A Model of Prefrontal Cortical Mechanisms for Goal-directed Behavior

Michael E. Hasselmo

Abstract

■ Many behavioral tasks require goal-directed actions to obtain delayed reward. The prefrontal cortex appears to mediate many aspects of goal-directed function. This article presents a model of the prefrontal cortex function emphasizing the influence of goal-related activity on the choice of the next motor output. The model can be interpreted in terms of key elements of the Reinforcement Learning Theory. Different neocortical minicolumns represent distinct sensory input states and distinct motor output actions. The dynamics of each minicolumn include separate phases of encoding and retrieval. During encoding, strengthening of excitatory connections forms forward and reverse associations between each state, the following action, and a subsequent state, which may include reward. During retrieval, activity spreads from reward states throughout the network. The interaction of this spreading activity with a specific input state directs selection of the next appropriate action. Simulations demonstrate how these mechanisms can guide performance in a range of goal-directed tasks, and provide a functional framework for some of the neuronal responses previously observed in the medial prefrontal cortex during performance of spatial memory tasks in rats. ■

INTRODUCTION

Numerous behavioral tasks involve goal-directed behavior based upon a delayed reward. For example, a rat in an instrumental task must generate lever presses to obtain food reward (Wyble, Hyman, Rossi, & Hasselmo, 2004; Corbit & Balleine, 2003; Killcross & Coutureau, 2003), and a rat in a T-maze must run down the stem of the maze to obtain food reward in one arm of the maze (Baeg et al., 2003; Ferbinteanu & Shapiro, 2003; Wood, Dudchenko, Robitsek, & Eichenbaum, 2000; Jung, Qin, McNaughtonm, & Barnesm, 1998). Lesions of the prefrontal cortex cause impairments in goal-directed behavior (Corbit & Balleine, 2003; Killcross & Coutureau, 2003; Miller & Cohen, 2001; Fuster, 1995), and prefrontal units show firing dependent upon the association of cues and future responses (Miller, 2000). The model presented here addresses how goal-directed behavior can be mediated by populations of neurons.

An extensive theoretical framework termed Reinforcement Learning (RL; Sutton & Barto, 1998; Sutton, 1988) describes how an agent can generate behaviors for delayed rewards in its environment. Current sensory input to the agent is represented by a ''state'' vector, and the output of the agent is represented by ''actions'' which alter the state vector (i.e., moving the agent to a different location). The selection of actions is guided by value functions (associating states with future reward) and state–action value functions (associating actions in specific states with future reward). These functions are often learned using variants of temporal difference (TD) learning (Sutton & Barto, 1998; Sutton, 1988).

Research has focused on the similarity between the error term in TD learning and the activity of dopaminergic neurons (Schultz, Dayanm, & Montague, 1997; Montague, Dayan, & Sejnowski, 1996). The basal ganglia have been proposed to provide circuitry for computation of TD learning (Houk, Adams, & Barto, 1995). Alternatives to TD learning have also been developed in models of the basal ganglia (Brown, Bullock, & Grossberg, 1999). Despite these links to biology, the mechanisms for many other aspects of RL have not been analyzed. Most RL models use simple look-up tables for the action-value function, without mapping these functions to the physiological properties of neurons. The state–action value mapping has been modeled with neural networks (Zhu & Hammerstrom, 2003; Barto & Sutton, 1981), but these hybrid models retain many algorithmic steps which are not implemented biologically.

In contrast, this article focuses on obtaining goal-directed behavior using a neurobiological circuit model with all functions implemented by threshold units and modifiable synaptic connections. This model demonstrates how action selection could be computed by activity in prefrontal cortical circuits. The model does not focus on dopaminergic activity and does not explicitly use the TD learning rule. Instead, this model obtains effective action selection using interacting neurons, and

Boston University

demonstrates how specific circuit dynamics with local Hebbian rules for synaptic modification can provide functions similar to TD learning. The activity of individual neurons in the simulation is described in relationship to experimental data on the prefrontal cortex unit firing in two different tasks: an open-field task and a spatial alternation task in a T-maze (Baeg et al., 2003; Wood et al., 2000; Jung et al., 1998). This model demonstrates how the activity of prefrontal cortical units can be interpreted as elements of a functional circuit which guide the actions of an agent on the basis of delayed reward.

RESULTS

Overview of Network Function

The model presented here contains a repeated subcircuit (Figure 1) intended to represent a repeating functional unit of neocortical architecture, such as the minicolumn (Rao, Williams, & Goldman-Rakic, 1999). Each local minicolumn includes a population of n input units, designated with the letter a, which receives input about the current state or the most recent action. Across the full model these units provide input to n minicolumns, forming a larger vector a (with size $n * n$). The vector a represents units in layer IV of cortical structures, which receive input from thalamic nuclei conveying information about sensory stimuli or proprioceptive feedback about an action, and also receive feedforward connections from cortical areas lower in the sensory hierarchy (Scannell, Blakemore, & Young, 1995; Felleman & Van Essen, 1991; Barbas & Pandya, 1989). The representations in this model are consistent with data showing that the prefrontal cortex neurons respond to a range of behaviorally relevant sensory stimuli, motor outputs, and reward (Koene & Hasselmo, in press; Mulder, Nordquist, Orgut, & Pennartz, 2003; Wallis, Anderson, & Miller, 2001; Schultz, Tremblay, & Hollerman, 2000; Jung et al., 1998; Schoenbaum, Chiba, & Gallagher, 1998; Schoenbaum & Eichenbaum, 1995).

Each minicolumn contains four populations of units that mediate associations with other minicolumns activated at different time points (see Figure 1). The reverse spread of activity from the goal (reward) minicolumn is mediated by connections W_g, and forward associations from current input are mediated by W_c and W_o. Populations g_i and g_o in each minicolumn are analogous to neurons in layers II and III (supragranular layers) in the neocortex, which have long-range excitatory connections (Lewis, Melchitzky, & Burgos, 2002). Population g_i receives input spreading from the goal via connections W_g. Population g_o receives input from g_i via internal connections W_{ig} and sends output to other minicolumns via W_g. These connections link each action with preceding states, and each state with preceding actions.

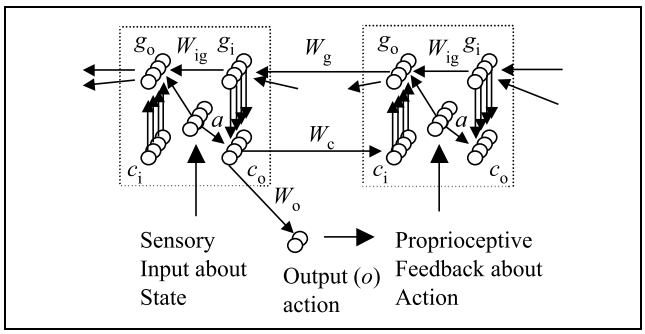

Figure 1. Components of each model minicolumn. Each minicolumn has multiple units in vector a receiving input from the thalamus about current sensory state or proprioceptive feedback about motor output (actions). The population g_i receives input spreading back from the goal via connections W_g from other minicolumns. The population g_o receives activity spreading from the goal via internal connections W_{ig} and sends output to other minicolumns. The population c_i receives forward spread from the current state or action. Population c_o is activated by current input from population a in the same minicolumn, which converges with input from population g_i. Population c_o projects via connections W_o to output (action) units o to generate the next action on the basis of activity in population c_o.

Populations c_o and c_i in each minicolumn are analogous to neurons in layers V and VI (infragranular layer) in the neocortex. These neurons have more localized connections and influence the cortical output to subcortical structures and lower levels of the neocortex, consistent with the role of population c_o in regulating the output of the circuits in this model. Population c_i receives input about the current state or action from other minicolumns, whereas population c_o receives input from population a in the same minicolumn, and sends output to other minicolumns and to the output vector o via connections W_o.

Each minicolumn receives inputs consisting of either sensory information from the environment (described with the term ''state'' in RL) or proprioceptive feedback about specific motor actions performed (described with the term ''action'' in RL). As shown in Figure 2, the network generates outputs which guide behavior of a virtual rat. During retrieval, the spread of activity within the network guides the selection of the next action of the virtual rat. Each output action causes input to specific minicolumns representing actions. The encoding of associations between actions and the preceding and following states occurs during an encoding phase which is distinct from the retrieval phase which guides action selection. These separate phases could correspond to phases of oscillatory dynamics within cortical structures (Manns, Alonso, & Jones, 2000).

Example: Movement on a Linear Track

As an example, Figure 2 shows a model guiding movements of a rat on a linear track with reward provided

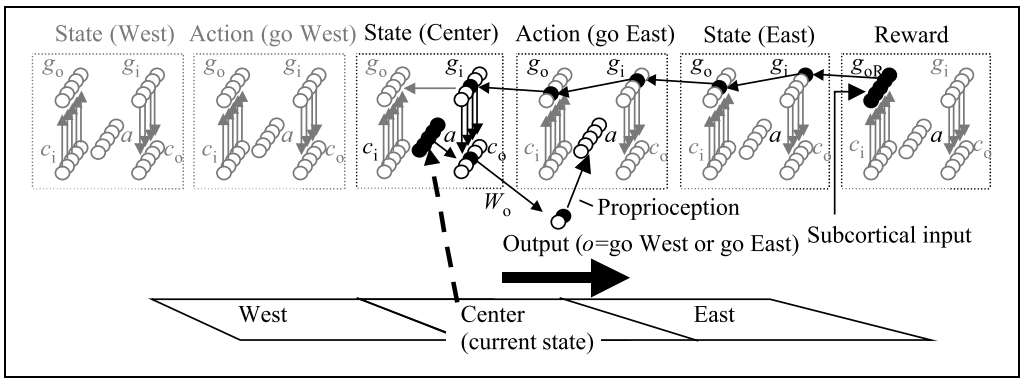

Figure 2. Retrieval activity in an example linear track task. Separate minicolumns represent the two actions ''go West'' and ''go East,'' and the three states ''West,'' ''Center,'' and ''East.'' One minicolumn represents goal (reward) in ''East.'' The selection of output action depends on convergence of reverse spread from goal with sensory input to the minicolumn representing the current state (Center). The desire for goal is represented by subcortical activation of the population g_{oR} in the reward minicolumn, which causes activity to spread back through the network (active units are black). Activity spreads to population g_i and then g_o in the East state minicolumn, then spreads through g_i and g_o in the ''go East'' action minicolumn before reaching the g_i population of the ''Center'' state minicolumn. In the center minicolumn, the subthreshold input from population g_i converges in population c_o with subthreshold input from population a. The activation of one unit in population c_o spreads over weights W_o to generate the appropriate action via the output vector o (which has two elements representing ''go East'' and ''go West'' actions.) Here, the activity of the output vector guides the rat to go East, and also returns as proprioceptive feedback to activate components of the a vector in the minicolumn for ''go East.''

consistently at one location (the ''East'' location). This resembles the RL example of an agent in a gridworld environment (Sutton & Barto, 1998), and resembles tasks used for studying neuronal responses in the hippocampus (Wyble et al., 2004; Gothard, Skaggs, & McNaughton, 1996). Here we use an allocentric representation of the state, but this framework can also be applied to egocentric representations.

This simple model consists of six minicolumns: three representing states (locations), two representing actions, and one representing reward. The ''states'' are labeled West, Center, and East in Figure 2, and provide input to three separate minicolumns. Each current ''state'' is represented by active elements in a. Here, the virtual rat has the option of two ''actions'' defined allocentrically, the actions ''go West'' and ''go East.'' Actions are generated by an output population, a two-element vector o which guides the movements of the rat (see Figure 2). Proprioceptive feedback about the active output in vector o activates elements of vector a in the corresponding action minicolumn representing ''go West'' or ''go East.'' The network also has a ''reward'' (goal) representation of the sensory input about food that becomes associated with physiological drive states such as hunger. The reward minicolumn is activated during encoding when food reward is first obtained, and provides the goal for selecting actions during retrieval. This example focuses on the retrieval process after encoding of the environment has been performed. The Methods section provides detailed equations for both the encoding and retrieval phases.

Retrieval Provides Action Selection

The following mechanism provides action selection when the rat is at the location in the center of the environment. The goal state is activated by subcortical drive mechanisms, represented in the model by diffuse activation of the population g_{oR} in the reward minicolumn (filled circles in Figure 2). In Figure 2, activity spreads over connections W_g from the g_{oR} population in the ''Reward'' minicolumn to the input population g_i in the ''East'' state minicolumn. These connections were strengthened during previous exploration of the environment (as described in the Methods section below), allowing units in g_o to activate a unit in g_i. The activity spreads over internal connections W_{ig} from population g_i to population g_o in the ''East'' state minicolumn. The spread continues over W_g from g_o in the ''East'' minicolumn to g_i in the ''go East'' action minicolumn, then over W_{ig} to g_o in the ''go East'' action minicolumn and from there over W_g to g_i in the ''Center'' state minicolumn. This continuous spread of activity traces possible reverse pathways from the goal back through sequences of states and actions leading to that goal.

The selection of an action depends upon the interaction of the spread from goal/reward with the input representing current state. The reverse spread from reward converges with input to the ''Center'' state minicolumn. Sensory input from the environment about the current state activates the units of a in the minicolumn representing the ''Center'' state, which send diffuse *subthreshold* activity to population c_o in that minicolumn. Activity in population c_o depends upon the convergence of this subthreshold input with subthreshold input from the unit in g_i which was activated by the reverse spread from reward. In Figure 2, this convergent input causes activity in unit 3 of population c_o in the ''Center'' state minicolumn, corresponding to the appropriate output ''go East.'' Previously strengthened connections W_o between this element of the

population c_o and the output population causes activity in the ''go East'' output unit, as shown in Figure 2. The activity of the output unit causes the virtual rat to move to the goal in the ''East'' location. Thus, the retrieval process performs the correct action selection for approaching the goal.

Separate input and output populations for reverse spread are required due to repeated use of actions in multiple contexts. The same action could result in different outcomes dependent upon the starting state. For example, a ''go East'' action could shift the state from West to Center, but also from Center to East. If there were only one population for both input and output, the network would map all inputs to every output. But with distinct populations of input and output populations, it is possible to make these mappings distinct. Minicolumn structure was chosen to be the same for both states and actions, just as the structure of the neocortex appears similar throughout the prefrontal cortex, where units respond to both sensory input and motor output (Mulder et al., 2003; Fuster, 1995).

Encoding Creates Necessary Synaptic Connectivity

The retrieval function described above depends upon prior modification of the appropriate pattern of connectivity in the synapses of the network. The process of encoding is summarized in Figures 8 and 9 and described in detail in the Methods section. The buffering of sensory input and timing of activity spread within the network allows encoding to occur with the time course of spike-timing-dependent synaptic plasticity (Markram, Lubke, Frotscher, & Sakmann, 1997; Levy & Steward, 1983), which requires postsynaptic spikes to occur immediately after presynaptic spikes. Encoding and retrieval phases alternate continuously in the model during all stages of behavior. Retrieval activity does not occur during encoding because there is no subcortical input to population g_o in the model and therefore no reverse spread. Modification of synapses occurs selectively on the encoding phase, based on data on phasic changes in LTP induction during theta rhythm (Hyman, Wyble, Goyal, Rossi, & Hasselmo, 2003). The effective learning of behavior results from an interaction of synaptic modification and the backward spread from goal, resulting in a function similar to that of TD learning.

The network starts with weak connectivity which does not generate learned actions. Outputs are initially generated randomly to move the animal from its prior location to a new location (state). Therefore, the initial encoding of the environment occurs as the virtual rat explores randomly, generating random sequences with a state followed by an action, which leads to another state. As the encoding process strengthens synaptic connections, the network begins to perform effective goal-directed behavior, as summarized in simulation results below using MATLAB.

Simulation Results

Goal Finding on a Linear Track

The guidance of goal-directed behavior by the prefrontal cortex circuit model was tested in a range of different behavioral tasks. The first task utilized a linear track, with reward located at one end. The virtual rat starts at the West end of the track, searches until it finds the reward at the East end, and is immediately reset to the West end. Figure 3 shows the movements of the virtual rat as it learns optimal performance in the linear track. The states (locations) of the rat over time are plotted as black rectangles in the top four rows of the plot. During the initial time steps of the simulation, the virtual rat explores back and forth randomly along the linear track (both West and East movements), and obtains infrequent rewards. As connections are modified within the network, the virtual rat gradually learns to run directly from the starting position to the goal location on each trial, thereby obtaining frequent rewards as shown on the right side of Figure 3. This gradual increase in goal-directed behavior results from the increase in reverse spread from the goal location as the rat learns the task and excitatory reverse connections are strengthened. The spread of activity across these reverse connections allows consistent selection of the correct response which guides the virtual rat to the goal location.

The simulations demonstrate that the encoding equations described in the Methods section allow formation of the necessary pattern of connectivity to encode potential pathways through the environment. The con-

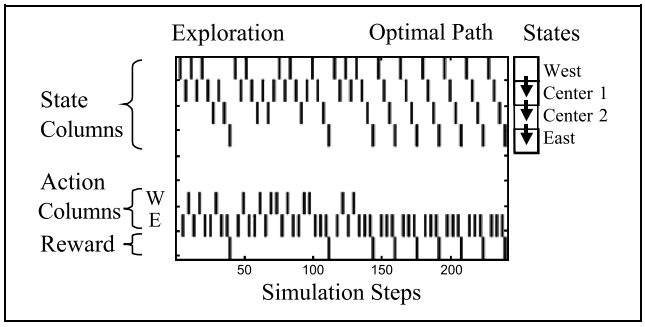

Figure 3. Movements of the virtual rat during learning and performance of the simple linear track task. Black vertical bars in the top four rows show the states (locations West, Center 1, Center 2, East) of the virtual rat during different simulation time steps plotted horizontally. Black rectangles in rows 7 and 8 show the actions go West (W) and go East (E). Black rectangles in the bottom row (row 9) show delivery of reward. Initially, the virtual rat moves randomly (Exploration), but as connections W_g, W_{ig}, and W_o are progressively strengthened, the activity becomes guided by the goal location (Optimal path). This results in generation of only ''go East'' movements (row 8) and greater regularity of reward delivery (shown by black rectangles in row 9).

vergence of sensory state input with the reverse spread of activity allows selection of actions which result in movement along optimal pathways within the environment in most cases (although in some cases the network settles into nonoptimal pathways). The effective goal-directed behavior can be seen in Figure 3, where the virtual rat learns to make Eastward movements only, thereby rapidly moving from the start location to the goal location and obtaining reward. The encoding process occurs during random exploration of the environment, so that the network does not have to be selectively structured for each environment, but can learn goal-directed behavior in a range of different environments.

Goal Finding in an Open Field

The model can guide movement of the virtual rat in environments of arbitrary shape and size, with different goal locations and barriers similar to the gridworld examples used in RL (Foster et al., 2000; Sutton & Barto, 1998). Exploration and effective performance in a two-dimensional environment can be seen in Figure 4. Here, the virtual rat starts in the middle left (location 4), searches until it finds the goal location in the middle right (location 6), and is reset to the start position when it finds the goal. The greater range of possible movements results in longer pathways during initial exploration (left side of Figure 4A and B1), but ultimately, the virtual agent discovers the reward location and on subsequent trials eventually starts taking the shortest path between start location and reward (as seen on the right side of Figure 4B1). Across 15 simulated rats, this results in an increase in the mean number of rewards received per unit time, as shown in Figure 4B2.

Note that these simulations use Equation E1b in the Methods section. In this equation, the activity of the g_o population during encoding depends on both the new state input and the reverse spread from the goal on the previous retrieval cycle. Although this slows down learning, it actually results in much better overall performance, because strengthening of connectivity progresses backward from the goal location, so that the virtual rat is much more likely to find an optimal pathway. In contrast, use of the alternate Equation E1 results in faster convergence to a single pathway to the goal location, but this pathway is more likely to be nonoptimal, because strengthening progresses forward from the start location without any dependence upon proximity to the goal location. The performance of the network with Equation E1 is shown in Figure 4C. With Equation E1, the network rapidly learns a single pathway to the goal (Figure 4C1), but this is usually a nonoptimal pathway, and can just be a local loop. Across 15 rats, these effects result in a much poorer final average performance well below the optimal level (Figure 4C2). In contrast, Equation E1b results in the network finding

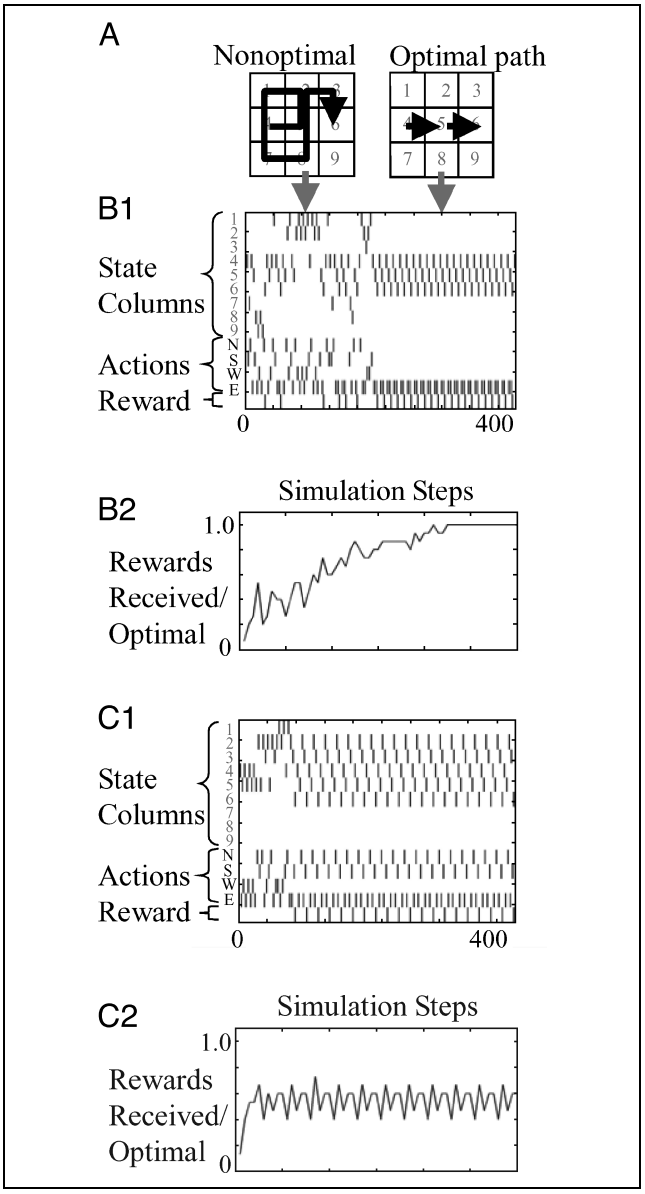

Figure 4. Performance of the network over time in an open-field task. (A) The network starts out with random exploration, resulting in pathways with many steps between the start point and goal (nonoptimal path). After effective encoding, retrieval guides movement along the shortest path directly across the environment (optimal path). (B1) Learning with Equation E1b. Black rectangles in top nine rows (States) show the location of the virtual rat in the environment. Middle four rows show actions (N, S, W, E). After learning of the optimal pathway, the virtual rat repeatedly moves along the shortest path from start to goal (three steps to East). (B2) Average number of rewards received in 15 different simulations versus optimal number per unit time (1 reward for every 3 steps). Initially, pathways are long resulting in a low-average rate of reward. As learning continues, the circuit consistently finds the optimal pathway, resulting in an optimal reward rate (value of 1.0). (C1) Learning with Equation E1. The network rapidly selects a pathway because each transition from state to action to state results in learning, so the first pathway to reward will be repeated. However, the model stays with this pathway which is usually nonoptimal, such as the longer five-step path shown here. (C2) The use of nonoptimal pathways results in a lower-average reward rate over the 15 different simulations. The network also falls into local loops, which contribute to the lower final average rate of reward (around 0.6).

the optimal pathway to reward in all cases (Figure 4B2). The encoding of the environment depends upon the relative amount of random exploration versus exploitation of previous encoding. Greater randomness of exploration in early stages results in a better learning of the environment, resulting in more optimal final behavior. However, this greater randomness results in slower learning and less initial reward. In a sense, greater speed of learning results in less complete understanding of the contingencies of the environment.

The speed of simulations slows down for larger numbers of states, but in real cortical circuits, the neuronal activity and synaptic weights would be computed in parallel, avoiding the sequential computation of large matrices. The sequential learning of states and actions may help to reduce the dimensionality of the matrices being learned. For transitions between n states, the network must modify $n * n$ matrices, but transitions between n states and n' actions requires $2 * n * n'$ connections, which could take advantage of the smaller dimensionality of actions. Learning of the task takes longer in a larger environment due to the time required to randomly encounter the goal and the intervening states. Future models should move beyond the location states used here. For example, in an open field a rat can see the goal location and approach it without needing to learn values for each intervening part of the open space. This may require multiple interacting representations of the environment which can simultaneously guide behavior, as in multiple models based RL (Doya et al., 2002).

Comparison of Model with Temporal Difference Learning

The speed of learning of the prefrontal cortex model was compared with the speed of learning of a traditional actor critic model using TD learning, as shown in Figure 5. This comparison was not performed with the expectation that the prefrontal cortex model would be faster. Instead, it was focused on determining whether the more biologically detailed implementation could attain learning at a behaviorally realistic rate. Real rats require multiple trials (and effective shaping by experimenters) to learn specific goal-directed behaviors in an environment, consistent with the requirement in the simulation for multiple learning trials before optimal performance is obtained. As shown in Figure 5, the prefrontal cortex model could obtain effective function at about one half the rate of TD learning. This was due to the incorporation of both state and action minicolumns into the network, requiring two steps of learning for each action selection, as opposed to the actor critic model where only one step of learning is necessary for each modification of action values at each state. The prefrontal model creates the connectivity necessary for action selection at each state with learning of associations between that state and each action, as well as connections from that action to the resulting subsequent state.

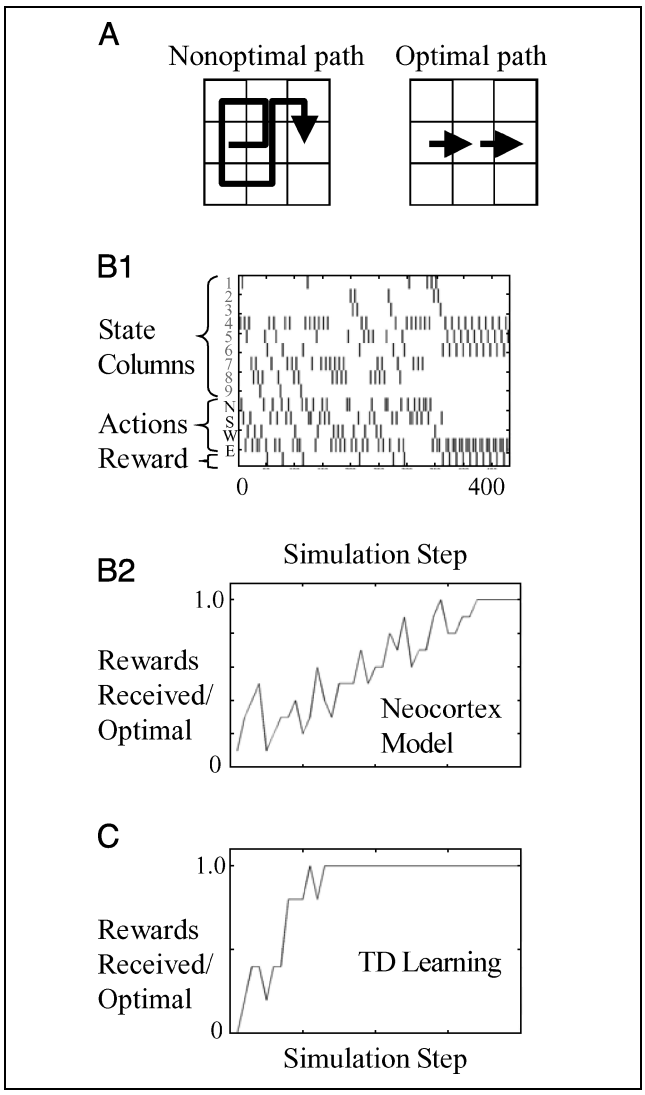

Figure 5. Comparison of neocortex circuit model with pure temporal difference learning. (A) With both mechanisms, the virtual rat starts out taking long pathways to the goal and eventually learns the optimal pathway directly from start point to goal. (B1) Example of rat movements guided by neocortex simulation in open field. (B2) Across 15 simulations, the neocortex model approaches optimal performance about twice as slowly as TD learning. This results from the necessity for two stages of encoding (state to action, and action to state) for connections which mediate the equivalent of an action-value function. (C) Temporal difference learning results in a rapid approach to optimal performance over 30 time steps (average of 15 simulations). Only one step is required for each update of the action-value function look-up table when using the more abstract temporal difference learning rule.

Comparison of Simulation with Neurophysiological Data

The structure of this simulation allows comparison with physiological data on the activity of neurons in the medial prefrontal cortex of rats performing spatial mem-

ory tasks, including movement in an open-field task (Jung et al., 1998) and a spatial alternation task in a figure-of-8 maze (Baeg et al., 2003; Jung et al., 1998). Note that movement in the open field was done with one reward location, corresponding to exploration before finding of one food pellet during foraging. The activity of simulated neurons was plotted in the same manner as experimental data, with shading in a specific location, indicating that the plotted neuron was active when the virtual rat was in that location. In Figure 6, the firing in the open field looks superficially as if it is place dependent, but most neurons do not respond on the basis of spatial location alone. This is similar to experimental data where few true place cells are found, and responses in specific locations are highly variable (Jung et al., 1998). Instead, the g_o neurons initially are active dependent on the prior movement into a particular state. For example, in Figure 6A, the unit codes a Northward movement into the Northwest (upper left) corner, but only fires after subsequent movements including Eastward or Southward movement. These simulations generate the specific experimental prediction that variability of neuronal responses in a specific spatial location should depend upon previous movements. Figure 6 shows that the activity of modeled neurons within the open field are initially relatively localized, but as backward connections from the goal are strengthened, the same neurons should be active when the rat is in a much larger range of spatial locations. The change in neuronal response over time has not been studied, but the distributed firing seen after learning is consistent with experimental data showing firing of medial prefrontal units in a wide range of locations in a familiar environment (Hyman & Hasselmo, unpublished data). Figure 6 also shows a cell from population g_i, which shows no activity before learning has occurred (Late), and a cell from population c_o, which shows activity only for goal-directed actions in specific locations.

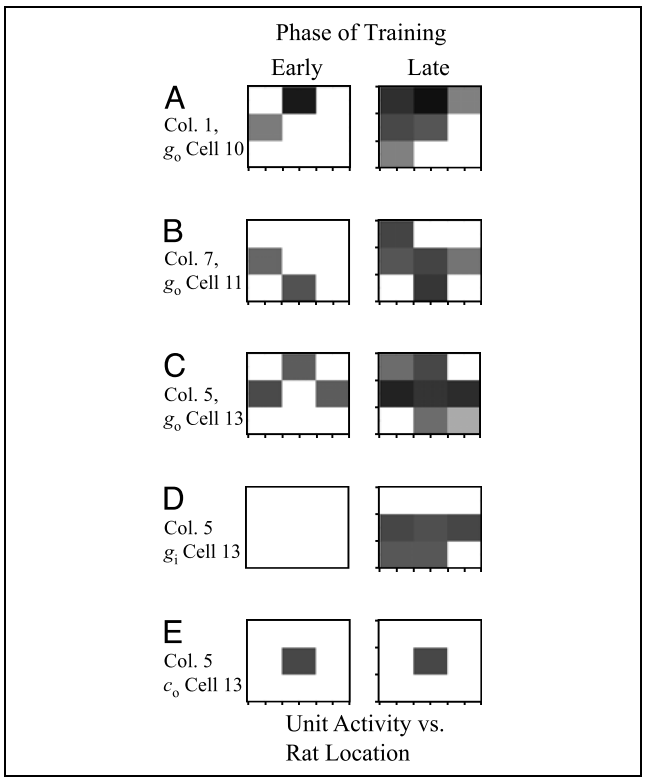

Figure 6. Activity of units plotted according to the location of the virtual rat in the open-field task. White indicates an absence of activity when the virtual rat is in that location, and darker shading indicates greater average activity of a neuron when the rat is in a specific location. Examples on the left show the localized distribution of activity during early stages of training (Early). Examples on the right show the more widely distributed activity after extensive training on the task (Late). Examples A–C show activity of cells in population g_o. D shows typical activity of a g_i cell. E shows activity of a c_o cell.

The same model was utilized to simulate behavior in a spatial alternation task, requiring the addition of a circuit representing hippocampal recall of the previously generated action at each state. This simulation was able to learn the spatial alternation task, as illustrated in Figure 7A, based on activity corresponding to action values for external states and memory states shown in Figure 7B. The firing of simulated units is shown for different locations of the virtual rat in Figure 7C. These simulations show some more consistent responses dependent on spatial location, primarily due to the more constrained nature of prior action at each location. These plots replicate the dependence of many experimentally recorded neurons on the goal location. The response in Figure 7C, Cell 3 resembles goal approach neurons (Jung et al., 1998), whereas the response in Figure 7C, Cell 1 resembles units which respond after visiting the goal location (alternating between bottom right and left). The prominence of goal-related firing arises directly from the dependence of synaptic modification on the backward spread from the goal, which causes units close to the goal to respond earlier and more prominently during learning of the task. The simulations again generate the prediction that the spatial distribution of firing should expand as the task is learned, consistent with the expansion of responses seen in some data (Baeg et al., 2003).

DISCUSSION

The model presented here demonstrates how local circuits of the prefrontal cortex could perform selection of action, and provides a functional framework for interpreting the activity of prefrontal units observed during performance of spatial memory tasks (Baeg et al., 2003; Jung et al., 1998). This circuit model contains populations of threshold units which interact via modifiable excitatory synaptic connections. The retrieval process described here shows how spreading activity in the prefrontal cortex could interact with

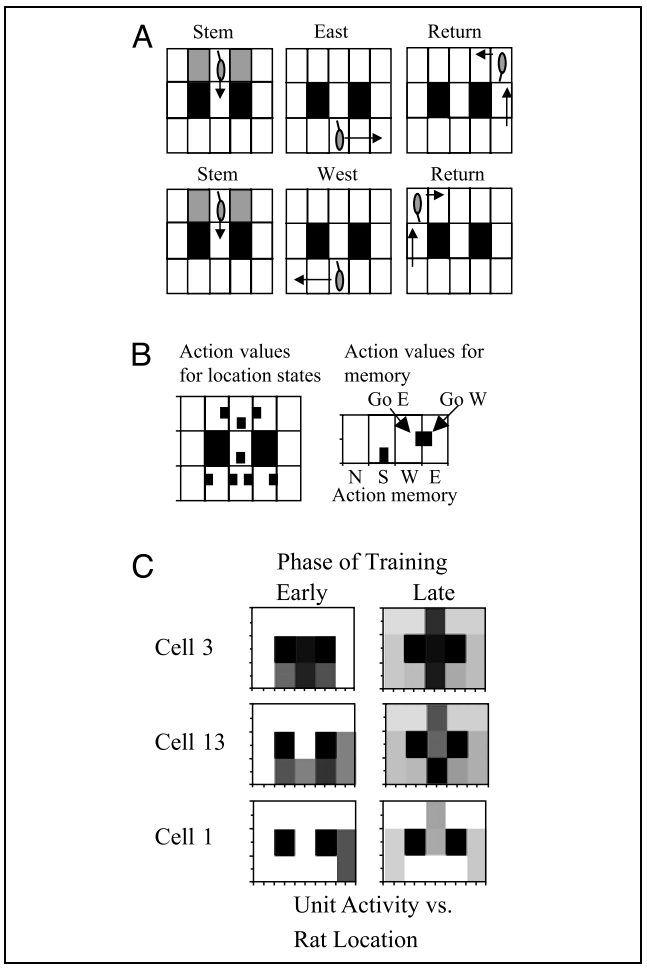

Figure 7. (A) The network guides correct behavior of the virtual rat in the spatial alternation task as shown for two trials on the figure-of-eight maze. (B) Action values plotted for individual locations (states) on the basis of the strength of connections from the c_o units of the state minicolumn with each of the four possible actions at that state. Note that output is ambiguous at the choice point (black rectangles indicate strong connections with both "go East" and "go West" actions). But separate action values from a memory vector of retrieved actions at the state show that the unit representing retrieval of "go East" (E) is associated with "go West" output, and the retrieval of "go West" (W) is associated with "go East" output. (C) Activity of units plotted according to location of the virtual rat in the spatial alternation task. Units show localized responses early in training which predominate in regions of the maze close to reward locations. Cell 3 resembles goal approach cells, whereas Cell 1 resembles cells responding after a visit to the goal location. Responses expand in Late training, but still show specific firing relative to goal.

current sensory input to regulate the selection of the next action necessary for goal-directed behavior. The encoding process described here shows how strengthening of synapses by spike-timing-dependent synaptic plasticity could provide the connectivity patterns necessary for goal-directed behavior. As shown in Figures 6 and 7, the activity of individual units in this model are consistent with some of the properties of neuronal firing activity determined by electrophysiological recordings from the prefrontal cortex (Baeg et al., 2003; Jung et al., 1998). Most simulated neurons show complex relationships to prior actions, rather than simple responses to state, consistent with the rarity of simple place cells in the prefrontal cortex. In addition, neurons in the simulation tend to fire more in regions close to the reward, consistent with evidence for neurons firing during the approach to reward (Baeg et al., 2003; Jung et al., 1998). Research with more detailed integrate-and-fire simulations (Koene & Hasselmo, in press) has replicated some properties of unit firing during performance of a cued response task in monkeys (Schultz et al., 2000). However, the slow speed of simulations with integrate-and-fire neurons does not yet allow learning with random exploration of the environment as utilized here, and that model is difficult to describe using simple equations as presented here. The different types of units used in this model are consistent with other neurophysiological data. Research shows that some units in the prefrontal cortex fire in response to specific sensory stimuli (Wallis et al., 2001; Schoenbaum & Eichenbaum, 1995), consistent with the state representations in a units used here. Research also shows units in the prefrontal cortex which fire during particular motor actions (Wallis et al., 2001; Schoenbaum & Eichenbaum, 1995), consistent with the o and c_o units. Some neurons in the prefrontal cortex change their response to specific stimuli based on changes in the association between stimuli and reward (Mulder et al., 2003; Schoenbaum, Chiba, & Gallagher, 2000; Thorpe, Rolls, & Maddison, 1983). These changes are consistent with the spread of activity from the reward representation across strengthened connections in this model. A change in reward location will cause a change in the pattern of reverse spread during retrieval in this model, resulting in a change in the firing properties of multiple neurons in the network.

Comparison with Functional Components of Reinforcement Learning

The model presented here can perform the same functions as elements of RL. The prefrontal model learns the environment and goal location during exploration, then guides the virtual rat as it follows the shortest pathway from the start location to the goal location. The elements performing this function were developed based on previous simulations of hippocampal function (Hasselmo, Cannon, & Konea, 2002; Hasselmo, Hay, Ilyn, & Gorchetchnikov, 2002), rather than on the elements of RL (Sutton & Barto, 1998). However, these components are clearly related to one another as follows.

In RL, the selection of the next action depends upon the action-value function, a look-up table that has values for all possible actions (4 in this case) in each state (9 in the open field used here). A similar function is obtained here by computing the strength of activity spreading

over output synapses W_o from population c_o. This provides action values for each state, as plotted in Figure 7B. The modification of W_o during encoding, and the strength of c_o during retrieval both depend on propagation of activity back from the goal across multiple connections W_g and W_{ig}, including the strength of connections W_g to a given state from multiple different action minicolumns.

In RL, the selection of a specific action at a specific state is determined by an algorithm which searches only the action values for the current state. This function has been obtained in the circuit model presented here by using an interaction of the sensory input for the current state with the backward spread. Thus, elements in population c_o only spike when they receive both the input from g_o (corresponding to action values) and the inputs from a (corresponding to current state). This allows a circuit model to select the action appropriate for the current state. Here, the unit with largest output activity is selected to guide output. However, rather than choosing the maximum of output activity, the selection of output could use mechanisms which select the first output which crosses the firing threshold. For example, the activity in forward output population c_o could be restricted if we ensure that the first unit which spikes inhibits the activity of other units (Koene & Hasselmo, in press).

In RL, action values are usually trained with TD learning (Sutton, 1988; Sutton & Barto, 1998), or related algorithms such as SARSA (Sutton & Barto, 1998), which propagate value back from the reward state, through adjacent states. A similar function is provided by Equation E1b in this article. During encoding with this equation, the activity of population g_o depends on the spread from the goal/reward. Therefore, reverse connections W_g are only strengthened for transitions to a minicolumn already receiving spread from reward. Because the action value corresponds to W_g, this means that the action value for one minicolumn only increases when a transition is made to another minicolumn with a larger action value or with direct reward input. This resembles the spread of action values through adjacent states caused by TD learning (Sutton, 1988; Sutton & Barto, 1998).

Previously, elements of RL theory have been linked to physiological mechanisms. The activity of dopamine neurons has been related to the error term in TD learning (Schultz et al., 1997). Mechanisms for computation of TD error have been attributed to the basal ganglia (Houk et al., 1995). Changes in parameters of exploration, learning rate, and discounting have been related to neuromodulators such as norepinephrine, acetylcholine, and serotonin (Doya, 2002). The explicit cortical circuit model presented here could allow the literature on RL theory to be extended to other specific physiological properties of neurons within cortical structures.

METHODS

Network Dynamics: Separation of Encoding and Retrieval

This section describes the detailed equations used in these simulations. During each step of a behavioral task, the network dynamics alternate between separate encoding and retrieval phases. This resembles the proposal for separate phases of encoding and retrieval within each cycle of the theta rhythm in the hippocampus (Hasselmo, Bodelon, et al., 2002), and could correspond to phases of theta rhythm observed in the medial prefrontal cortex (Hyman et al., 2002; Manns et al., 2000). The input a is maintained during a full cycle of processing (both encoding and retrieval). During the encoding phase of each cycle, the network sets up sequential forward associations between the previous sensory input (state) and the current motor output (action), as well as associations between the motor output (action) and the subsequent resulting state of sensory input. During each encoding period, thalamic input represents either the current action or the current state. During a period of action input, encoding strengthens reverse associations on synapses W_g between the current motor action and the previous sensory state. In addition, during this period, encoding strengthens output connections W_o, between the selectively activated units in the forward output population c_o and the active elements of the output population o. During a period of state input, encoding forms connections between prior motor actions and the ensuing sensory states, and forms reverse associations W_g between the current sensory state and the previous motor action resulting in that sensory state.

During retrieval, activation of the goal representation causes activity which propagates along reverse connections. The reverse spread of activity from the goal converges with the current state input to activate elements of population c_o that activate a specific appropriate action in the output vector o. In the absence of specific retrieval guiding output, the network reverts to random activation of the output vector to explore states of the environment associated with different actions.

Equations of the Model

"Retrieval": Reverse Spread from Goal

During retrieval, goal-directed activity is initiated by input g_{oR} to the g_o units in the goal minicolumn as shown in Figure 2. This represents a motivational drive state due to subcortical activation of prefrontal cortical representations. This retrieval input g_{oR} in the goal minicolumn is distinct from activation of the unit a in the goal minicolumn during encoding when the goal/reward is actually encountered in the environment.

The activity caused by g_{oR} then spreads back through a range of other minicolumns representing sequences

of states and actions that could lead to the goal. The reverse flow of activity from the goal involves two populations g_i and g_o in each minicolumn, entering a minicolumn through the input population g_i and exiting from the output population g_o. These populations contain one unit for interaction with each other minicolumn, so each minicolumn has n units in g_i and n units in g_o. Because there are n minicolumns in the network, this results in n^2 units in each population: g_i, g_o.

The reverse spread from one minicolumn to a different minicolumn takes place via a matrix W_g providing reverse connections from g_o to g_i. Reverse connections within a minicolumn take place via a matrix W_{ig} providing connections from population g_i to population g_o. The full set of connections across all minicolumns consists of individual matrices defined within individual minicolumns (so W_{ig} is a matrix of matrices) and individual connections between minicolumns (W_g). The reverse flow spreads transiently through units g_i and g_o, but does not persist in these units. The spread of activity from output vector at the previous time step $g_o(t_r - 1)$ across reverse synapses W_g to input vector g_i takes the form:

$$g_i(t_r) = [W_g(g_o(t_r - 1) + g_{oR}]_+ \qquad (R1)$$

Where t_r represents steps of retrieval during one retrieval phase. g_{oR} represents input to elements of g_o in the goal minicolumn during the full period of retrieval. $[\ ]_+$ represents a step function with value zero below the threshold and 1 above threshold. The threshold is set to 0.7 in the simulations presented here. Because this is a binary threshold function and activity spreads at discrete time steps through the network, this network can be replicated relatively easily with integrate-and-fire neurons, but runs much more slowly and cannot be described with simple equations (Koene & Hasselmo, in press).

Reverse spread from the input g_i to the output g_o within a minicolumn involves the matrix W_{ig}, which has modifiable all-to-all connections between g_i and g_o in each minicolumn. To prevent excessive spread of reverse activity, each minicolumn has inhibitory interneurons responding to the sum of excitatory activity in the input g_i, which acts on the output g_o. Both effects are combined in the equation:

$$g_o(t_r) = [(W_{ig} - W_H)g_i(t_r) + g_{oR}]_+ \qquad (R2)$$

Where W_{ig} is the matrix of modified excitatory feedback synapses between the input g_i and the output g_o within each minicolumn, and the matrix W_H consists of elements of strength H (set here to 0.4) for all n by n connections within a minicolumn, but has strength zero between minicolumns.

On each retrieval cycle, retrieval is repeated for R steps. In real cortical structures, the total retrieval steps R would probably be determined by the speed of excitatory synaptic transmission at feedback synapses relative to feedback inhibition and by externally imposed oscillations of activity, such as the theta rhythm (Manns et al., 2000).

Convergence of Forward and Reverse Activity

The network performs a comparison of the reverse flow from goal with activity at the current state, in the form of a summation of two inputs followed by thresholding. The forward output population c_o receives a subthreshold input from the backward input population $g_i(t_r)$ within a minicolumn (via an identity matrix). The forward population c_o also receives subthreshold activity from the units of vector $a(t_r)$ in that minicolumn representing current sensory input. To make them subthreshold, both inputs are scaled by a constant μ weaker than threshold ($\mu = 0.6$).

$$c_o(t_r) = [\mu a(t_r) + \mu g_i(t_r)]_+ \qquad (R3)$$

Thus, an individual unit in the vector c_o will spike only if it receives input from both a and g_i sufficient to bring c_o over threshold. The retrieval dynamics are similar to those used previously (Gorchetchnikov & Hasselmo, in press; Hasselmo, Hay, et al., 2002), in which reverse flow of activity from the goal converges with forward flow from current location. But here the function uses two populations for input and output, allowing multiple pathways through one minicolumn representing a state or an action.

Selection of Output

The choice of one output during retrieval is mediated by the spread of activity from units that were activated in population c_o at the final step of the retrieval period ($t_r = R$). This activity spreads across a set of output connections W_o, which link the populations c_o with the output units o. For the simple example presented in Figure 2, the output vector o consists of two units representing movements of the agent: ''go East'' or ''go West.'' For other simulations, the output population consists of units representing movements of an agent in four directions within a grid: North, South, West, and East.

$$o(t_r) = \max[W_o c_o(t_r)] \qquad (R4)$$

The output (next action) of the network is determined by the selection of the output unit on the basis of the maximum activity spreading across W_o from the population c_o. This equation was also used to com-

pute the action values for each state shown in Figure 7B. The connectivity matrix W_o involves convergence of a relatively large number of units in c_o onto a small number of output units. After effective encoding, each state minicolumn ends up with appropriate connections from units c_o to output units o, similar to action values (Sutton & Barto, 1998). The competitive selection process used here could reflect the function of the basal ganglia, which receive convergent input from the cortex, and contain GABAergic projection cells with inhibitory interactions. If retrieval based on prior experience is strong, then the next action of the virtual rat will primarily depend upon retrieval (i.e., the largest output activity), but the network also has some small probability of generating a random output, in order to allow exploration of all possible actions in the environment. Early in encoding, this random output dominates behavior, allowing exploration, but even after encoding, the output is occasionally chosen as the maximum from a separate random vector. This mechanism represents the effect of stochastic firing properties of neurons within cortical structures (Troyer & Miller, 1997). Random output activity allows the network to explore a range of different possible actions in order to find the best actions for obtaining reward within a given environment (Doya, 2002; Sutton & Barto, 1998).

"Encoding": Formation of Connectivity

Encoding occurs in each state during a separate phase from retrieval. The activity pattern or synaptic connectivity modified by each of the following equations is labeled in Figure 8. The encoding phase consists of a single time step t_e, during which the multiple encoding equations shown below are implemented. Thus, $t_e - 1$ refers to activity retained from the previous encoding phase. This contrasts with the retrieval phases, each of which involves multiple steps t_r up to the maximum R (thus, $t_r - 1$ refers to a previous time step in the same retrieval phase). Encoding modifies the matrices W_g, W_{ig}, and W_c. These matrices start with specific patterns of connectivity representing sparse connections from the output population in minicolumn number o (units 1 to n) to the input population in minicolumn number i (units 1 to n), as follows: $W_g = g_{(i-1)n+o} g^T_{(o-1)n+i} = 0.5$. (The same connectivity was used for W_c). The internal connections W_{ig} start with more extensive connectivity representing intracolumn excitatory connections, as follows: $W_{ig} = g_{(o-1)n+(1...n)} g^T_{(i-1)n+(1...n)} = 0.5$.

Activity in g_o and g_i

For the equations of encoding, first imagine the association between a state (location) which arrives as input on step $t_e - 1$ and a new action generated in this state (which arrives at time step t_e). The state minicolumn has

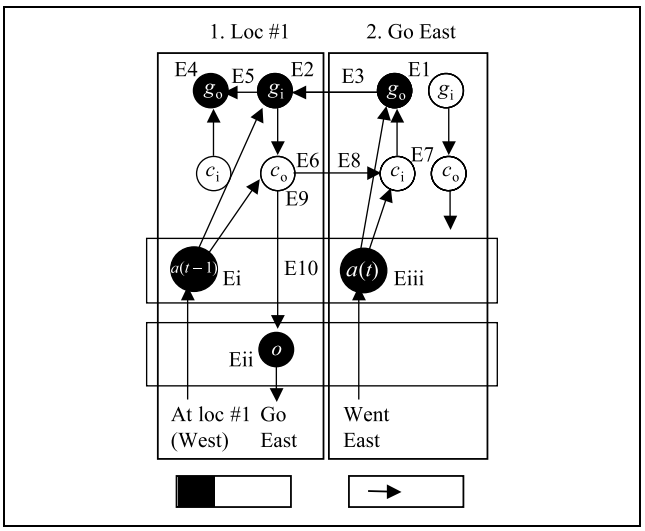

Figure 8. (E1b) Encoding of necessary connections between a state and the action randomly generated at this state. Each step of encoding is listed numerically. (Ei) A buffer in the state minicolumn holds activity from previous input $a(t_e - 1)$. (Eii) A randomly generated output o generates a movement. (Eiii) The proprioceptive feedback of this output causes activity $a(t_e)$ in the action minicolumn. (E1) This activates the g_o population (Equation E1). (E2) Activity spreads to the g_i population in the state minicolumn. (E3) Connections W_g are modified between g_o and g_i. (E4) Buffering of previous activity $c_i(t_e - 1)$ activates g_o in the state minicolumn. (E5) Connections W_{ig} are modified. (E6) The buffer of previous input $a(t_e - 1)$ causes activity in c_o. (E7) Activity in c_o causes activity in c_i. (E8) Connections W_c between c_o and c_i are modified. (E9) Activity in g_i causes activity in c_o. (E10) Connections between c_o and output vector o are modified.

a buffer which holds the prior state input $a(t_e - 1)$, labeled with Ei in Figure 8. Subsequently, the random output $o(t_e)$ is generated, labeled with Eii in Figure 8. This causes proprioceptive input about the action $a(t_e)$, labeled with Eiii in Figure 8. These inputs are then encoded.

Two different forms of encoding were tested, as shown in Figure 4. In one model, the new input vector a causes suprathreshold activation of the g_o population in a specific action minicolumn (Figure 4C).

$$g_o(t_e) = a(t_e) \tag{E1}$$

A different version (Equation E1b) works better and is used in most simulations (Figure 4B). This is:

$$g_o(t_e) = [\mu a(t_e) + \mu g_o(t_r)]_+ \tag{E1b}$$

μ is a constant ($\mu = 0.6$) which must be smaller than the threshold (of 0.7), but large enough for combined input (2μ) to be suprathreshold. In this version, the effect of the input vector a on the reverse output population g_o is made subthreshold, and activity in the population g_o at time t_e (note different time index) only crosses threshold if it converges with backward spread

from the goal as computed by the activity during the final retrieval step on that cycle ($t_r = R$) in the population $g_o(t_r)$. This version of encoding gives the neocortex model properties similar to the TD learning algorithm proposed by Sutton and Barto (1998) and Sutton (1988). This learning rule is not equivalent to TD learning, but does cause modification of connections dependent on an interaction of current state and action with the backward spread from goal (which plays a role similar to the value function in TD learning).

Modification of W_g

Once activity has been induced in the g_o population of the newly activated minicolumn, this activity spreads in the reverse direction back to the minicolumn activated by the previous state, which is activated by a separate buffer holding $a(t_e - 1)$. Spiking network simulations suggest that intrinsic afterdepolarization properties can provide this buffer function in a variety of regions including the prefrontal cortex (Koene & Hasselmo, in press; Koene, Gorchetchnikov, Cannon, & Hasselmo, 2003; Fransén, Alonso, & Hasselmo, 2002; Haj-Dahmane & Andrade, 1998; Klink & Alonso, 1997; Lisman & Idiart, 1995). The population g_i in the previous state minicolumn receives subthreshold input from the buffered representation of $a(t_e - 1)$, and receives subthreshold input from g_o across reverse connections W_g, which start out with weak initial strength. These two subthreshold inputs cause activity in a single unit in g_i which receives both inputs.

$$g_i(t_e) = [\mu a(t_e - 1) + \mu W_g g_o(t_e)]_+ \quad \text{(E2)}$$

In a network with higher time resolution (Koene & Hasselmo, in press), spiking in g_i would follow spiking in g_o by a short delay, allowing spike-timing-dependent Hebbian synaptic plasticity to modify the connections W_g according to:

$$\Delta W_g = g_i g_o^T \quad \text{(E3)}$$

In these simulations, the strength of existing connections started at 0.5 and was limited to a maximum of 1.0, which was reached in a single step when both presynaptic and postsynaptic activities were present. The connectivity of W_g has a specific form meant to represent sparse connectivity between cortical columns. There is only one connection W_g between each pair of minicolumns.

Modification of W_{ig}

In order to link the association between the previous state and the new action with the association between previous action and previous state, the modification of W_g needs to be followed by modification of internal reverse connections W_{ig}, that are all-to-all connections within each minicolumn. Modification of these connections occurs due to persistence of activity in the forward input population of the previous state minicolumn $c_i(t_e - 1)$. This forward input population then supplements the activity of the g_o population, as follows:

$$\Delta g_o(t_e) = c_i(t_e - 1) \quad \text{(E4)}$$

This allows the activity of g_o to be selective for the connection to the minicolumn which received input $a(t_e - 2)$ (due to Equation E6 below). Activity induced in g_o by this buffer follows activity in g_i by a short delay, allowing spike-timing-dependent Hebbian plasticity to modify connections W_{ig} as follows:

$$\Delta W_{ig} = g_o g_i^T \quad \text{(E5)}$$

Activity in c_o and c_i

Population c_o is updated by the buffer of prior input $a(t_e - 1)$:

$$c_o(t_e) = a(t_e - 1) \quad \text{(E6)}$$

This activity then spreads forward over the weak initial strength of forward connections to converge with subthreshold input of current input $a(t_e)$ to induce activity in specific units of the population c_i in the new minicolumn as follows:

$$c_i(t_e) = [\mu a(t_e) + \mu W_c c_o(t_e)]_+ \quad \text{(E7)}$$

Modification of W_c

The modification of forward connections does not play a strong functional role in the examples presented here, but will be important for forward planning evaluating possible forward pathways. The modification of the forward connections W_c uses the new activity c_o and c_i:

$$\Delta W_c = c_i c_o^T \quad \text{(E8)}$$

Modification of Output Weights W_o

Finally, the output population $c_o(t_e)$ is associated with the current activity in the output population $o(t_e)$. The activity in the output population was previously generated by the action currently being encoded by the network. Initially, these outputs are generated randomly. On each step, the network learns the association between activity in a specific unit of c_o (which is activated by the reverse connection input to g_i) and the element of the output vector which caused this output. This would allow effective learning of the map-

ping between internal representations and output populations even without highly structured connectivity. The activity in the output forward population is set by the input reverse population:

$$c_o(t_e) = g_i(t_e) \tag{E9}$$

Then the output weights are modified according to the activity at this time step:

$$\Delta W_o = o c_o^T \tag{E10}$$

These stages of encoding allow spike-timing-dependent synaptic plasticity to strengthen the connections necessary for the retrieval process described in the earlier section. As shown in Figure 9, the representation of each movement from one location to another requires two steps of encoding. The first forms associations between the prior location [vector $a(t_e - 1)$] and the proprioceptive feedback of the randomly generated action [vector $a(t_e)$]. The second forms an association between the proprioceptive representation of the randomly generated action [the action vector which is now $a(t_e - 1)$], and the new state [now represented by the vector $a(t_e)$].

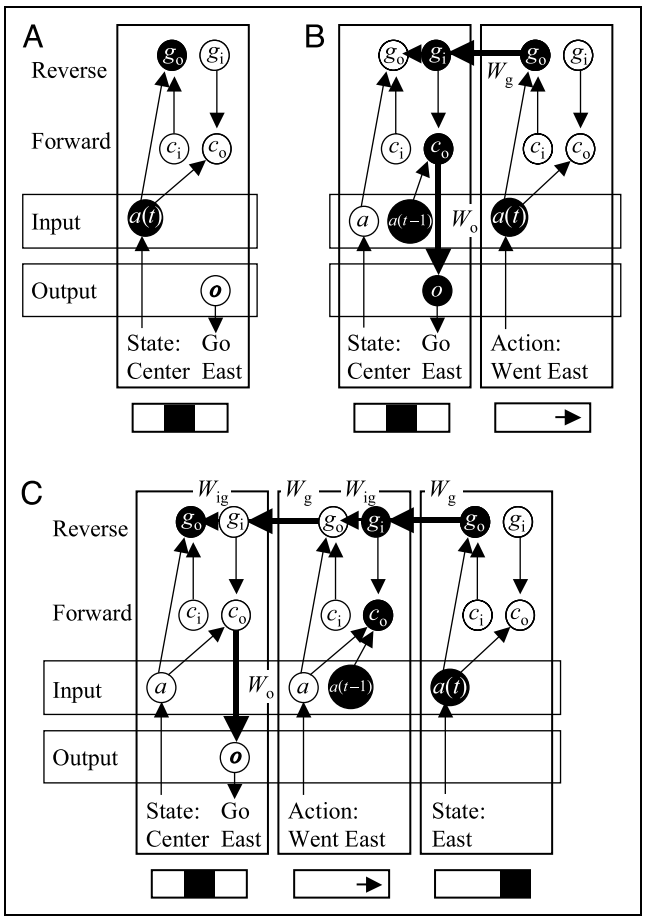

Figure 9. Simple schematic illustration of the steps of synaptic modification during encoding. (A) The network receives state input which activates $a(t_e)$. (B) An output action is generated randomly. The network receives proprioceptive feedback about the new action $a(t_e)$ which activates g_o in a separate action minicolumn. Connections W_g are strengthened between g_o and population g_i for the previous state, which buffers activity from $a(t_e - 1)$, followed by strengthening of W_{ig} between g_i and g_o. Population g_i activates population c_o, which strengthens connections W_o with the output o. (C) The network receives sensory input $a(t_e)$ about the next state which activates a new minicolumn. Connections are strengthened between this new state minicolumn population g_o and the preceding g_i and g_o of the action minicolumn.

Acknowledgments

This work was supported by NIH DA16454, DA11716, NSF SBE-0354378, NIH MH61492, and NIH60013.

REFERENCES

Baeg, E. H., Kim, Y. B., Huh, K., Mook-Jung, I., Kim, H. T., & Jung, M. W. (2003). Dynamics of population code for working memory in the prefrontal cortex. *Neuron, 40,* 177–188.

Barbas, H., & Pandya, D. N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology, 286,* 353–375.

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

Brown, J., Bullock, D., & Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience, 19,* 10502–10511.

Corbit, L. H., & Balleine, B. W. (2003). The role of prelimbic cortex in instrumental conditioning. *Behavioural Brain Research, 146,* 145–157.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks, 15,* 1347–1369.

Doya, K., Smejima, K., Katagiri, K., & Kawato, M. (2002). Multiple model-based reinforcement learning. *Neural Computation, 14,* 1347–1369.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Ferbinteanu, J., & Shapiro, M. L. (2003). Prospective and retrospective memory coding in the hippocampus. *Neuron, 40,* 1227–1239.

Foster, D. J., Morris, R. G., & Dayan, P. (2000). A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus, 10,* 1–16.

Fransén, E., Alonso, A. A., & Hasselmo, M. E. (2002). Simulations of the role of the muscarinic-activated calcium-sensitive nonspecific cation current INCM in entorhinal neuronal activity during delayed matching tasks. *Journal of Neuroscience, 22,* 1081–1097.

Fuster, J. M. (1995). *Memory in the cerebral cortex.* Cambridge: MIT Press.

Gorchetchnikov, A., & Hasselmo, M. E. (in press). A model of prefrontal, septal, entorhinal and hippocampal interactions to solve multiple goal navigation tasks. *Connection Science.*

Gothard, K. M., Skaggs, W. E., & McNaughton, B. L. (1996). Dynamics of mismatch correction in the hippocampal ensemble code for space: Interaction between path integration and environmental cues. *Journal of Neuroscience, 16,* 8027–8040.

Haj-Dahmane, S., & Andrade, R. (1998). Ionic mechanism of the slow afterdepolarization induced by muscarinic receptor activation in rat prefrontal cortex. *Journal of Neurophysiology, 80,* 1197–1210.

Hasselmo, M., Cannon, R. C., & Koene, R. A. (2002). A simulation of parahippocampal and hippocampal structures guiding spatial navigation of a virtual rat in a virtual environment: A functional framework for theta theory. In M. P. Witter & F. G. Wouterlood (Eds.), *The parahippocampal region: Organisation and role in cognitive functions* (pp. 139–161). Oxford: Oxford University Press.

Hasselmo, M. E., Bodelon, C., & Wyble, B. P. (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation, 14,* 793–817.

Hasselmo, M. E., Hay, J., Ilyn, M., & Gorchetchnikov, A. (2002). Neuromodulation, theta rhythm and rat spatial navigation. *Neural Networks, 15,* 689–707.

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

Hyman, J. M., Wyble, B. P., Goyal, V., Rossi, C. A., & Hasselmo, M. (2003). Stimulation in hippocampal region CA1 in behaving rats yields LTP when delivered to the peak of theta and LTD when delivered to the trough. *Journal of Neuroscience, 23,* 11725–11731.

Hyman, J. M., Wyble, B. P., Rossi, C. A., & Hasselmo, M. E. (2002). Coherence between theta rhythm in rat medial prefrontal cortex and hippocampus. *Society of Neuroscience Abstracts, 28,* 476–477.

Jung, M. W., Qin, Y., McNaughton, B. L., & Barnesm, C. A. (1998). Firing characteristics of deep layer neurons in prefrontal cortex in rats performing spatial working memory tasks. *Cerebral Cortex, 8,* 437–450.

Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex, 13,* 400–408.

Klink, R., & Alonso, A. (1997). Muscarinic modulation of the oscillatory and repetitive firing properties of entorhinal cortex layer II neurons. *Journal of Neurophysiology, 77,* 1813–1828.

Koene, R. A., Gorchetchnikov, A., Cannon, R. C., & Hasselmo, M. E. (2003). Modeling goal-directed spatial navigation in the rat based on physiological data from the hippocampal formation. *Neural Networks, 16,* 577–584.

Koene, R. A., & Hasselmo, M. E. (in press). An integrate and fire model of prefrontal cortex neuronal activity during performance of goal directed decision-making. *Cerebral Cortex.*

Levy, W. B., & Steward, O. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience, 8,* 791–797.

Lewis, D. A., Melchitzky, D. S., & Burgos, G. G. (2002). Specificity in the functional architecture of primate prefrontal cortex. *Journal of Neurocytology, 31,* 265–276.

Lisman, J. E., & Idiart, M. A. (1995). Storage of 7 +/− 2 short-term memories in oscillatory subcycles. *Science, 267,* 1512–1515.

Manns, I. D., Alonso, A., & Jones, B. E. (2000). Discharge profiles of juxtacellularly labeled and immunohistochemically identified GABAergic basal forebrain neurons recorded in association with the electroencephalogram in anesthetized rats. *Journal of Neuroscience, 20,* 9252–9263.

Markram, H., Lubke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science, 275,* 213–215.

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews: Neuroscience, 1,* 59–65.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16,* 1936–1947.

Mulder, A. B., Nordquist, R. E., Orgut, O., & Pennartz, C. M. (2003). Learning-related changes in response patterns of prefrontal neurons during instrumental conditioning. *Behavioural Brain Research, 146,* 77–88.

Rao, S. G., Williams, G. V., & Goldman-Rakic, P. S. (1999). Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: Evidence for microcolumnar organization in PFC. *Journal of Neurophysiology, 81,* 1903–1916.

Scannell, J. W., Blakemore, C., & Young, M. P. (1995). Analysis of connectivity in the cat cerebral cortex. *Journal of Neuroscience, 15,* 1463–1483.

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neuroscience, 1,* 155–159.

Schoenbaum, G., Chiba, A. A., & Gallagher, M. (2000). Changes in functional connectivity in orbitofrontal cortex and basolateral amygdala during learning and reversal training. *Journal of Neuroscience, 20,* 5179–5189.

Schoenbaum, G., & Eichenbaum, H. (1995). Information coding in the rodent prefrontal cortex: I. Single-neuron activity in orbitofrontal cortex compared with that in pyriform cortex. *Journal of Neurophysiology, 74,* 733–750.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.

Schultz, W., Tremblay, L., & Hollerman, J. R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex, 10,* 272–284.

Sutton, R. S. (1988). Learning to predict by the method of temporal differences. *Machine Learning, 3,* 9–44.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning (adaptive computation and machine learning).* Cambridge: MIT Press.

Thorpe, S. J., Rolls, E. T., & Maddison, S. (1983). The orbitofrontal cortex: neuronal activity in the behaving monkey. *Experimental Brain Research, 49,* 93–115.

Troyer, T. W., & Miller, K. D. (1997). Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell. *Neural Computation, 9,* 971–983.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature, 411,* 953–956.

Wood, E. R., Dudchenko, P. A., Robitsek, R. J., & Eichenbaum, H. (2000). Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron, 27,* 623–633.

Wyble, B. P, Hyman, J. M, Rossi, C. A., & Hasselmo, M. (2004). Analysis of theta power in hippocampal EEG during bar pressing and running behavior in rats during distinct behavioral contexts. *Hippocampus, 14,* 368–384.

Zhu, S., & Hammerstrom, D. (2003). Reinforcement learning in associative memory. In D. Wunsch & M. E. Hasselmo (Eds.), *International Joint Conference on Neural Networks* (pp. 1346–1350). Portland, OR: IEEE.