which are embedded in the parietal cortex.

We relate the taxon to the notion of affordances for movement. Just as a rat may have basic taxes for approaching food or avoiding a bright light, so does it have a wider repertoire of affordances for possible actions associated with the immediate sensing of its environment. In this sense, affordances would then include: for a dark hole - hide; for visual sighting of a corridor - go straight ahead; for sensed branches in a maze or the sight of the end of the corridor - turn; for food as sensed generically - eat; etc.

McNaughton et al. (1989; McNaughton et al., 1994a) found cells in rat posterior parietal cortex (PPC) that fired before or as the animal was executing a specific kind of movement. Roughly 40% of the cells had responses discriminating whether the animal was turning left, turning right, or moving forward. Very few cells required a

conjunction of movement and location, e.g., one parietal cell fired more for a right turn at the western arm of a cross-maze than for a right turn at the eastern arm, and these firings were far greater than for all left turns. However, in general this location information is more generic than that provided by the hippocampal place cells. McNaughton et al. reported that turn-direction information was varied, with a given cell responding to a particular subset of vestibular input, neck and trunk proprioception, visual field motion and possibly efference copy from motor commands. For Arbib et al. (1997), however, this location information is, when it precedes the movement, better seen as affordance information, signaling recognition of situations where a certain range of motor behaviors is possible, i.e., the rat parietal cortex would encode affordances for movement. On the other hand, those cells that fire after the rat has begun to turn can be thought of as turn detectors, since they are similar to rotation sensitive neurons found in area 7a of the monkey posterior parietal cortex (Sakata et al., 1994). Arbib (1997) argues that these cells monitor the animal's progress against its expectations so that the animal can correct its actions and/or internal representations when discrepancies arise.

The idea of PPC cells computing affordances for locomotion was inspired by the situation for grasping (the FARS model: Fagg, 1996 and Fagg & Arbib, 1998). In grasping, parietal cells can use vision of an object to determine an appropriate grip, which can then be signaled to premotor cortex. Further computations, including frontal activity modulated by the basal ganglia, are needed to determine which action will be executed. The FARS model includes several cortical regions in monkey that are involved in planning and execution of a coordinated motor program for grasping, such as the anterior, posterior and ventral intraparietal areas (AIP, PIP, and VIP), the supplementary motor area, frontal area 46, inferior premotor cortex areas F5 and F4, and dorsal premotor cortex. Within AIP, cells are selective to a variety of grasp-related object properties, as are F5 neurons (Rizzolatti et al., 1988). AIP cells also show significant movement-related modulation (based on a return path from F5) as the grasp is executed. In the FARS model, AIP serves a dual role of first computing a set of affordances that are appropriate for the object being attended, and then maintaining an active memory of the single affordance as the corresponding grasp is executed.

However, we now leave affordances aside for a while, as we consider the notion of motivation, and its implementation as a motivational schema with a dual role: setting current goals and providing reinforcement for learning in TAM (Section 4) and in the WG model (Section 5).

3 Motivation

A motivational system is required to select goals to be pursued and to organize the animal's commerce with appropriate goal objects. Motivated behavior is usually goal-oriented; the goal may be associated with a drive, such as hunger or thirst. However, the effect of motivation is closely tied to sensory stimuli, e.g., an animal will not usually exhibit eating behavior unless food is present. Moreover, the motivation to eat is not controlled solely by feelings of hunger (when presented with food, animals will often eat in anticipation of hunger and continue to eat after satiation to maintain themselves until the next meal). An external factor, like the sight of food, can play a role in stimulating motivation, and for this reason can be called an incentive. For Toates (1986), motivation arises as a function of both internal state and incentive; a motivational system is one that selects goals to be pursued and organizes the animal's commerce with appropriate goal objects. Motivation played a significant role in many theories of behavior, especially Hull's (1943) theory which proposed that motivation is the initiation of learned, or habitual, patterns of movement or behavior. In Hull's theory, events that threaten survival give rise to internal drive states, and behaviors that reduce the drive are rewarding. In this way, lack of food causes an increase in the hunger drive, and the consumption of food is rewarding because it leads to a reduction in the hunger drive.

Neurally speaking, the basic motivation-related circuitry is located in the hypothalamus. Moreover, the hypothalamus has been divided into functionally related circuitry (subsystems) for eating, drinking, sex, and temperature regulation (Swanson and Mogenson, 1981). The hypothalamus can then be assumed to be the primary place where the information about the internal state of the animal is combined with incentive information, although this role can be shared with other systems. The outcome of such a process would be represented as drives with specific levels. In the systems level model proposed by Arbib et al. (1997), goals and motivational information are passed to prefrontal cortex and basal ganglia. These are considered to form a "potential field" planner in the sense that food, for example, is coded as an attractant field if the animal is hungry. In favor of this hypothesis, we mention the triangle of connections between nuclei of hypothalamus, thalamus, and prefrontal cortex (frontal eye fields) that

appears to be involved in attentional behavior and in head and eye movements (Shibata & Kato, 1993; Risold & Swanson, 1995).

So far, most models of spatial learning and navigation do not incorporate the role of motivation into their computation. By contrast, the WG theory posits a set d_1, d_2, \dots, d_k of discrete drives to control the animal's behavior. At time t , each drive d_i in (d_1, \dots, d_k) has a value $d_i(t)$. Drives can be *appetitive* or *aversive*. The idea is that each appetitive drive spontaneously increases with time towards d_{max} while aversive drives are reduced towards 0, both according to a factor α_d intrinsic to the animal. An additional increase occurs if an incentive $I(d, x, t)$ is present such as the aroma of food in the case of hunger. Drive reduction $a(d, x, t)$ takes place in the presence of some substrate - ingestion of water reduces the thirst drive. If the animal is at the place it recognizes as being node x at time t , and the value of drive d at that time is $d(t)$, then (in the model of Arbib and Liebllich, 1977) the value of d for the animal at time $t+1$ will be

$$d(t+1) = d(t) + \alpha_d |d_{max} - d(t)| - a(d, x, t) |d(t)| + I(d, x, t) |d_{max} - d(t)|$$

This is the drive dynamics incorporated into the motivational schema.

Moreover, our motivational schema is not only involved in setting current goals for navigation; it is also involved in providing reinforcement for learning. Midbrain dopamine systems are crucially involved in motivational processes underlying the learning and execution of goal-directed behavior. Schultz et al. (1995; Mireniewicz and Schultz, 1996; Apicella et al., 1991) found that dopamine neurons in monkeys are uniformly activated by unpredicted appetitive stimuli such as food and liquid rewards and conditioned reward-predicting stimuli (see Lavoie and Mizumori, 1994, and Wiener, 1993 for reward neurons in rats). Schultz et al. (1995) hypothesized that dopamine neurons may mediate the role of unexpected rewards to bring about learning. For them, the information conveyed by the dopamine neurons can be viewed as an error signal, which suggests that the dopamine signal could be an important substrate for reinforcement learning. In contrast to the responses of dopamine neurons to unexpected rewards, neurons in the striatum and in structures projecting to the striatum, such as the amygdala and orbitofrontal and anterior cingulate cortex, respond to primary rewards in well-established behavioral tasks and are not affected by the uncertainty of reward. This suggests that different aspects of the reward signal are distributed over different neuronal systems in the brain and that the function of reward signals accordingly may vary among these systems: "The neuronal response to reward in striatal, amygdalar, and cortical neurons during well-established behaviors suggests a function in maintaining rather than bringing about learning."

In the present work, reward information is used to learn *expectations* of future reinforcements. Furthermore, the motivational schema implements the idea that the amount of reward is dependent on the current motivational state of the animal. If the rat is extremely hungry, the presence of food might be very rewarding, but if not, it will be less and less rewarding. In this way, the influence of a reward $r(t)$ in the WG model and TAM will depend on the value of the animal's current dominant drive $d(t)$ (the maximum drive amongst all drives at a particular time t), its corresponding maximum value, and the actual drive reduction, according to:

$$r(t) = \frac{d(t)}{d_{max}} a(d, x, t)$$

In this way, reward will bring about or maintain learning in the WG model and TAM depending on the animal's current motivational state. On the other hand, if the rat reaches the end of a corridor and does not find food, the drive reduction factor $a(d, x, t)$ is set equal to -0.1 times the normal value $a^*(d, x)$ it would have if food were present, and the motivational schema uses the above formula to generate a negative value of $r(t)$, a frustration factor, that will cause unlearning in TAM and in the WG model.

4 TAM: The Taxon-Affordances Model

We now develop our TAM model of the rat's Taxon-Affordances system. Surprisingly, it is based on a model of detour behavior in frogs. Figure 1 shows the assemblage of TAM components. Sensory input provides the system with incentives and current affordances, as well as the sensing of drive relevant stimuli, like food. Parietal

affordances, motivational information and drive-related stimuli are then combined in the premotor cortex where the selection of the next movement takes place.

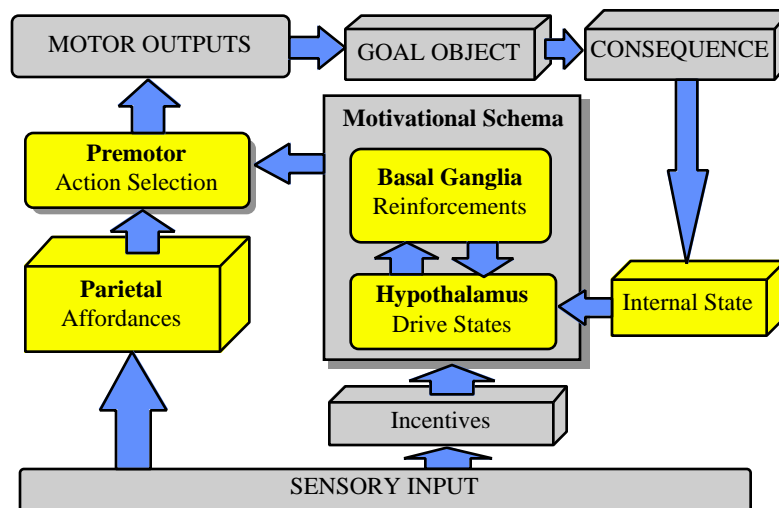


Figure 1: Assemblage of TAM components. Parietal affordances, motivational information and drive-related stimuli are combined in the premotor cortex where the selection of the next movement takes place.

Ingle (1980) and Collett (1982) have observed that a frog/toad's approach to prey or avoidance of a threat are also influenced by the stationary objects in the animal's surround. Collett (1982) presented toads with different prey-barrier configurations and analyzed the number of trials in which toads elected to head directly towards the prey versus the number of trials in which they elected to detour around either of the fence ends (Figure 2).

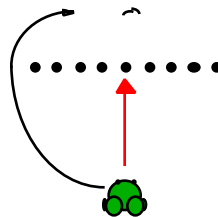


Figure 2: Toad detour behavior. Arrows show two different trajectories elected by toads while approaching the prey behind a barrier (the row of dots).

The determination of movement direction in the present model is implemented by a reinterpretation of the models of Arbib & House (1987) and Corbacho & Arbib (1995). In both, each prey object was represented by a peak of excitation coarsely coded across a population, while barriers to movement were encoded by inhibition whose extent was closely linked to the retinotopic extent of each barrier. The resultant sum of excitation was passed through a winner-take-all circuit to yield the choice of movement direction. Similarly, Schmajuk and colleagues (1990, 1992, 1993) describe global and local cognitive mapping using a neural network that assumes animals perceive the visual angles of different objects and build a spatial generalization surface around appetitive and aversive stimuli. Based on this generalization surface, and making use of a stimulus approach principle, the network is able to describe place learning independently of the responses used to approach the goal.

In the current work, we use a model similar to the frog models described above for the behavioral orientation system by replacing the direction of the prey (frog) by the direction of the orientation vector (rat), while the barriers correspond to directions in which no arm of the maze is visible, i.e., the presence of an obstacle or wall constrains the possible movements of the animal. Visible arms are encoded by the use of excitatory fields, since this seemed to accord with the idea of affordances, e.g., go straight if an opening exists.

The implemented model employs coarse coding in a linear array of cells, coded for turns from -180° to $+180^\circ$. In this way, when in the dais of an eight-arm radial maze, the rat is able to sense eight different visible arms and

consequently, eight different affordances (since a rat's visual field covers 320°, we assume it is able to visually sense the turning affordance given by the arm behind it) - Figure 3.

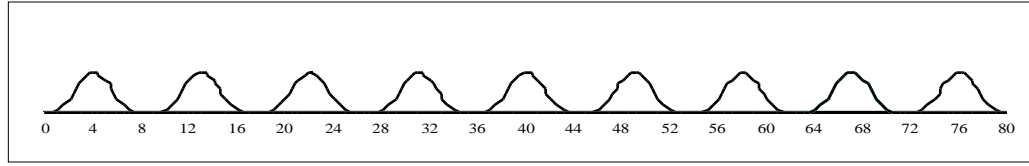


Figure 3: Available affordances when the rat is on the dais of the simulated eight-arm radial maze. Each peak of activity represents a different affordance. Note that activity is coded in an array of 80 cells - the x-coordinate is thus cell index, not number of degrees. The left most peak codes for turning 180° to the left and the right most peak codes for turning 180° to the right. The remaining peaks of activity code for turns in between -180° and +180° in 45° intervals.

Recall the discussion in Section 1 of reversal of the T-maze habit in normal and fornix-lesioned rats. From the relatively smooth changing of the orientation vector shown by the fornix-lesioned animals, one might expect that, in their case, only the cells close to the preferred behavioral direction are excited, and that learning “marches” this peak, called “rewardness expectation” in the present work, from the old to the new preferred direction. However, the peak in our model does not actually march. Rather, it is the mean of 2 peaks that moves. During reversal of the T-maze, the reinforcement learning rule, implemented in the present model by the use of temporal differences in an actor-critic architecture (Barto et al., 1983; Barto, 1994; Sutton, 1988), will “unlearn” -90° (turn to the left in the T-junction), by reducing the peak there, while at the same time “building” a new peak at the new direction of +90° (turn to the right in the T-junction). If the old reinforcement peak has “mass” $p(t)$ and the new reinforcement peak has “mass” $q(t)$, then as $p(t)$ declines toward zero while $q(t)$ increases steadily from zero, the center of mass, the animal’s rewardness expectation:

$$\frac{(-90)p(t) + 90q(t)}{p(t) + q(t)}$$

will progress from -90 to +90.

As described in Section 3, hunger level in our model is represented as a state in the motivational schema, with an initial value, a maximum value, and a rate of increase. In wandering around the maze, the rat becomes hungrier and hungrier. When food is detected, it serves as an incentive and the rat will become even more hungry. The presence of food will also be represented by excitation coarsely coded across a population of neurons, which will influence the choice of movement direction. Consumption of food reduces the hunger level (Figure 4). The finding of food is also used as the only rewarding event by the reinforcement learning rule. In this case, the $r(t)$ value, supplied by the motivational schema, is used as the reinforcement signal. On the other hand, if the rat reaches the end of a corridor and does not find food, the frustration factor $u(t)$, also supplied by the motivational schema, is used by the reinforcement rule as a punishment signal.

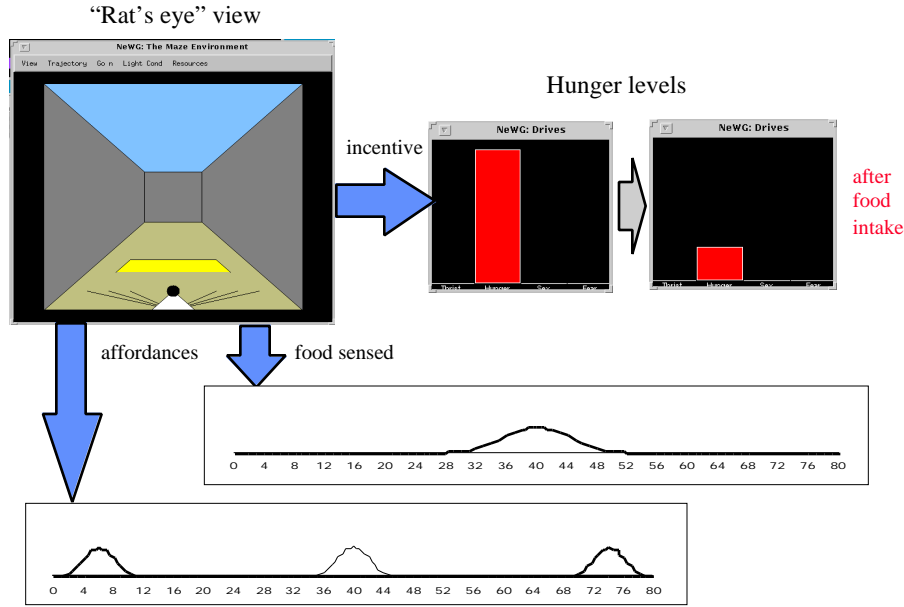


Figure 4: “Rat’s eye” view at the end of an arm (in this case, food is present and is represented as a rectangle in front of the rat). When food is sensed, it is represented by excitation coarsely coded across a population of neurons. Its presence is also coded as an incentive which will raise the hunger level. In the situation depicted, three affordances (peaks of activity) are available: go straight ahead or turn 180° right or left (the rat’s eye view shown here does not reflect the actual computed field of view of our simulated rat; it is used for interface purposes). By deciding to go straight ahead and consequently eating the food, the rat will reduce its hunger level.

In TAM, reinforcement learning is used as a way of learning expectations of future reinforcement associated with state/action pairs, e.g., the expectation associated with turning left at the T-junction is increased if the rat finds food at the end of the left arm. The reinforcement learning system tries to influence the behavior of the rat to maximize the sum of the reinforcement that will be received over time. We follow the usual convention in applying temporal differences in an actor-critic architecture. Because the sum of future reinforcement might be infinite in some problems, and because the learning system usually has control only over its expected value, an expected discounted sum is considered instead:

$$E\{r(t) + \gamma r(t+1) + \gamma^2 r(t+2) + \dots\} = E \sum_{k=0}^{\infty} \gamma^k r(t+k)$$

where E is the expectation over all possible future behaviors of our simulated rat and γ is a discount factor that determines the present value of future reinforcement. The output $P(t)$ of an adaptive critic unit is a prediction of this discounted sum and, for this reason, should satisfy:

$$P(t) = E\{r(t)\} + \gamma P(t+1)$$

An estimate of the error by which any two adjacent predictions fail to satisfy this consistency condition is called the temporal difference (TD) error (Sutton, 1988):

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

In an actor-critic architecture, the predictions formed by an adaptive critic act as reinforcement for an associative reinforcement learning component, called the actor. To distinguish the adaptive critic’s signal from the reinforcement signal $r(t)$ supplied by the nonadaptive critic (the motivational schema in our case), Barto and Sutton (1983) call $\hat{r}(t)$ the internal reinforcement signal. The actor tries to maximize the immediate internal reinforcement signal, while the adaptive critic tries to predict total reinforcement. The adaptive critic unit adjusts its weights according to the following learning rule:

$$w(t) = \alpha \hat{r}(t) e(t)$$

where α is the learning rate, and $e(t)$ is an eligibility trace - Klopff (1972) suggested that a pathway should reach maximum eligibility a short time after the occurrence of a pairing of a nonzero input signal on that pathway with the firing of the element. Eligibility should decay thereafter toward zero. In the current model, the eligibility trace associated to a state/action pair decays exponentially according to:

$$e(t + 1) = \delta e(t) + (1 - \delta)y(t)x(t)$$

where δ , $0 < \delta < 1$, determines the trace decay rate, $y(t)$ is the action chosen at time t , and $x(t)$ is the current state.

When in the T-junction, the animal will learn to go towards the arm containing the food reward, since, for example, if the rat finds food on the left arm, the action of turning to the left will be eligible to be reinforced. In the case shown in Figure 5, the rat has already learned that food can be encountered in the left arm, which is represented by the presence of a peak of activity in the same position in the reinforcement array as the affordance coding for -90° (turn to the left). As the rat has not yet experienced the reversal case, its “rewardness expectation” activity will be the same as its reinforcement activity.

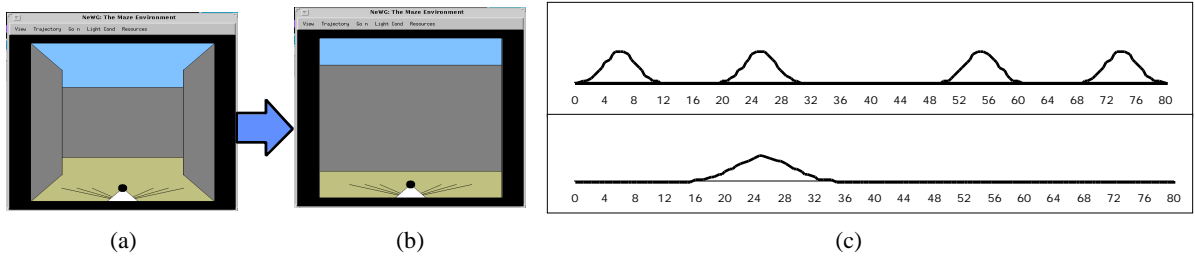


Figure 5: (a) Rat's eye view one simulated step before the animal arrives at the T-junction. (b) Situation when the rat arrives at the T-junction of the maze. (c) Coarse coded representations of (top) affordances (four peaks of activity can be seen: 90° and 180° - left and right) and (bottom) expectation of future reinforcement (a peak of activity in the same position as the affordance to turn 90° leftward).

The decision to turn to a certain angle (the orientation vector) is given by a winner-take-all process performed over the integration of fields produced by the available affordances (AF), drive relevant stimuli (ST), e.g., food if sensed, and rewardness expectation (RE), which contains an associated noise factor, $noise_re$. In our model, then, the total input I_{in} to the action selection module becomes

$$I_{in}(i, t) = AF(i, t) + ST(i, t) + (RE(i, t) + noise_re)$$

where i varies from 1 to N , the length of our population of cells in the linear array. In the figures show in this paper, N is always equal to eighty.

Under the influence of the rewardness expectation activity, the animal's turn will then be the available direction closest to that encoded by the “center of mass” of the neural activity in its reinforcement field as shown in Figure 6a. In this way, during reversal, the animal will march the peak from left to right (Figure 6b).

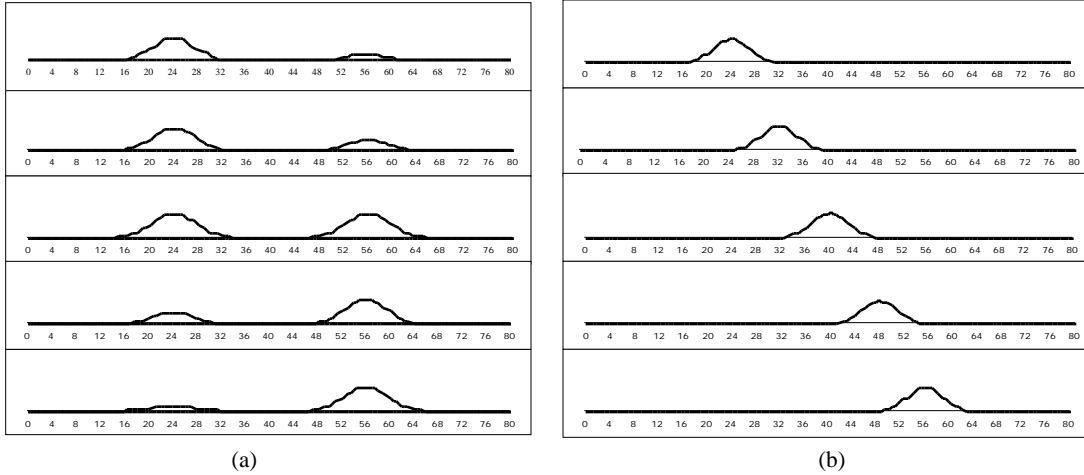


Figure 6: (a) Peaks of reinforcement activity during reversal of the T-maze. Each activity bump (computed from the reinforcement weight w) represents a previously reinforced action associated with the current state. The activity bump on the left is in fact decreasing. However, due to a normalization process used to generate the activity plots, it appears to decrease only when the activity bump in the right passes it in magnitude. Activity plots are depicted in intervals of four trials, approximately. (b) Marching from left to right of the rewardness expectation activity, $RE(i, t)$. Each activity bump represents the coarse coded center of mass of its correspondent (from top to bottom) peaks of activity displayed in (a), where the maximum values of the left and right bumps are q and p , respectively.

Figure 7a depicts the performance of fornix-lesioned animals (O'Keefe, 1983). The simulation results (Figure 7b) replicate these data by the use of coarse coding and reinforcement learning. In Figure 7, the top graphs show that the shift from the incorrect to the correct arm of the T-maze is reasonably abrupt in real and simulated rats; the bottom graphs show that the orientation bias gradually swings from the direction of the originally rewarded turn through straight ahead and over to the reversed turn. Figure 8 shows the performance of the fornix-lesioned rat which most clearly showed the incremental shift in its orientation vector. We stress however that we see the orientation vector as, in this case, representing the mean of two peaks of activity.

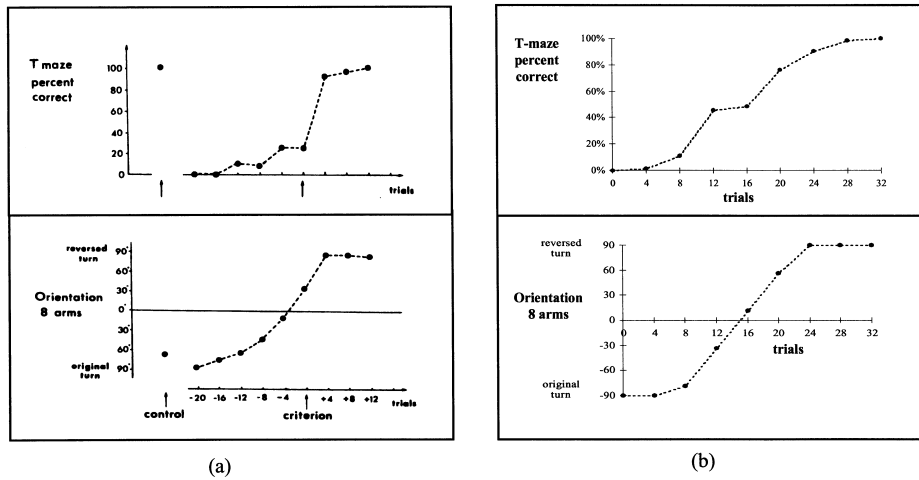


Figure 7: Top graph: Percentage of correct choices on the T-maze task. Bottom Graph: Choice of arm in the radial maze task. (a) The performance of six rats with fornix lesions during reversal on a T-maze. Each graph was obtained by averaging the graphs of the individual animals backwards and forwards from the beginning of the criterion run. Extracted from O'Keefe (1983). (b) The performance of ten TAM simulated rats during reversal on a T-maze. The bottom graph was obtained by computing the medians amongst animals and trials (four by four). The top graph was obtained by averaging the performance of each animal over periods of four trials.

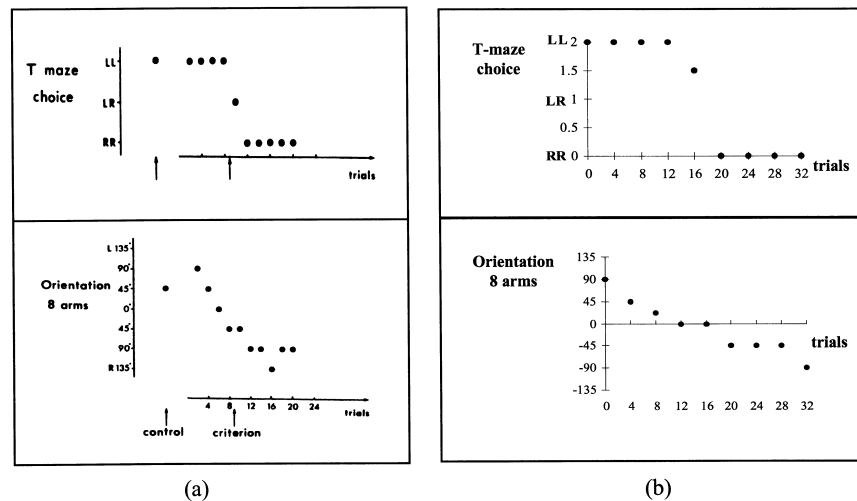


Figure 8: Top graph: Choice of left vs. right in the T-maze task. Bottom Graph: Choice of arm in the radial maze task. (a) The reversal performance of the fornix-lesioned rat which most clearly showed the incremental shift in its orientation vector. Extracted from O'Keefe (1983). (b) The reversal performance of the TAM simulated rat which most clearly showed the incremental shift in its orientation vector. In the top-graph, the values were obtained by computing the average over periods of four trials. In this case, a value of 2 is obtained for eight left-turns, 1.5 for 6 right-turns, and so on.

For O'Keefe (1983), the fact that the shift in behavior on the T-maze occurs around the point that the orientation bias crosses the midline suggests that the former is caused by the latter. The same shift in behavior can also be observed in the data gathered from simulated animals. However, the rats in O'Keefe's experiment did go to the right side of the maze (a few times) even before the orientation vector had crossed the midline (Figure 7a). In our simulations, the rat showed the same behavior, but this behavior actually accelerated the march of the reinforcement peak towards the right. Based on this, we prefer to assume that the shift in behavior and the shift in the orientation vector are tightly related to each other in a mutually causal relationship.

5 The World Graph Model

The World Graph model incorporates in its architecture and dynamics the role of head direction and place cells and the notion of drives. It is strongly attached to the WG theory, which gives the model many of its features. In order to describe the WG model, we first describe the kinds of inputs presented to it, how these are processed, and how these inputs are finally incorporated into the model dynamics.

Like Poucet (1993), we hold that the topological and metric properties of space provide the basic information of animal spatial representations. Topology provides the relations of compactness, neighborhood, enclosure, and connectivity. In the present model, as well as in the system proposed by Poucet (1993), topological information affords a loose, yet operational representation that specifies the connectivity of space and its overall arrangement. This basic representation is enhanced by providing metric information about angles and distances. Poucet (1993) provides extended discussion of how animals might compute and represent metric and topological information. Some animals appear to integrate sensory signals provided by angular and linear displacements as they traverse a path so that they can return directly to their starting point even in total darkness (Mittelstaedt & Mittelstaedt, 1980). In such path integration, signals generated during locomotion allow the subject to update its position relative to its point of departure. This kinesthetic information includes at least vestibular, somatosensory, and proprioceptive components (Save and Moghaddam, 1996). However, as pointed out by Etienne et al. (1996), without the help of external spatial cues, and in particular without an external compass, path integration by itself is not precise enough to be used beyond limited excursions.

Cells that show a unimodal tuning to head direction independent of location and relative to a "compass direction" have been reported in several areas of the rat brain, including the postsubiculum (Taube et al., 1990a,b) and the posterior parietal cortex - PPC (Chen et al., 1994a,b). Head direction cells provide the animal with an internal representation of the direction of its motion. When landmarks in a familiar environment are rotated about the animal,

tuning curves of head direction cells rotate accordingly (Taube et al., 1990b; McNaughton et al., 1994b; Goodridge and Taube, 1995; Knierim et al., 1995; Taube, 1995), which suggests that the compass direction can be determined by familiar landmarks. However, in unfamiliar environments, vestibular information dominates and the rat does not respond to the rotation of landmarks. Like place cells (McNaughton et al., 1989; Quirk et al., 1990; Markus et al., 1994), head direction cells continue to fire in the dark. But if the animal wanders about in a darkened circular arena with no other positional cues available, they eventually drift. Hippocampal place fields appear to drift in synchrony with the head direction system (Knierim et al., 1993; McNaughton et al., 1994b). For Touretzky and Redish (1996), the information conveyed by head direction cells contributes to the operation of a path integrator. As proposed by McNaughton et al. (1996; Samsonovich, 1997; Samsonovich & McNaughton, 1997), head direction is responsible for the implementation of angular path integration.

Although incorporating the role of head direction cells, the model presented here does not implement path integration. However, it is devised to do so in the near future by means of dynamic remapping (Dominey and Arbib, 1992). Arbib (1997) notes the existence of a strong similarity of dynamic remapping (updating the representation of a target relative to the fovea as the eye moves) with path integration (updating the representation of “home base” relative to the animal as the animal moves), in each case shifting relative position in a direction equal and opposite to the intervening movement.

The WG model receives three kinds of inputs. These are given by three different systems: the place layer, the motivational schema, and the action selection schema. The WG model is composed of three distinct layers (Figure 9). The first layer contains coarse coded representations of the current sensory stimuli. This layer comprises four different perceptual schemas, which are responsible for coding head direction, walls, and landmarks (bearings and distances), respectively. Perceptual schemas incorporate “what” and “where” information, the kind of information assumed to be extracted by the parietal and infero-temporal cortices (Mishkin et al., 1983; Kolb et al., 1994) to represent the perception of environmental landmarks and maze walls. In this model, perceptual schemas, like affordances, employ coarse coding in a linear one-dimensional array of cells to represent pertinent sensory input. In this case, the height of a wall or an obstacle (high enough that the rat could not jump over it), for example, does not affect the rat’s behavior, and so only the extent of the wall needs to be coded.

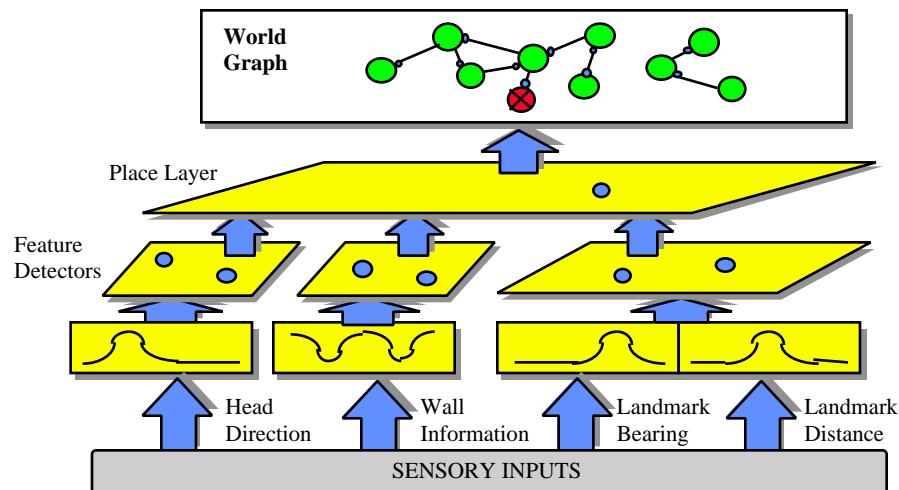


Figure 9: The WG model components. Sensory inputs are coarse coded over linear arrays of cells, our perceptual schemas. These send information to feature detector layers where the information is contrast-enhanced by a competitive process. The set of feature detector winners are, in turn, used by the Place Layer to determine the current place. Current place information is then used by the World Graph layer to determine its current active node, or, conversely, to determine if a new node must be created. In the current implementation, the World Graph is non-neural, and is represented as nodes connected by edges. Edges are directional and this is represented by a little circle attached to the nodes they point to. The currently active node is darker than the other nodes and has an X inside. In case of more than one landmark, new landmark feature detectors and perceptual schemas are added to the model for each landmark.

In the WG model, two distinct perceptual schemas are responsible for presenting landmark information to the network. We show the case for a single landmark. In general, new perceptual schemas are added to the model for each landmark. However, since the experiments described in this paper are landmark free, the importance of landmarks for navigation and their further WG modeling will be described in another paper. In the current implementation, each

landmark will have two perceptual schemas associated to it: one to code landmark bearing and the other to code landmark distance. These will together form the schema responsible for presenting landmark bearing and distance information to the corresponding landmark feature detector layer. A landmark is visible if it is inside a 320° field of view. The separate coding of bearings and distances to landmarks is also used by Touretzky and Redish (1996). In their model, bearings and distances to landmarks together with Cartesian coordinates and bearing differences between pairs of landmarks are used to build a place code representation of the environment.

Moreover, walls are also responsible for the formation of the rat's local view. O'Keefe and Burgess (1996) show place cells responsive to the walls of a box environment that modify their place fields in relation to different configurations of the environment. Muller et al. (1987) observed that, in some cases, place cells had a crescent-shaped field and were assumed to be related to the arena wall. In the WG model, if the wall is close to the rat (less than a body length in the current implementation), it will be represented in the "wall perceptual schema". Figure 10 (top) shows the wall perceptual schema when the rat is in the middle of a corridor. In fact, the wall representation is the opposite of the affordances representation (Figure 10, bottom).

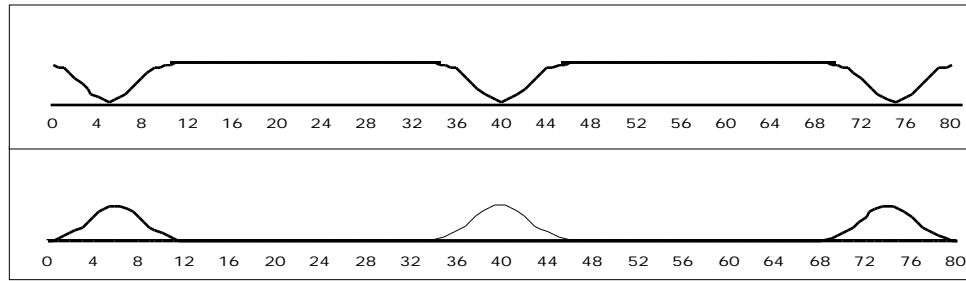


Figure 10: Top diagram shows the activation of the wall perceptual schema when the rat is in the middle of a corridor. Affordances, which are the opposite representation of the walls, are shown in the bottom diagram.

As mentioned above, place cells show direction-dependent firing fields only in some specific types of tasks, particularly those with restricted movement paths. For example, place fields tend to be directional on a radial arm maze (McNaughton et al., 1983), while directional place fields are rarely seen in an open arena (Muller et al., 1991; Muller et al., 1994). In this case, however, a task requiring restricted movement paths produces more directional place fields than one with unrestricted paths (Markus et al., 1995). Although the directionality of place fields is yet to be fully understood, the WG model incorporates the role of head direction cells into its computation. This information, like the other perceptual inputs, will be coarse coded as a bump of activity over a linear array of cells, the head direction perceptual schema. Figure 11 shows the activation of the head direction perceptual schema when the rat is facing 90° west (west may not mean, however, the magnetic west, but the allocentric direction based on the rat's internal, e.g., landmark-based, compass).

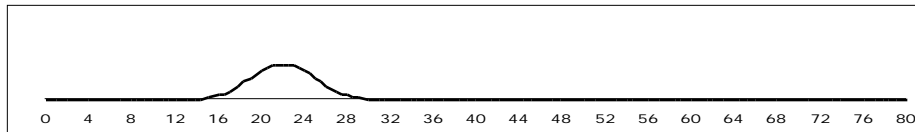


Figure 11: Peak of activation in the head direction perceptual schema when the simulated rat is facing 90° west.

Each perceptual schema projects in turn to its respective feature detector layer. Each of these layers is composed, in the current implementation, of 100 nodes which are organized in a two dimensional matrix. Each node in the perceptual schema connects to all nodes in its corresponding feature detector layer, i.e., the two are fully connected. The activity level of the feature detector unit (F_j) is computed according to the following equation:

$$F_j = \sum_i (V_i w_{ij}) + N_j$$

where V_i is the activity level of input unit i , N_j is a random signal that is injected into feature detector j , and w_{ij} is the weight of the connection from input i to detector j .

These feature detector units then interact through a local competition mechanism to contrast-enhance the incoming activity pattern, i.e., a neuron in a feature detector layer produces a non-zero output (G_j) if and only if it is the most active neuron within a neighborhood. The adjustments to the connection weights between any pair of perceptual schema-feature detector layers are done using a normalized Hebbian learning algorithm. This ensures that the same set of feature detectors is activated by increasing the connection strength from active units (V_i) in the perceptual schema to active feature detectors thereby increasing their response level the next time. This rule is captured in the following connection strength update equation:

$$\dot{w}_{ij} = \alpha V_i G_j w_{ij}$$

where \dot{w}_{ij} is the change to the connection strength w_{ij} and α is the learning rate parameter. In a second step, the weights are normalized according to:

$$w_{ij} = \frac{w_{ij} + \dot{w}_{ij}}{(w_{ij} + \dot{w}_{ij})_i}$$

The combined weight update and normalization processes not only maintains individual weights within a limited range, but also implements a form of competition between the different connections. When a set of input connections are incremented relative to a single feature detector unit, the remaining connections are subsequently decremented. This encourages the development of feature detector units that are responsive to a small number of situations.

Placed atop the feature detector layers is the Place Layer. This layer will keep track of the feature detector winners and will activate a different place/node if the set of winners that becomes active in the feature detector layers does not activate any previously activated node in the place layer. (The matching of winners is based on a pre-determined threshold.) As mentioned above, in the WG theory there is a node for each distinct place/situation the animal experiences. This is implemented in the WG model by a functional layer, called the World Graph layer, placed above the Place Layer. In this way, for each “new” set of feature detector winners, a “new” node is activated in the place layer and, consequently, a node x is created in the world graph layer of the WG model.

In the WG theory, there is an edge from node x to node x' in the graph for each distinct and direct path the animal has traversed between a place it recognizes as x and a place it recognizes as x' . This is implemented in the WG model as a link between two nodes of the World Graph layer. If, for example, a node x is currently active at time t , but by deciding to move towards north the animal activates a “different” set of feature detectors and so a distinct node x' in the world graph at time $t+1$, a link will be created from node x to node x' . Appended to each link/edge will be sensorimotor features associated with the corresponding movement/path. In the current model, these features consist of a motor efference copy (e.g., movement towards north) supplied to the world graph via a feedback connection from the action selection schema.

The current view and the head-direction information will together let the animal determine whether it is experiencing a place already represented by a particular node x in the world graph or whether a new node must be allocated. In the latter case, a new node x' will be added to the world graph to represent the new place, together with an edge from x to x' .

Edge information allows the animal to perform goal-oriented behaviors. In the WG theory, this is obtained by incorporating expected drive reduction information to each node of the world graph in the form a vector. The WG model, however, is able to learn expectations of future reward. As in TAM, this is implemented by the use of temporal differences learning in an actor-critic architecture (Barto et al., 1983; Barto, 1994; Sutton, 1988). In the WG model, however, expectations of future reinforcement are associated with pairs of nodes/edges, and not with state/actions as is the case in TAM. For this reason, when a particular node of the world graph is active and, for example, the simulated animal is hungry, all it needs to do is to select the edge containing the biggest hunger reduction expectation and follow its direction towards the node it points to. In the WG model, each drive (hunger, thirst, sex, and fear) will have a drive reduction expectation value associated to it. In this way, if the rat is thirsty, it

will not base the next node selection on hunger reduction expectations, but on thirst reduction expectations. Reinforcement learning schemes for navigation (see Barto and Sutton, 1981; Dayan, 1991) can produce optimal trajectories, even in cluttered environments. However, reinforcement learning, if used as the main learning strategy, requires almost exhaustive exploration of the environment. Whenever a goal or obstacle is added to the environment, learning must begin again from scratch. Some latent learning is incorporated in Dayan (1991) that would reduce learning time somewhat if a goal were added, but not if an obstacle were added. As pointed out by Burgess et al. (1994), the rat's navigation is better characterized by fast learning and good, but not necessarily optimal, trajectories.

6 The Integrated Model: TAM and the WG model Working Together

The integrated model, TAM-WG, joins together TAM (Figure 1) and the WG model (Figure 9). The result is depicted in Figure 12.

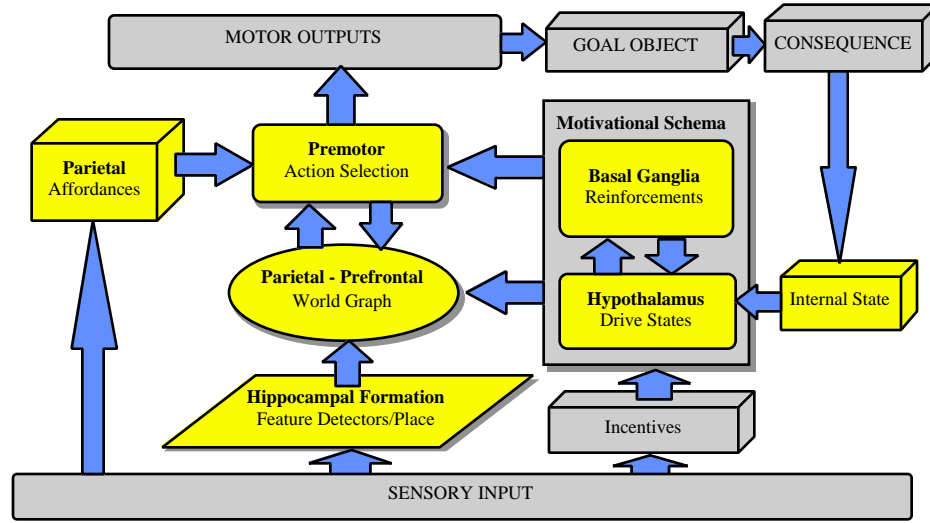


Figure 12: Integrated TAM-WG model of rat navigation.

In the integrated model, sensory inputs are used to compute the affordances (taxons), for world graph node selection and updating, and to trigger incentives. These inputs are then used to allow the animal to successfully solve and reverse the T-maze, as seen in O'Keefe's experiments (1983). Available map information ("where to go") is now available to be used by the "Action Selection" schema. In TAM-WG, the decision to turn to a certain angle (the orientation vector) is also given by a winner-take-all process performed over the integration of fields produced by the available affordances (AF), drive relevant stimuli (ST), rewardness expectation (RE), map information from the world graph (MI), and a curiosity level (CL) - to be described below. MI , as well as RE , contains an associated noise factor. In this paper, we are assuming that the expectation of future reward given by the WG model is noisier than the rewardness expectation given by TAM. The combined noise factor is equal to the sum of $noise_re$ and $noise_mi$ for the integrated model). If a node x is active in the world graph, there will be expectations associated to each edge coming from node x . Since associated to each edge is the direction it points to, a peak of activity with the magnitude of the expectation associated to that node/edge pair will be coarse coded over a linear array of neurons in the location respective to the edge's direction. In TAM-WG, then, the total input I_{in} to the action selection module becomes

$$I_{in}(i, t) = AF(i, t) + ST(i, t) + (RE(i, t) + noise_re) + (MI(i, t) + noise_mi) + CL(i, t)$$

Eventually, the selected motor outputs will enable the system to reach the goal object. When a selected action is executed, it will produce consequences in the animal's internal state and in the way it perceives the world (cf. the action-perception cycle - Arbib, 1972). The internal state will alter the drive levels, which will in turn influence the selection of the next node in the world graph, and so on.

A crucial aspect of the WG theory and of the TAM-WG model is that an animal may go to places that are not yet represented in its world graph. To see how this occurs, we must clarify what is meant by a node x' immediately adjacent to a node x . Certainly, if there is an edge from x to x' in the world graph, then x' is immediately adjacent to

x . However, the WG theory also considers as adjacent to x those world situations x' not yet represented in the graph, but that can be reached by taking some path from x . In the WG model, these situations will have a value attached to them, which can be seen as a “curiosity level” attached to the unknown. If, for example, an animal visited some arms of a maze and did not find anything interesting in them then, when faced with a new arm, it will prefer to visit it rather than the arms previously visited. The curiosity level is based on the value of the animal’s biggest drive at a particular given time. In the current model, if not motivated to go towards a previously experienced place, the simulated rat will tend to choose, based on its curiosity level, an affordance that leads to a place not yet represented in its world graph, i.e., from the current node x , there are no edges pointing in the direction signaled by the not yet selected affordance.

7 Experiments and Results

As mentioned before, the behavior of the control rats during reversal training in the experiments reported by O’Keefe (1983) was different from that obtained for fornix-lesioned animals. Although performance on the T-maze proceeded in roughly the same pattern as in the lesioned rats, the underlying orientation vector of the group did not shift in a smooth fashion but jumped around. Examination of the individual curves failed to reveal any systematic shifts (Figure 13a). O’Keefe (1983) concluded that the reversal performance of the controls was not based solely on a shift in their orientation system but also on the rat’s use of its hippocampal cognitive mapping system. He also assumed that this system can operate on the basis of the shape of the maze, since no landmarks were available during the task.

Figure 13b shows the performance of ten TAM-WG simulated rats. Like O’Keefe’s rats, our “control rats” did not show any systematic shifts in behavior. They randomly chose different arms to orient to when in the dais of the radial maze. However, by analyzing the simulation results, we saw that the orientation vector given by the rewardness expectation signal shifts from the left quadrant to the right quadrant, just as in the TAM simulated animals.

In the present situation, however, when the simulated rat is learning to go to the left arm, the rewardness expectation signal given by TAM and the expectation of future reinforcement given by the world graph will coincide. In this case, the simulated rat will learn to go to the left consistently and will eventually meet the criterion once the combined expectation is bigger than noise. In fact, the animal will tend not to go to the right arm once it is represented in the rat’s world graph, since the simulated rat already “knows” that food was not found at that particular arm (there are no expectations of future reinforcement associated with any node/edge pair leading to the right arm). In this case, the rat’s curiosity level for the right arm is decreased.

During reversal of the T-maze, the expectations of future reinforcement for the left arm will decrease continuously, since the absence of food is coded as a frustrating event by the reinforcement learning rule. When the combined expectation activity given by TAM and the world graph is smaller than the curiosity level, the TAM-WG simulated rat will persist in going to the left arm in the T-maze, but not in the radial arm maze. Since its curiosity level for the extra available arms is bigger than the one for the right and left arms, the rat will start to explore the arms not yet represented in its world graph. Once the combined expectation is smaller than noise, the rat will visit the right arm in the T-maze and will continuously create a peak of activity in the world graph and in TAM for the right arm. Since in the beginning the peak for the right arm is smaller than the combined curiosity and noise levels, the rat will still tend to choose arms randomly in the 8-arm maze. Although shifting steadily, the rewardness expectation activity in this case is too small to influence the behavior of the animal. The slow increase of activity in TAM can be explained due to a difference in the eligibility decay traces in the two systems. While the expectations in the world graph obey a fast-learning/fast-decay principle, expectations of future reinforcement in TAM obey a slow-learning/slow-decay mechanism.

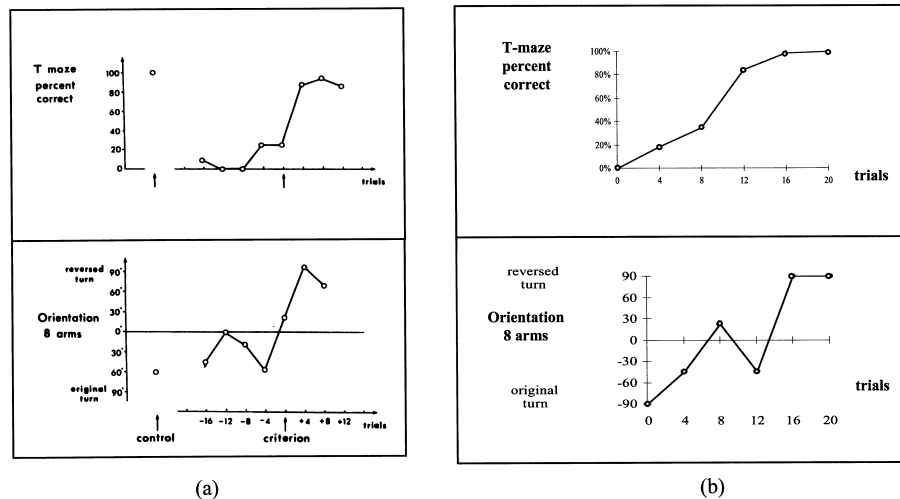


Figure 13: (a) The performance of four control rats during reversal on a T-maze. Each graph was obtained by averaging the graphs of the individual animals backwards and forwards from the beginning of the criterion run. Extracted from O'Keefe (1983). (b) The performance of ten TAM-WG "rats" during reversal on a T-maze. The bottom graph was obtained by computing the medians amongst animals and trials (four by four). The top graph was obtained by averaging the performance of each animal over periods of four trials.

The world graph constructed by a typical simulation of TAM-WG is depicted in Figure 14. This figure was taken after the simulated rat finished the reversal task. It is easy to see in the figure that the world graph formed for the T-maze is contained in the world graph formed for the 8-arm radial maze. The node in the center of the graph represents the T-junction and the dais of the radial arm maze. In fact, in the experiments conducted by O'Keefe (1983), the animals are always tested in the same 8-arm radial maze. However, for each T-trial five arms are pulled away from the center so that the remaining three arms form the T-maze (there is no reason to think the animals do not recognize the dais of the maze for the T and radial maze conditions as the same place). During the simulated experiments, the animals were always placed in the same starting position in the south arm for the T and the 8-arm radial maze. If a new node is eventually created to represent the dais of the radial arm maze in the world graph, this is merged to the node representing the T-maze junction. As previously described, if there is an edge from x to x' in the world graph, then x' is considered to be immediately adjacent to x . If, on leaving x , a new node x'' is created in the same direction of node x' , x'' and x' are merged to a single node. The world graph depicted in Figure 14 has also three nodes for each arm visited, with the exception of the "north arm", which does not have any nodes associated to it, since it is represented by a single node, the one the animal was released from in the south arm (the rat senses each of these two arms in the same direction - north). The three nodes for each of the other arms represent a conjunction of head direction and wall information. Nodes in the extremities are coding for the end of an arm and head directionality. Since in this case no landmarks are available, the head direction information is used to distinguish the several arms.

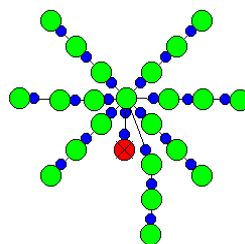


Figure 14: World graph formed by the WG model after completion of the reversal T-maze task. The shape of the graph reflects the topology of the 8-arm radial maze. In the figure, the world graph is represented as nodes connected by edges. Edges are directional and this is represented by a little circle attached to the nodes they point to. The currently active node is darker than the other nodes and has an X inside.

We now describe another test of our model, inspired by the Hirsh et al. (1978) study of a task in which animals were required to turn one way in a T-maze for food and the other way for water. Animals were alternately hungry and

thirsty. Sessions of three food and three water trials were run on alternate days until the animal was able to obtain food and water on not less than 11 out of 12 trials (for each food and water). Normal animals learned this task “easily”, going towards food when hungry and towards water when thirsty (Figure 17a). The median number of trials required for controls was 128. However, fornix-lesioned animals could not learn both tasks. They learned one task completely and then unlearned that task as they learned the other (Figure 16a): cross-overs required from 6 to 93 trials for completion (Hirsh et al., 1978). According to Hirsh et al., the observed switch in behavior is due to a tendency of performance in obtaining one stimulus to decline and in obtaining the other stimulus to improve, which indicates a lack of reward differentiation. In fact, a few of the fornix-lesioned animals eventually learned the task. In their case, the median number of trials was 222 trials (74 days).

In Hirsh et al.’s experiments, the intertrial interval within sessions was 20 minutes and the intersession interval was 23 hours. However, according to O’Keefe (1983) the effect of a turn on the orientation vector is time dependent, being large immediately after the event and decaying to an asymptote within a short time. For O’Keefe (1983), if the one minute interval used in his experiments is short compared with the asymptotic time constant of the orientation system, then the larger intervals, like the ones used by Hirsh et al., would produce much smaller increments and the lesioned rats would take much longer to shift their behavior in the radial maze, i.e., the longer the interval, the smaller the recall rate associated to it.

The model replicates Hirsh et al.’s data for fornix-lesioned (Figure 16b) and control animals (Figure 17b). Figures 16 and 17 show data for individual animals. In Hirsh et al.’s experiments and in our simulations, if the data for animals within groups are pooled, salient stylistic features are obscured because they occur at different times for different individuals. Because of the long intervals used in between trials and sessions, the model implements distinct recall rates in TAM and in the WG model (Figure 15). These are applied over the rewardness expectation array of cells in TAM and over the expectation of future reward in the WG model. Note that in Figure 15 we are assuming that the world graph (our extended locale system) has an initial small recall rate of its expectation of future reward and that its curve raises up fast when training trials proceed if compared to the curve used to model the recall rates for the TAM rewardness expectation signal. Hirsh et al. do not mention the existence of possible distal cues around the environment. However, they do report that the maze walls were 10 cm high and that the whole T-maze was covered by Plexiglas. This would make it difficult for the control animals to relate themselves to possible distal stationary cues sitting around the environment. In this case, the map construction would have to rely mainly on the head-direction information and wall information (no landmarks were used in our simulated experiments). This fact, together with longer intersession intervals would diminish the amount of recall for the mapping system. We hypothesize that the WG intersession recall rate is dependent not only on the time interval between sessions but also on the availability of reliable landmarks the animal could relate to when building its cognitive map.

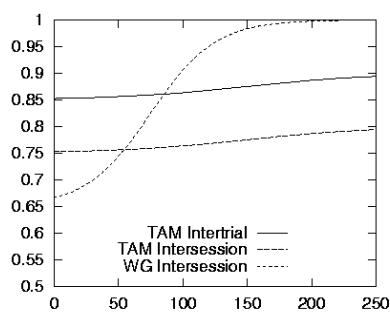


Figure 15: Recall rates applied over TAM and the WG model during the simulation of Hirsh et al.’s task. The x-axis represents the number of trials. Simulations were executed up to 250 trials. The TAM intertrial and the TAM intersession rates represent the recall rates applied in TAM over the rewardness expectation array of cells for each intertrial and intersession interval, respectively. The WG intersession rate represents the recall rate applied over the expectation of future reward in the WG model. The intertrial recall rate for the WG model was considered to be 1 at all times.

For the TAM simulated animals (Figure 16b), when the rat discovers food or water on a determined arm, it will be biased to turn towards that arm in the next trial. For example, if the simulated rat finds food on the left arm in the first trial of a session, it will be biased to turn left in the next two trials (each session is composed of three trials). In the fourth trial, water is available on the right arm and not in the left arm. In this way, if turning left again, the

animal will get frustrated and its confidence to turn left next time will be lower. The model assumes that fornix-lesioned animals are indifferent to the kind of reward experienced, i.e., in TAM water and food are coded as reinforcers devoid of contextual information. In this way, the finding of water when thirsty and of food when hungry will determine together in TAM the center of mass of the rewardness expectation peak of activity. In Figure 16b, the intersession and intertrial intervals recall rates will keep the bias to turn left or right below noise and, for this reason, the rat will eventually turn in a different direction than the one to which he has been previously biased. The simulated animal will then have the chance to build a bump of activation for the other arm, which according to our example means turning right for water. By doing this, there will be a shift in the centroid of the rewardness expectation peak of activity towards the right side. When the peak swings from the left to the right quadrant, it will influence less and less the turning left behavior and will eventually start to influence more and more the turning right behavior. However, there will be a time the rewardness expectation signal will provide an influence neither right nor left, since it may be in between (in this case, if an affordance to go north existed, as in the case of the eight-arm radial maze, this would be selected, just as in the O'Keefe experiments). In this situation, the selected turn will be dependent solely on noise, what would explain the big cross-over variability observed by Hirsh et al. In this way, the TAM-simulated animals, like the fornix-lesioned animals, will be constantly switching behaviors from time to time, learning one task completely and then unlearning that task as they learn the other.

However, as mentioned before, some of the fornix-lesioned animals eventually learned the task, while TAM simulated animals never do. In a sense, fornix transection is not the same as hippocampal-lesion: a fornix lesion severs the cholinergic input to the hippocampus, the connections from the subiculum to the nucleus accumbens, and the postsubicular connections to the lateral mammillary nuclei, but it does not sever the hippocampal/entorhinal connections nor all the connections between hypothalamic nuclei and cortical areas. We hypothesize that with the extended training procedure (~3 months), Hirsh et al's fornix-lesioned animals were able to change strategy, maybe by extracting information from the hippocampal formation via cortical connections, and were then able to use the expectation of future reward given by their "world graph" to eventually learn the task.

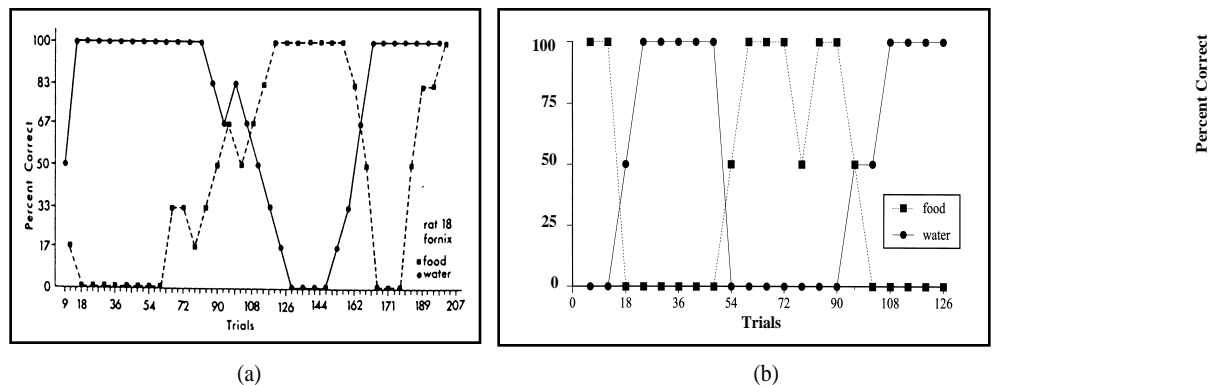


Figure 16: (a) The performance of fornix-lesioned rat 18 during adjoining hunger and thirst sessions plotted as a function of trials for the individual animal requiring the median number of trials to reach criterion. The first point from the ordinate represents performance during a block composed of trials 1-3 and 7-9. A sliding scale of trials was used. The other curve is plotted in an analogous manner, except that the first point on it represents trials 4-6 and 10-12. Extracted from Hirsh et al. (1978). (b) The performance of a TAM simulated animal during adjoining hunger and thirst sessions plotted as a function of trials. The amount of trials shown is equally divided in between food and water sessions (a sliding scale of trials was not used). Each point represents the median of medians computed over two consecutive sessions, i.e., the first point for food represents food trials 1-3 and 7-9, while the first point for water represents water trials 1-3 and 7-9.

In Figure 17b, due to the lack of reward differentiation in TAM, control animals will be biased towards one response or the other. The strong influence of the rewardness expectation signal from the taxon system, at least in the beginning, is seen partly because of the small recall rate associated to the WG expectation of future reward. This explains why in the beginning one of the behavioral conditions dominates, while the other tends to remain in the chance level. As the number of trials increase, the WG signal gets stronger and stronger (Figure 15) and will eventually dominate the behavior selection process. Since in the World Graph the expectation of future reward is context dependent, food will not in this case interfere with water or vice-versa, and the rat will eventually learn, for example, to turn left for food when hungry and to turn right for water when thirsty. By the time the animal starts turning an even amount of times left and right, the peak of activity of the TAM rewardness expectation signal will

tend to remain in the middle of its correspondent array of cells, and will influence neither the turning left nor the turning right behavior.

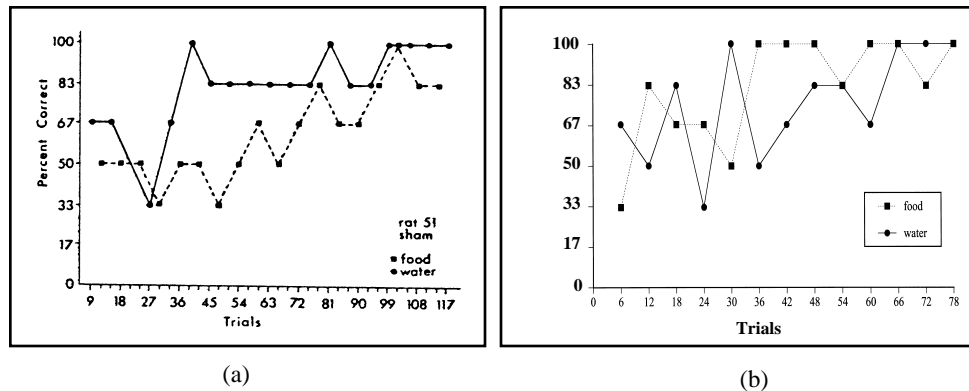


Figure 17: (a) The performance of control rat 51 during adjoining hunger and thirst sessions plotted as a function of trials for the individual animal requiring the median number of trials to reach criterion. The first point from the ordinate represents performance during a block composed of trials 1-3 and 7-9. A sliding scale of trials was used. The other curve is plotted in an analogous manner, except that the first point on it represents trials 4-6 and 10-12. Food trials can be interpreted as hunger trials when the animal is looking for food, while water trials can be interpreted as thirst trials when the animal is looking for water. Extracted from Hirsh et al. (1978). (b) The performance of a TAM-WG simulated animal during adjoining hunger and thirst sessions plotted as a function of trials. The set of trials shown is equally divided between food and water sessions (a sliding scale of trials was not used). Each point represents the median of averages computed over two consecutive sessions, e.g., the first point for food represents food trials 1-3 and 7-9, while the first point for water represents water trials 1-3 and 7-9.

8 Conclusions

TAM-WG provides a model that can be used to answer how behavioral and spatial representations are compiled and how they are selected for use. Moreover, TAM-WG can be used to explain much spatial behavior without requiring an a priori system for organizing sensations into an absolute spatial framework, as postulated by O'Keefe and Nadel (1978). For this reason, the model can be used as a modeling environment for the investigation of motivated spatial behavior. It can also be viewed as a tool for development of biologically testable models of the cooperative computation of multiple brain regions as the animal explores its world in a goal-driven way.

Acknowledgments

This work was partially supported by CAPES Research Foundation - Brazil, and by the Human Brain Project (with funding from NIMH, NASA, and NIDA) under a P20 Program Project Grant HBP: 5-P20-52194 for work on "Neural Plasticity: Data and Computational Structures" (M. A. Arbib, Director).

References

- Apicella, P., E. Scarnati, T. Ljungberg and W. Schultz (1992). Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *Journal of neurophysiology* 68(3): 945-960.
- Arbib, M. A. (1972). *The Metaphorical Brain: Cybernetics as Artificial Intelligence and Brain Theory*. New York, Wiley-Interscience.
- Arbib, M. A. (1997). From Visual Affordances in Monkey Parietal Cortex to Hippocampo-Parietal Interactions Underlying Rat Navigation. *Phil. Trans. Roy. Soc. (London) B* (in press).
- Arbib, M. A., P. Érdi and J. Szentágothai (1997). *Neural Organization: Structure, Function, and Dynamics*. Cambridge, MA, A Bradford Book/The MIT Press.
- Arbib, M. A. and D. H. House (1987). Depth and detours: An essay on visually guided behavior. In M. A. Arbib and A. R. Hanson (Eds.), *Vision, brain and cooperative computation*. Cambridge, MA, Bradford Book/MIT Press.
- Arbib, M. A. and I. Lieblich (1977). Motivational Learning of Spatial Behavior. In J. Metzler (Eds.), *Systems Neuroscience*. New York, Academic Press. 221-39.
- Barto, A. G. (1994). Adaptive critics and the basal ganglia. In J. C. Houket al (Eds.), *Models of Information Processing in the Basal Ganglia*. Cambridge, MA, MIT Press. 215-232.

Barto, A. G., S. Richard and C. W. Anderson (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Trans. on Systems, Man, and Cybernetics* SMC-13: 834-846.

Barto, A. G. and R. S. Sutton (1981). Landmark learning: an illustration of associative search. *Biological Cybernetics* 42: 1-8.

Chen, L. L., L. H. Lin, C. A. Barnes and B. L. McNaughton (1994b). Head-Direction Cells In the Rat Posterior Cortex .2. Contributions Of Visual and Ideothetic Information to the Directional Firing. *Experimental Brain Research* 101(1): 24-34.

Chen, L. L., L. H. Lin, E. J. Green, C. A. Barnes and B. L. McNaughton (1994a). Head-Direction Cells In the Rat Posterior Cortex .1. Anatomical Distribution and Behavioral Modulation. *Experimental Brain Research* 101(1): 8-23.

Collett, T. (1982). Do toads plan routes? A study of detour behavior of *B. viridis*. *Journal of Comparative Physiology [A]* 146: 261-271.

Corbacho, F. J. and M. A. Arbib (1995). Learning to detour. *Adaptive Behavior* 3(4): 419-468.

Dayan, P. (1990). Navigating through temporal difference. In R. P. Lippmann et al (Eds.), *Neural Information Processing Systems 3*. San Mateo, Morgan Kaufmann. 464-470.

Dominey, P. F. and M. A. Arbib (1992). A cortico-subcortical model for generation of spatially accurate sequential saccades. *Cerebral Cortex* 2: 135-175.

Dorman, C. and P. Gaudiano (1995). Motivation. In M. A. Arbib (Eds.), *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA, The MIT Press. 591-594.

Etienne, A. S., R. Maurer and V. Seguinot (1996). Path integration in mammals and its interaction with visual landmarks. *J Exp Biol* 199(Pt 1): 201-9.

Fagg, A. H. (1996) *A computational model of the cortical mechanisms involved in primate grasping*. Ph. D. Thesis, University of Southern California.

Fagg, A. H. and M. A. Arbib (1998). Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks* (to appear).

Gibson, J. J. (1966). *The Senses Considered as Perceptual Systems*. Allen and Unwin.

Goodridge, J. P. and J. S. Taube (1995). Preferential use of the landmark navigational system by head direction cells in rats. *Behav Neurosci* 109(1): 49-61.

Hirsh, R., B. Leber and K. Gillman (1978). Fornix fibers and motivational states as controllers of behavior: A study stimulated by the contextual retrieval theory. *Behavioral Biology* 22:463-478.

Hull, C. L. (1943). *Principles of behavior*. New York, Appleton-Century-Crofts.

Ingle, D. (1980). The frog's detection of stationary objects following lesions of the pretectum. *Behavioral Brain Research* 3: 151-173.

Klopf, A. H. (1972). Brain function and adaptive systems - A heterostatic theory. Air Force Cambridge Res. Lab. Res. Rep., AFCRL-72-0164, Bedford, MA. (A summary appears in *Proc. Int. Conf. Syst., Man, Cybern.*, 1974)

Knierim, J. J., H. S. Kudrimoti and B. L. McNaughton (1995). Place cells, head direction cells, and the learning of landmark stability. *Journal of Neuroscience* 15(3 Pt 1): 1648-59.

Knierim, J. J., B. L. McNaughton, C. Duffield and J. Bliss (1993). On the binding of hippocampal place fields to the inertial orientation system. *Soc. Neurosci. Abst.* 19: 795.

Kolb, B., K. Buhrmann, R. McDonald and R. J. Sutherland (1994). Dissociation of the medial prefrontal, posterior parietal, and posterior temporal cortex for spatial navigation and recognition memory in the rat. *Cerebral Cortex* 4(6): 664-80.

Lieblich, I. and M. A. Arbib (1982). Multiple representations of space underlying behavior. *The Behavioral and Brain Sciences* 5: 627-659.

Markus, E. J., C. A. Barnes, B. L. McNaughton, V. L. Gladden and W. E. Skaggs (1994). Spatial information content and reliability of hippocampal CA1 neurons: effects of visual input. *Hippocampus* 4(4): 410-21.

McNaughton, B. L., C. A. Barnes, J. L. Gerrard, K. Gothard, M. W. Jung, J. J. Knierim, H. Kudrimoti, Y. Qin, W. E. Skaggs, M. Suster and K. L. Weaver (1996). Deciphering the hippocampal polyglot: the hippocampus as a path integration system. *J. Exp. Biol.* 199(Pt 1): 173-85.

McNaughton, B. L., C. A. Barnes and J. O'Keefe (1983). The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Experimental Brain Research* 52(1): 41-9.

McNaughton, B. L., J. J. Knierim and M. A. Wilson (1994b). Vector encoding and the vestibular foundations of spatial cognition: Neurophysiological and computational mechanisms. In M. Gazzaniga (Eds.), *The Cognitive Neurosciences*. Boston, MIT Press. 585-595.

McNaughton, B. L., B. Leonard and L. Chen (1989). Cortico-hippocampal interactions and cognitive mapping: A hypothesis based on reintegration of the parietal and inferotemporal pathways for visual processing. *Psychobiology* 17: 230-235.

McNaughton, B. L., S. J. Mizumori, C. A. Barnes, B. J. Leonard, M. Marquis and E. J. Green (1994a). Cortical representation of motion during unrestrained spatial navigation in the rat. *Cerebral Cortex* 4(1): 27-39.

Mirenowicz, J. and W. Schultz (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* 379: 449-451.

Mishkin, M., L. G. Ungerleider and K. A. Mack (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neuroscience* 6: 414-417.

Mittelstaedt, M. L. and H. Mittelstaedt (1980). Homing by path integration in a mammal. *Naturwissenschaften* 67: 566-567.

Morris, R. G., P. Garrud, J. N. Rawlins and J. O'Keefe (1982). Place navigation impaired in rats with hippocampal lesions. *Nature* 297(5868): 681-3.

Morris, R. G. M., J. J. Hagan and J. N. P. Rawlins (1986). Allocentric spatial learning by hippocampectomised rats: a further test of "spatial mapping" and "working memory" theories of hippocampal function. *Quarterly Journal of Experimental Psychology* 38B: 365-395.

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

O'Keefe, J. and J. Dostrovsky (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. *Brain Research* 34: 171-175.

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

Olton, D. S., J. T. Becker and G. E. Handelmann (1980). Hippocampal function: working memory or cognitive mapping? *Physiological Psychology* 8: 239-246.

Packard, M. G. and J. L. McGaugh (1992). Double dissociation of fornix and caudate nucleus lesions on acquisition of two water maze tasks: further evidence for multiple memory systems. *Behavioral Neuroscience* 106(3): 439-446.

Poucet, B. (1993). Spatial cognitive maps in animals: new hypotheses on their structure and neural mechanisms. *Psychological Review* 100(2): 163-82.

Quirk, G. J., R. U. Muller and J. L. Kubie (1990). The firing of hippocampal place cells in the dark depends on the rat's recent experience. *Journal of Neuroscience* 10(6): 2008-17.

Risold, P. Y. and L. W. Swanson (1995). Evidence for a hypothalamo-thalamocortical circuit mediating pheromonal influences on eye and head movements. *Proc. Natl. Acad. Sci. U S A* 92(9): 3898-902.

Rizzolatti, G., R. Camarda, L. Fogassi, M. Gentilucci, G. Luppino and M. Matelli (1988). Functional Organization of Inferior Area 6 in the Macaque Monkey II. Area F5 and the Control of Distal Movements. *Experiment Brain Research* 71: 491-507.

Sakata, H., H. Shibutani, Y. Ito, K. Tsurugai, S. Mine and M. Kusunoki (1994). Functional properties of rotation-sensitive neurons in the posterior parietal association cortex of the monkey. *Experimental Brain Research* 101(2): 183-202.

Samsonovich, A. (1996) *Attractor Map Theory of the Hippocampal Representation of Space*. Ph. D. Thesis, University of Arizona.

Samsonovich, A. and B. L. McNaughton (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience* 17(15): 5900-5920.

Save, E., M. C. Buhot, N. Foreman and C. Thinus-Blanc (1992). Exploratory Activity and Response to a Spatial Change In Rats With Hippocampal or Posterior Parietal Cortical-Lesions. *Behavioural Brain Research* 47(2): 113-127.

Save, E. and M. Moghaddam (1996). Effects Of Lesions Of the Associative Parietal Cortex On the Acquisition and Use Of Spatial Memory In Egocentric and Allocentric Navigation Tasks In the Rat. *Behavioral Neuroscience* 110(1): 74-85.

Schenk, F. and R. G. M. Morris (1985). Dissociation between components of spatial memory in rats after recovery from the effects of retrohippocampal lesions. *Experimental Brain Research* 58: 11-28.

Schmajuk, N. A. (1990). Role of the hippocampus in temporal and spatial navigation: an adaptive neural network. *Behavioral Brain Research* 39(3): 205-29.

Schmajuk, N. A. and H. T. Blair (1993). Place learning and the dynamics of spatial navigation: A neural network approach. *Adaptive Behavior* 1: 355-87.

Schmajuk, N. A. and A. D. Thieme (1992). Purposive behavior and cognitive mapping: a neural network model. *Biol. Cybernetics* 67: 165-174.

Schultz, W., R. Romo, T. Ljungberg, J. Mirenowicz, J. R. Hollerman and A. Dickinson (1995). Reward-related signals carried by dopamine neurons. In J. R. Houket al (Eds.), *Models of information processing in the basal ganglia*. Cambridge, MA, The MIT Press. 233-248.

Shibata, H. and A. Kato (1993). Topographic relationship between anteromedial thalamic nucleus neurons and their cortical terminal fields in the rat. *Neuroscience Research* 17(1): 63-9.

Sutherland, R. J. and A. J. Rodriguez (1989). The role of the fornix/fimbria and some related subcortical structures in place learning and memory. *Behav. Brain Res.* 32(3): 265-277.

Sutton, S. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning* 3: 9-44.

Swanson, L. W. and G. J. Mogenson (1981). Neural mechanisms for the functional coupling of autonomic, endocrine and somatomotor responses in adaptive behavior. *Brain Research Reviews* 3: 1-34.

Taube, J. S., R. U. Muller and J. B. Ranck, Jr. (1990a). Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *Journal of Neuroscience* 10(2): 436-47.

Taube, J. S., R. U. Muller and J. B. J. Ranck (1990b). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. *J. Neuroscience* 10: 420-435.

Taube, J. S. J. (1995). Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *Journal of Neuroscience* 15(1 Pt 1): 70-86.

Thinus-Blanc, C. (1996). *Animal Spatial Cognition: Behavioral and Brain Approach*. Singapore, World Scientific Publishing.

Toates, F. (1986). *Motivational Systems*. Cambridge University Press.

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

Touretzky, D. S. and A. D. Redish (1996). Theory of Rodent Navigation Based on Interacting Representations of Space. *Hippocampus* 6: 247-270.

Wiener, S. I. (1993). Spatial and behavioral correlates of striatal neurons in rats performing a self-initiated navigation task. *Journal of Neuroscience* 13(9): 3802-3817.