An integrate and fire model of prefrontal cortex neuronal activity during performance of goal-directed decision making

Randal A. Koene
Michael E. Hasselmo
Center for Memory and Brain
Department of Psychology and Program in Neuroscience
Boston University
64 Cummington Street, Boston, MA 02215, U.S.A.
tel: 1-617-358-2769, fax: 1-617-353-1424
randalk@bu.edu

November 23, 2005

Running title: Prefrontal cortex model

Abstract

The orbital frontal cortex appears to be involved in learning the rules of goal-directed behavior necessary to perform the correct actions based on perception to accomplish different tasks. The activity of orbitofrontal neurons changes dependent upon the specific task or goal involved, but the functional role of this activity in performance of specific tasks has not been fully determined. Here we present a model of prefrontal cortex function using networks of integrate-and-fire neurons arranged in minicolumns. This network model forms associations between representations of sensory input and motor actions, and uses these associations to guide goal-directed behavior. The selection of goal-directed actions involves convergence of the spread of activity from the goal representation with the spread of activity from the current state. This spiking network model provides a biological implementation of the action selection process used in reinforcement learning theory. The spiking activity shows properties similar to recordings of orbitofrontal neurons during task performance.

Keywords: Orbitofrontal, minicolumns, selective activity, reinforcement learning

The orbitofrontal cortex plays an important role in goal-directed behavior (Wallis *et al.*, 2001). Lesions of the orbitofrontal cortex impair the ability of animals to learn which stimuli are associated with reward (Pears *et al.*, 2003; Izquierdo and Murray, 2004; Miller and Cohen, 2001; Frey and Petrides, 1997; Bechara *et al.*, 1994, 1997). Recordings from orbitofrontal cortex neurons demonstrate that spiking activity in response to sensory stimuli changes dependent upon the association of a stimulus with a reward in humans (Rolls, 1999), non-human primates (Thorpe *et al.*, 1983; Schultz *et al.*, 2000; Wallis and Miller, 2003) and rats (Mulder *et al.*, 2003; Schoenbaum and Eichenbaum, 1995a,b; Schoenbaum and Ramus, 2003). The orbitofrontal cortex appears to be particularly important when the generation of specific actions depends upon the context of particular sensory stimuli (Miller and Cohen, 2001). Here we focus on behavior directed toward a specific goal, we do not yet deal with decisions about the relative value of different goals (Balleine and Dickinson, 1998; Tremblay and Schultz, 1999).

Here we present a computational model that is applicable to multiple regions of the prefrontal cortex (PFC), demonstrating how populations of spiking neurons could mediate goal-directed behavior. In particular, we demonstrate how representations of specific motor actions can be used for goal-directed behavior in multiple different circumstances, dependent upon the context of specific sensory stimuli. This modeling effectively simulates the behavior and pattern of activity of orbitofrontal cortex neurons described in an experiment by Schultz, Tremblay and Hollerman (Schultz *et al.*, 2000), namely neurons that show response to sensory stimuli, response to reward and to expectation of reward. This task involves the differential generation of Go versus NoGo responses to randomly presented visual cues. Recordings demonstrated that some neurons in orbitofrontal cortex do indeed fire selectively for the transition from one specific state to another. Schultz *et al.* (Schultz *et al.*, 2000) identified these neurons, labeling them selective for the instruction that initiates a specific trial, as well as predictive for a specific action.

Previous models of frontal cortex function have used neurons with sigmoid input-output functions which represent firing of populations of neurons (Cohen and Servan-Schreiber, 1992; O'Reilly and Munakata, 2000). In order to more directly model the patterns of spiking activity during behavioral tasks, we use integrate-and-fire neurons (Stein, 1967; Gerstner, 2002; Gerstner and Kistler, 2002) with Hebbian spike-timing dependent synaptic plasticity (STDP) (Levy and Stewart, 1983). Integrate-and-fire neurons simulate the membrane potential response to the build-up of synaptic input over time and emit a spike when the potential crosses threshold. The model shows how integrate-and-fire neurons can perform the functions described in equations for a circuit model of prefrontal cortex (Hasselmo, 2005). The structure of the model was motivated by anatomical evidence suggesting the organization of neural circuits into minicolumns (Lund *et al.*, 1993), cell assemblies of highly interconnected neurons found in PFC. In our model, different minicolumns responded to both sensory input and motor actions, consistent with evidence (Fuster, 1973; Fuster *et al.*, 1982; Funahashi *et al.*, 1989; Quintana and Fuster, 1992; Fuster, 2000) that

activity in the prefrontal cortex represents two types of perception — 1.) the perception of past sensory stimuli available due to short-term buffers and current sensory stimuli, and 2.) the proprioceptive sensation and prediction of motor actions. The organization into minicolumns was motivated by evidence for strong excitatory and inhibitory connectivity within local circuits of cortical neurons (Mountcastle, 1997; Lübke and von der Malsburg, 2004). The rapid strengthening of associations between sensory states, motor actions and reward is motivated by studies showing rapid changes in functional interactions between populations of prefrontal neurons during learning (Thorpe *et al.*, 1983; Schoenbaum *et al.*, 2000; Mulder *et al.*, 2003).

The structure of this model closely resembles features of reinforcement learning (Sutton and Barto, 1998; Schultz *et al.*, 1997), so we will commonly refer to sensory information from the environment as "state". We will refer to motor output as "actions" and to the desired goal as "reward". However, this model does not focus on the temporal difference learning rule (Sutton, 1988), a rule that uses the difference between successive outputs as error measure. Instead it focuses on mechanisms of action selection associated with specific sensory states and reward. This demonstrates how integrate-and-fire neurons can perform the circuit mechanism of action selection proposed in a more abstract model of prefrontal cortex (Hasselmo, 2005).

In the following sections we simulate the proposed mechanism of the prefrontal minicolumn circuitry and apply that to the delayed Go/NoGo task with its reward protocol for different stimuli. We focus on explaining selective neuronal activity, as recorded by W.Schultz *et al*, with our model.

1 METHODS

This model focused on replicating neuronal activity and behavior in the experiments by Schultz *et al.*. In these experiments, an initial visual stimulus indicates one of three possible trials (fig. 1A): (1) rewarded movement stimulus (Srm), reward is given if the monkey presses a key; (2) rewarded non-movement stimulus (Srnm), reward is given if the monkey chooses not to press the key; (3) unrewarded movement stimulus (Surm), the reward is not given, but the key press is required. Unless the movement is performed in the Surm trial, another unrewarded Surm trial follows. The decision to move or not to move followed a delay of 2 seconds, when a trigger signal was given, which was identical in each trial. Schultz *et al.* found that orbitofrontal neurons that showed task related activity fired selectively. Some responded with increased firing rates to a specific instruction cue, some responded with increased firing rates predictive of Go/NoGo choice according to the expectation of reward, and some responded with increased firing rates to reward received.

We propose that goal directed behavior is learned by associating states and actions that are separately represented by the population of neurons of individual minicolumns. A state is indicated by the perception of specific sensory stimuli or the perception of reward received, while an action is indicated by pro-

prioceptive input about motor activity. According to our hypothesis, the initial states **Srm**, **Srnm** and **Surm**, as well as the **Reward** state are represented by activity in individual minicolumns in PFC, while activity in a further two minicolumns represents action selections **Go** (move to press a key) or **NoGo**. During learning of goal-directed behavior, STDP strengthens connections within and between minicolumns so that state and action representations are associated. Because activity that corresponds to consecutive states and actions may appear at arbitrary time intervals, a short-term buffer based on persistent spiking due to after-depolarization (ADP) of membrane potential (Andrade, 1991; Klink and Alonso, 1997b) is used to enable encoding with STDP (Lisman and Idiart, 1995; Jensen *et al.*, 1996; Koene *et al.*, 2003).

We propose that the retrieval of goal-directed behavior depends on the spread of activity through strengthened connections from a minicolumn that represents the reward state and from the specific state minicolumn activated by current input. Consistent with this hypothesis, experimental evidence indicates that retrieval in PFC produces goal-directed activity that is initiated by the desire for a goal (Schultz, 1998; Schultz and Dickinson, 2000; Miller and Cohen, 2001). In our model, the spread of activity from the representation of current state is gated by the spread from a desired goal. When the gated spread produces output from the minicolumn that represents current state the correct next action is selected. Hence, the convergence of activity from a current state representation and from a goal representation governs goal-directed behavioral responses.

Given the representation of states and actions, the transition from one state to another state via a specific action can be encoded uniquely if there is specific neural activity that occurs only for that action and only when the action is initiated in a particular state. This requirement leads to the presupposition that a functional minicolumn contains populations of input neurons and populations of output neurons that form connections with other minicolumns, and that the neurons in those populations are connected in a structured manner to other minicolumns (in this simulation to exactly one). Since the combination of activity at a specific input neuron and a specific output neuron of an action minicolumn represents the transition from a preceding state to a following state, that information gives the model the Markov property (Sutton and Barto, 1998). With this property, one-step dynamics enable us to predict the next state and expected reward for a specific action.

We developed simulations of the Schultz *et al.* task with Catacomb2 (Cannon *et al.*, 2003) that replicated the actions of an agent (monkey) within an environment, as well as integrate-and-fire neuron dynamics in PFC. With our approach[1], data from a simulated operant task protocol was linked with simulated neuronal circuitry for sensory processing and functions of the prefrontal cortex (see fig. 1B). Further details of the neurophysiology were modeled explicitly where needed for specific functional requirements, such as the after-depolarization experienced by specific neuron populations that may enable per-

[1]We call the method "design-based" modeling.

sistent firing.

The integrate-and-fire neurons in our model of PFC minicolumns have a resting and reset potential of -60 mV and an exponential decay time constant of 10 ms. The firing threshold is -50 mV and action potentials have a duration of 1 ms, followed by a 2 ms refractory period and subsequent strong after-hyperpolarization with reversal potential -90 mV and exponential decay time constant 30 ms. We used dual-exponential functions for the responses of synaptic conductances. Unless the description of a specific synaptic connection indicates otherwise, the time constant for the rise of the dual-exponential response function was 2 ms and the time constant for the fall was 4 ms. Excitatory synaptic connections had a reversal potential of 0 mV and inhibitory synaptic connections had a reversal potential of -70 mV.

In the simulation of the operant task environment, stimuli produced by visual cues and reward, as well as proprioceptive sensation of motor activity are conveyed as spike trains (top of fig. 2) that are produced by specific neurons (signal pathway (a) in fig. 1B). The simulation of perceptual processing circuitry receives those spike trains and transforms them into reliable sequences of **state-action** spike pairs (bottom of fig. 2). Every time that a spike train corresponding to a new state or a new motor action is detected, a pair of spikes is generated that represents the most recent state and the most recent action. The individual spike times of a state-action spike pair are separated by several cycles of the theta rhythm to insure that persistent spiking of the most recent two spike inputs to the short-term buffer occurs over a sufficient duration to achieved strong associative connections through STDP. To simplify the readability of the graphs, an identity matrix is used for input connections to the set of PFC minicolumns instead of a learned mapping (signal pathway (b) in fig. 1B). Motor action in the operant task is driven by the output of prefrontal minicolumns (signal pathway (c) in fig. 1B). In this manner, the seven trials shown in figure 2 are simulated during encoding so that all relevant rules are learned in the network of prefrontal minicolumns.

Specific neuron populations within prefrontal minicolumns achieve the gating of the forward spread of activity by spread from the goal

Retrieval and encoding of associations between prefrontal minicolumns that represent states and actions are assumed to take place in opposite phase intervals of rhythmic modulation at 8Hz (Hasselmo *et al.*, 2002) that represents the theta rhythm found in the prefrontal cortex and hippocampus (Manns *et al.*, 2000). This enables both to occur at any time during a task. The modulation supports different dynamics in the two modes. We will therefore discuss the distinct functions of encoding and retrieval separately, even though they alternate continuously during a simulated task. The modulating rhythm also serves to insure that activity in different simulated brain regions is properly synchronized, as described in our previous work (Koene *et al.*, 2003). The plot of membrane potential for the buffer neuron $\mathbf{a}_{buf}(Rew)$ in fig. 6B provides an example of the modulation by theta rhythm and clearly demonstrates rhythmic changes at 125

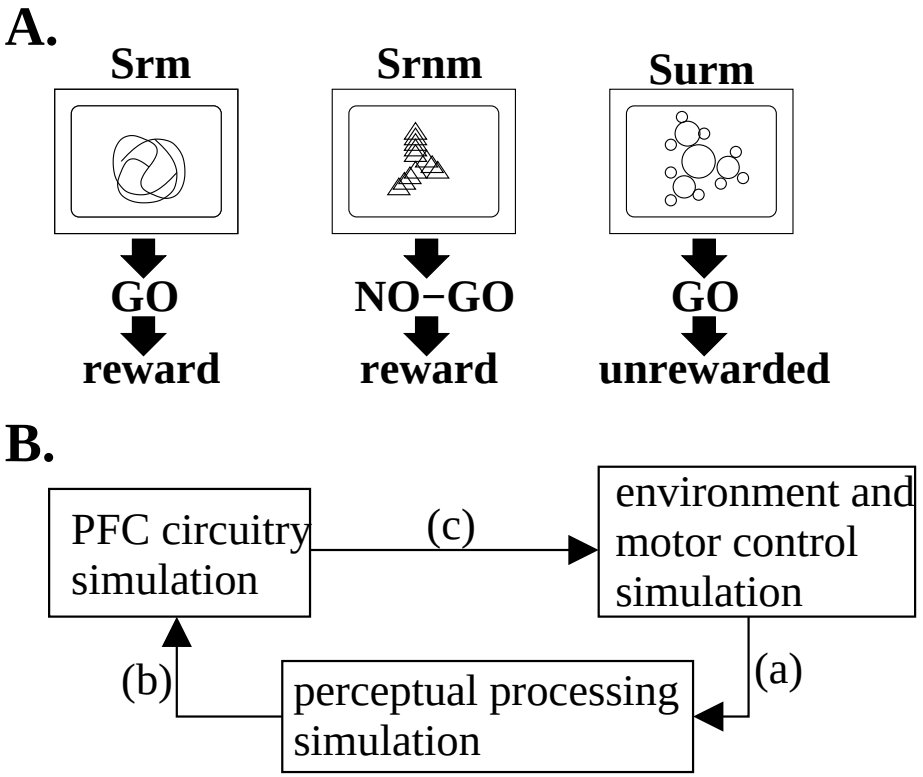


Figure 1: **A.** Summary of the Schultz *et al.* task. Three visual stimuli indicated above (fractal images) and different types of behavioral trials as in the simulation. **B.** Design of the simulation. The simulation includes the experimental environment of the operant task in terms of the task protocol, visual stimuli and motor actions. (a) The output of that simulation goes into the perceptual segment of the simulation. Perceptual stimuli are represented by spike trains, which are processed to produce spike pairs that are used as an internal representation. (b) The resulting neuronal spikes cause activity in a simulation of minicolumns in prefrontal cortex that includes specifics of relevant neurophysiology and neuroanatomy. (c) Feedback from the output of the simulated prefrontal cortex directs motor action in the operant task. The functions of integrate-and-fire neurons and other essential components were implemented in Catacomb2.

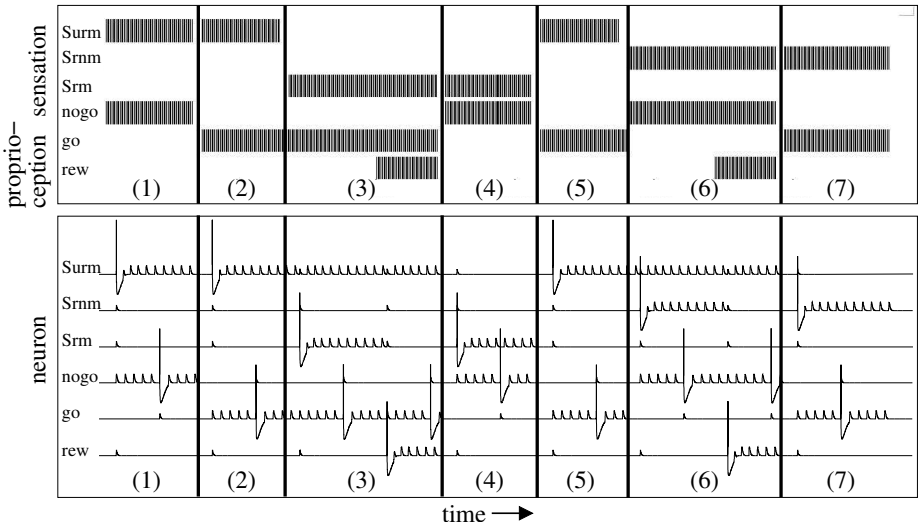


Figure 2: Input spike trains of sensory input (top) and membrane potential showing spike pairs that are the internal representation of changes of state or action (bottom). Vertical lines separate trials (after which buffers are cleared). Rules are learned by exposure to both rewarded and non-rewarded conditions in seven different trials: (1) NoGo following Surm does not lead to reward, (2&5) Go following Surm leads to rewarded trial, (3) Srm and Go leads to reward, (4) Srm and NoGo does not lead to reward, (6) Srnm and NoGo leads to reward, (7) Srnm and Go does not lead to reward.

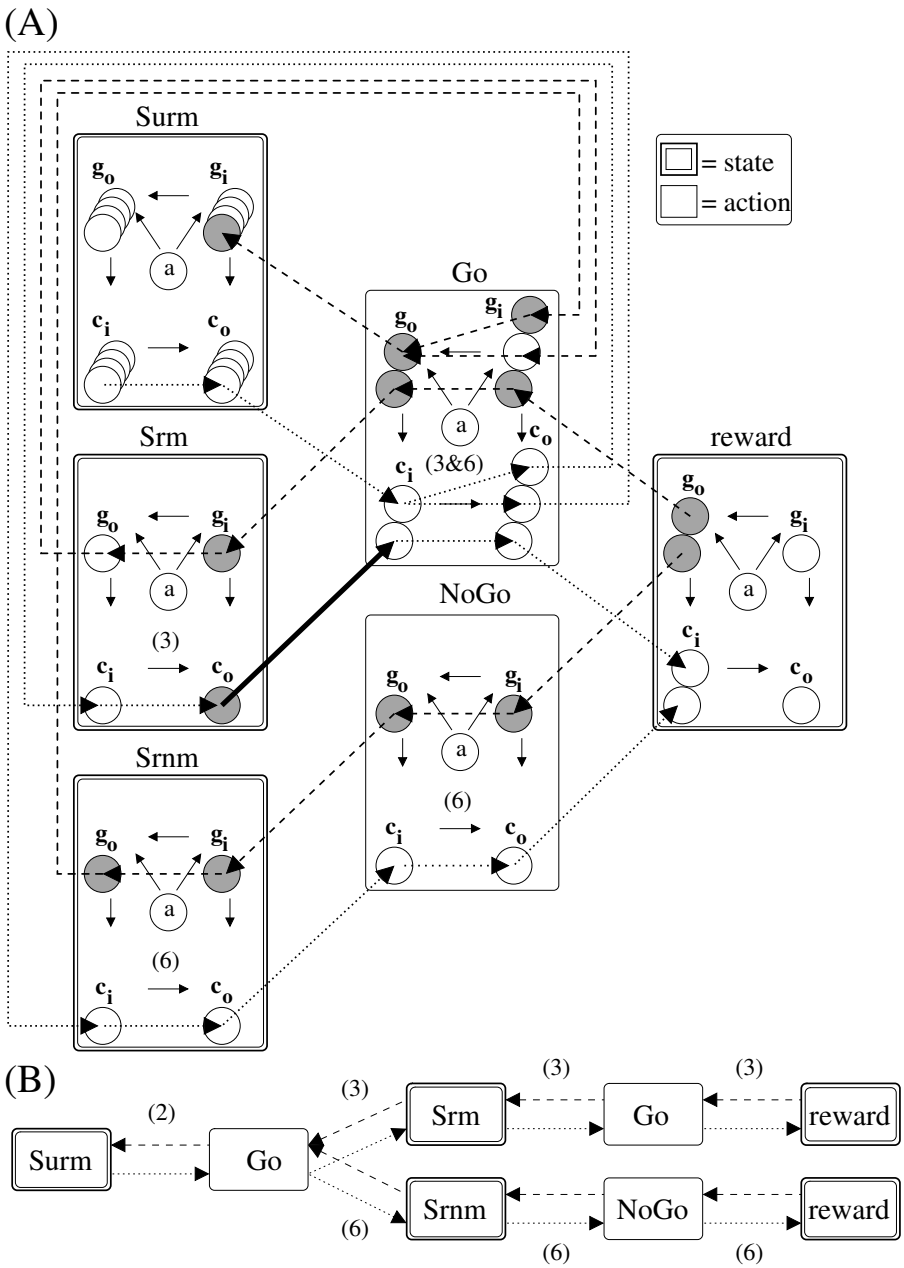
ms intervals.

As shown in fig. 3, we distinguish five populations of pyramidal neurons in each presupposed functional minicolumn of PFC: $\mathbf{a}$, $\mathbf{g}_i$, $\mathbf{g}_o$, $\mathbf{c}_i$ and $\mathbf{c}_o$. Of these, each $\mathbf{a}$ neuron connects exclusively to other neurons within the same minicolumn and play an important role during encoding of associations between minicolumns. These $\mathbf{a}$ neurons represent neurons that receive thalamic input in layer IV of PFC. The neurons of a population labeled $\mathbf{g}_o$ experience suprathreshold depolarization during encoding in response to input from $\mathbf{a}$ (with a fixed conductance of 5.2 nS and time constants 1 ms for the rise and 2 ms for the fall of the synaptic response), but during retrieval $\mathbf{g}_o$ is inhibited by an interneuron network that is driven by $\mathbf{a}$. A spike in $\mathbf{a}$ during encoding also provides subthreshold depolarization to all neurons of a population labeled $\mathbf{g}_i$ (with a fixed conductance of 1.0 nS and time constants 12 ms for the rise and 20 ms for the fall of the synaptic response).

The output of each neuron in the $\mathbf{g}_o$ population projects to one of the other minicolumns in the PFC network. In the $\mathbf{g}_i$ population, each neuron receives one connection from a $\mathbf{g}_o$ neuron located in another minicolumn. Synaptic weights are modifiable on these connections between different minicolumns and are the elements of a matrix $\mathbf{W}_g$. When strengthened, the $\mathbf{W}_g$ connection can fire a unit $\mathbf{g}_i$ if the presynaptic unit $\mathbf{g}_o$ is active. Such a connection indicates that a rule was learned that expresses the knowledge that activity in the minicolumn containing the presynaptic neuron $\mathbf{g}_i$ preceded activity in the minicolumn of the connected $\mathbf{g}_o$ neuron.

Similarly, each neuron of a population $\mathbf{c}_o$ makes one connection to a neuron in a $\mathbf{c}_i$ population of another minicolumn, so that activity in the $\mathbf{c}_o$ population can target any one of the other minicolumns specifically. Again, the synaptic strengths of such connections are modifiable and make up elements in a matrix

Figure 3: During training, associations are learned between state and action minicolumns. The network of minicolumns (A) is shown with the connections between them. Activity spreads along associations directed both from the minicolumn representing the goal (dashed arrows) and forward from the minicolumn representing the current state (dotted arrows). To simplify the schematic, populations of neurons, $\mathbf{g}_i$, $\mathbf{g}_o$, $\mathbf{c}_i$ and $\mathbf{c}_o$ as shown in the Surm minicolumn were reduced in the other minicolumns to display only those neurons that are involved in encoded associations. The numbers in brackets correspond to the marked training trials in figure 2, in which an associative connection is established by STDP. Here, activity in the neuronal populations of the minicolumns is indicated by shaded neurons. This is shown for retrieval of the correct action that leads to reward from a current state, Srm, in which the rewarded move stimulus was perceived. Neurons that spike are circles shaded gray. A separate diagram (B) shows a linear representation of the associative connections that are strengthened during rule learning (numbers in brackets again correspond to training trials in fig. 2). The Go and Reward minicolumns each fulfill two roles in the encoded rules.


(A)
Surm
Srm
Srnm
Go
NoGo
reward
= state
= action
(3)
(3&6)
(6)
(B)
Surm
Go
Srm
Srnm
Go
NoGo
reward
(2)
(3)
(3)
(3)
(6)
(6)
(6)

$\mathbf{W}_c$. Unlike the effect of synaptic weights in $\mathbf{W}_g$, postsynaptic depolarization due to input through a connection with the maximum strength in $\mathbf{W}_c$ is subthreshold, so that spiking in $\mathbf{c}_i$ remains dependent on additional input. The additional input to neurons in $\mathbf{c}_i$, which can elevate their membrane potential over threshold, is supplied by one-to-one connections[2]from neurons in $\mathbf{g}_o$ (with a conductance of 2.5 nS and time constants 1 ms for the rise and 2 ms for the fall of the synaptic response). The activity of $\mathbf{g}_o$ therefore fulfills a gating role with regard to spike propagation to $\mathbf{c}_i$.

Within a minicolumn, every neuron in $\mathbf{g}_i$ connects to every neuron in $\mathbf{g}_o$ through modifiable synapses with weights in $\mathbf{W}_{ig}$, while every neuron in $\mathbf{c}_i$ connects to every neuron in $\mathbf{c}_o$ through modifiable synapses with weights in $\mathbf{W}_{ic}$. The maximum depolarization caused by a connection encoded in $\mathbf{W}_{ig}$ is suprathreshold, while depolarization caused by strengthened connections in $\mathbf{W}_{ic}$ is limited to subthreshold values. Additional depolarization is provided to $\mathbf{c}_o$ by one-to-one connections from neurons in $\mathbf{g}_i$ (with a conductance of 2.5 nS and time constants 1 ms for the rise and 2 ms for the fall of the synaptic response). This provides a gating function for decisions about which action is selected based on convergence. The fan-out of connections within a minicolumn between $\mathbf{g}_i$ and $\mathbf{g}_o$ and between $\mathbf{c}_i$ and $\mathbf{c}_o$ enables the encoding of multiple routes between minicolumns. The following sections will first describe the retrieval process and then describe encoding.

Retrieving behavioral rules in prefrontal cortex

Miller and Cohen propose that the top-down processing in which behavior is guided by internal states or intentions (cognitive control) stems from the active maintenance of patterns of activity in PFC that represent goals and the means to achieve them. They suggest that these patterns provide a bias that guides activity affecting behavior, a gating function, and support their theory with neurobiological, neuroimaging and computational studies (Miller and Cohen, 2001).

In our simulation, associations that form known rules are encoded in PFC. A desire for reward then elicits a spread of activity from the minicolumn representing that reward state (see dashed lines in fig. 3a and left arrows in fig. 3b). The neurons of the $\mathbf{g}_o$ population within that Reward minicolumn spike simultaneously in response to rhythmic input at an 8Hz theta frequency. Those spikes propagate along connections with strengthened synaptic weights in $\mathbf{W}_g$ and produce a spike in the targeted $\mathbf{g}_i$ neurons of minicolumns that immediately preceded the Reward minicolumn in a known rule. Within such a preceding minicolumn (a minicolumn that represents an action) a spike elicited at a neuron in the $\mathbf{g}_i$ population fans out across strengthened connections to neurons in the $\mathbf{g}_o$ population of that minicolumn. Through those connections with strengthened synaptic weights in $\mathbf{W}_{ig}$, suprathreshold depolarization is elicited at the target $\mathbf{g}_o$ neuron. This same process is repeated in other consecutive mini-

[2]An identity matrix.

columns to spread activity through the $\mathbf{g}_i$ and $\mathbf{g}_o$ populations of consecutive action and state minicolumns. As the spread branches out, it follows multiple reverse paths through connections that associate states and actions. Once the spread of activity reaches the minicolumn that represents the current state, the convergence of current state and goal spread allows selection of action. In addition, spikes in $\mathbf{g}_o$ neurons are inhibited ("end-stopping") by the synchronous activity of interneurons (with time constants 1 ms for the rise and 10 ms for the fall of the synaptic response of the input) elicited by input that identifies the current state.

The selection of action is indicated by an interaction of the goal spread with current state. The input that identifies the current state also targets the neurons in the $\mathbf{c}_o$ population of the same current state minicolumn. The excitatory input produces a subthreshold depolarization of $\mathbf{c}_o$ neurons. In addition to this input, the spiking of neurons in the $\mathbf{c}_o$ population is gated by population $\mathbf{g}_i$ activity in the same minicolumn due to the spread of activity from the goal. Those $\mathbf{c}_o$ neurons that receive additional depolarization from spiking neurons in the $\mathbf{g}_i$ population fire.

The present simulation uses only the first step of the forward spread to determine output that controls goal-directed behavior in the task, so the forward gating only has an effect on the $\mathbf{c}_o$ of the minicolumn representing current state. The output of neurons in the $\mathbf{c}_o$ populations of state minicolumns that target action minicolumns is connected to the motor circuitry of the simulation. A spike in $\mathbf{c}_o$ thereby drives motor output of the corresponding action (thick black arrow in figure 3a). A spike in $\mathbf{c}_o$ also causes spiking in interneurons that provide lateral inhibition to the remaining neurons in $\mathbf{c}_o$, so that a clear winner-take-all behavioral response is obtained.

For other applications, the minicolumn model also enables a forward spread of activity for known associations encoded in PFC (see dotted lines in fig. 3a and right arrows in fig. 3b). The spikes that propagate through connections with strengthened synaptic weights in $\mathbf{W}_c$ cause subthreshold depolarization of a $\mathbf{c}_i$ neuron in the associated action minicolumns. Again, forward spread of activity is gated by the spread from the goal, since a neuron in the $\mathbf{c}_i$ population needs additional depolarization from a corresponding neuron in the $\mathbf{g}_o$ population to fire. The spike of a $\mathbf{c}_i$ neuron fans out through connections with strengthened synaptic weights in $\mathbf{W}_{ic}$ to $\mathbf{c}_o$ neurons that are gated by the dependence on activity in $\mathbf{g}_i$ neurons in the same minicolumn.

Figure 3a includes an example of rule retrieval in a rewarded move trial. Neurons that spike as activity spreads are represented by gray circles. The example points out the importance of neuron populations $\mathbf{g}_i$, $\mathbf{g}_o$, $\mathbf{c}_i$ and $\mathbf{c}_o$, in which individual neurons make connections with other minicolumns. As shown in fig. 3a, desire for reward causes all neurons in the $\mathbf{g}_o$ population of the Reward minicolumn to fire. The activity then spreads to associated minicolumns, including Go, NoGo and all sensory input minicolumns. In the same trial, when the Srm stimulus is perceived, the $\mathbf{c}_o$ population of the Srm minicolumn is depolarized. In the Srm minicolumn, the specific depolarized $\mathbf{c}_o$ neuron that corresponds with a spiking neuron of the $\mathbf{g}_i$ population fires, so that activity

spreads forward along a route from minicolumn Srm to minicolumn Go. The firing of the $\mathbf{c}_o$ neuron is used to generate the Go response. An analogous approach would be to use the spikes of a $\mathbf{c}_i$ neuron in the Go minicolumns to generate the Go response. During this process, the $\mathbf{g}_o$ population of the Srm minicolumn is inhibited (end-stopping). Figure 3a shows that the spread of activity from the goal is stopped there.

In the example, spreading activity from the Reward minicolumn involves two different known paths that include the Go minicolumn. One path retrieves the associated items Reward–Go–Srm, the other retrieves the associated items Reward–Go–Surm[3](and a separate path through NoGo retrieves Reward–NoGo–Srnm). Since the spread of activity through different known paths elicits spikes at separate $\mathbf{g}_i$ neurons, they do not interfere with each other. And since the neurons in $\mathbf{c}_i$ and $\mathbf{c}_o$ populations also maintain separate connections with other minicolumns, the activity in $\mathbf{g}_i$ correctly allows the gated forward spread to propagate only on a path from a state receiving current input. Thus, the structure of our model allows mapping through the same action from different states. While retrieval activity spreads forward along known paths to reward, those spikes elicited in the $\mathbf{c}_o$ population of the current state minicolumn that target action minicolumns also trigger the output of PFC. In fig. 3a, the spike propagation through the connection from minicolumn Srm to minicolumn Go is therefore marked as a thick black arrow. This output generates the correct "Go" response, thereby guiding successful goal-directed behavior.

Encoding behavioral rules in prefrontal cortex

The above section described retrieval. This section describes encoding. During encoding, the neuron labeled **a** in the model of a minicolumn fires when input that matches the item represented by the minicolumn is received. For example, when an input spike indicates that a rewarded-move stimulus, Srm, is detected, that input causes neuron $\mathbf{a}(Srm)$ to spike. Here, it is assumed that stimuli activate minicolumn n after minicolumn $n-1$. Encoding is achieved by STDP (Levy and Stewart, 1983; Markram *et al.*, 1997; Bi and Poo, 1998) that corresponds to the long-term potentiation (LTP) of synaptic responses (Bliss and Lømo, 1973; Bliss and Collingridge, 1993). The four steps described below take place sequentially in each encoding cycle.

Reverse associations between minicolumns are encoded in weight matrix $\mathbf{W}_g$ at synapses from $\mathbf{g}_o(n)$ onto $\mathbf{g}_i(n-1)$: A short-term memory (STM) buffer maintains spiking that corresponds with the two most recent inputs to the network of minicolumns. During this reactivation in encoding phases of PFC minicolumns, $\mathbf{a}(n)$ spikes less than 20 ms after $\mathbf{a}(n-1)$. As shown in figure 4a, the neuron $\mathbf{a}(n-1)$ provides subthreshold depolarization to all the

[3]The retrieval of rules resembles the sequence of transitions in a finite state machine (Harel, 1987), and the recurrent connections that lead to two visits of the Go minicolumn in trials initiated by the Surm stimulus are reminiscent of connectionist Elman networks (Elman, 1990, 1991).

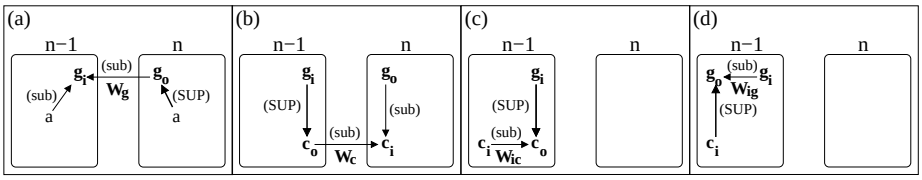


Figure 4: The four steps, (a) to (d), of rule encoding in prefrontal cortex. Rectangles indicate the nth minicolumn that activates and the one that precedes it at $n-1$. Thin arrows indicate connections between neuron populations (lower case letters within the rectangles) that may result in subthreshold postsynaptic depolarization (marked sub), while thick arrows indicate connections that may result in suprathreshold depolarization (marked SUP). The matrix of synaptic weights that is updated in an encoding step is indicated by $\mathbf{W}_g$, $\mathbf{W}_c$, $\mathbf{W}_{ic}$ and $\mathbf{W}_{ig}$ below an arrow that represents connections with synapses that are being modified.

neurons of the $\mathbf{g}_i$ population in minicolumn $n-1$. And all neurons in the $\mathbf{g}_o$ population in minicolumn n receive suprathreshold depolarization through synapses from $\mathbf{a}(n)$. As the neurons in $\mathbf{g}_o(n)$ spike, that neuron in the $\mathbf{g}_i$ population of minicolumn $n-1$ which is connected to a neuron in $\mathbf{g}_o(n)$ receives subthreshold depolarization, due to the initial value of synaptic strengths in weight matrix $\mathbf{W}_g$. The neuron in $\mathbf{g}_i(n-1)$ that receives input from both $\mathbf{a}(n-1)$ and $\mathbf{g}_o(n)$ spikes a few milliseconds later than the presynaptic neuron in $\mathbf{g}_o(n)$, so that STDP is elicited. Thus, the amplitude of the corresponding synaptic response is increased in $\mathbf{W}_g$. After several repetitions in the STM buffer, encoding establishes a suprathreshold connection between $\mathbf{g}_o(n)$ and $\mathbf{g}_i(n-1)$ (Fig. 4a).

Forward associations between minicolumns are encoded in weight matrix $\mathbf{W}_c$ at synapses from $\mathbf{c}_o(n-1)$ onto $\mathbf{c}_i(n)$: Rhythmic input modulates the membrane potential of neurons in $\mathbf{c}_o$. During the encoding phase, the rhythmic depolarization of neurons in $\mathbf{c}_o(n-1)$ is such that excitatory input through one-to-one connections from $\mathbf{g}_i(n-1)$ in the same minicolumn causes postsynaptic spiking. The spiking in $\mathbf{g}_i(n-1)$ that is described in the encoding step above therefore drives spiking in $\mathbf{c}_o(n-1)$, as shown in figure 4b. The neurons in $\mathbf{c}_i(n)$ receive subthreshold (gating) depolarization through one-to-one input from neurons in $\mathbf{g}_o(n)$. In the presence of rhythmic depolarization as above and given small initial values in $\mathbf{W}_c$, the neuron in $\mathbf{c}_i(n)$ that is connected to a neuron in the $\mathbf{c}_o$ population of minicolumn $n-1$ spikes due to the combined subthreshold inputs from both $\mathbf{g}_o(n)$ and $\mathbf{c}_o(n-1)$. Again, STDP is elicited, since the postsynaptic neuron in $\mathbf{c}_i(n)$ spikes a few milliseconds after it receives input from the presynaptic neuron in $\mathbf{c}_o(n-1)$. After repetition, a subthreshold connection is established between $\mathbf{c}_o(n-1)$ and $\mathbf{c}_i(n)$, which propagates spikes if input is received from the corresponding neuron in the gating $\mathbf{g}_o(n)$ population, even when rhythmic depolarization is absent in retrieval phases.

Rules that associate preceding with possible ensuing activity are encoded within a minicolumn by the weight matrix $\mathbf{W}_{ic}$ at synapses from $\mathbf{c}_i(n-1)$ onto $\mathbf{c}_o(n-1)$: During encoding, the activity of the $\mathbf{c}_i$ population is driven by a STM buffer that maintains the activity of $\mathbf{c}_i$ populations of the two[4]most recently active minicolumns. As figure 4c shows, neurons in $\mathbf{c}_i(n-1)$ spike several milliseconds before spiking of neurons in $\mathbf{c}_o(n-1)$ is driven by corresponding spikes in population $\mathbf{g}_i(n-1)$ (with a synaptic conductance of 6.0 nS), as described above. STDP is elicited and repetition increases synaptic strengths in $\mathbf{W}_{ic}$ from initial values near zero to subthreshold amplitudes.

Associations that enable the spread of activity from the representation of a goal are encoded by the weight matrix $\mathbf{W}_{ig}$ at synapses from $\mathbf{g}_i(n-1)$ onto $\mathbf{g}_o(n-1)$ within a minicolumn: During encoding, spiking in a sub-population of $\mathbf{g}_o$ that is identified as $\mathbf{g}_o^{specific}$ in minicolumn $n-1$ is driven by input from $\mathbf{c}_i(n-1)$, as shown in figure 4d. A delay in the synaptic transmission from $\mathbf{c}_i(n-1)$ insures that the spikes at $\mathbf{g}_o^{specific}$ occur several milliseconds after spiking in $\mathbf{g}_i(n-1)$. At connections that repeatedly experience STDP due to this sequence of spiking, the synaptic strength in $\mathbf{W}_{ig}$ is increased from near zero to suprathreshold values.

The population $\mathbf{g}_o^{specific}$ and a population of neurons known as $\mathbf{g}_o^{diffuse}$ provide separate encoding functions, but as shown in figure 5, they act together as $\mathbf{g}_o$ during retrieval. In the retrieval mode, transmission from $\mathbf{c}_i(n-1)$ to neurons in $\mathbf{g}_o^{specific}$ is suppressed, while input from $\mathbf{g}_i$ is received through connections with synaptic strengths $\mathbf{W}_{ig}$. The pattern of spikes in $\mathbf{g}_i$ and suprathreshold synaptic strengths established in $\mathbf{W}_{ig}$ therefore determines retrieval spiking in $\mathbf{g}_o^{specific}$. That spiking is duplicated in $\mathbf{g}_o^{diffuse}$ during retrieval, since transmission is then enabled through strong one-to-one input connections from $\mathbf{g}_o^{specific}$. By contrast, all neurons in the $\mathbf{g}_o^{diffuse}$ population of a minicolumn are driven by $\mathbf{a}$ during encoding modes, so that they provide the diffuse output of $\mathbf{g}_o(n)$ that is used to encode $\mathbf{W}_g$ and $\mathbf{W}_c$, as described above. In this manner, the two sub-populations of $\mathbf{g}_o$ can spike in separate patterns that satisfy the different needs of encoding protocols for synapses within a minicolumn ($\mathbf{W}_{ig}$) and between minicolumns ($\mathbf{W}_g$ and $\mathbf{W}_c$). This function could alternatively be obtained by very tightly regulating the activity of $\mathbf{g}_o$ at different phases.

Short-term memory based on persistent spiking enabled spike timing dependent potentiation to encode associations

As described, encoding in our model of PFC depends on STDP in $\mathbf{W}_g$, $\mathbf{W}_c$, $\mathbf{W}_{ig}$ and $\mathbf{W}_{ic}$, and on the buffered activity of populations $\mathbf{a}$ and $\mathbf{c}_i$. A Hebbian model of STDP that is based on the long-term potentiation observed at many synapses requires multiple instances in which presynaptic spiking precedes postsynaptic spiking by less than 40 ms (Levy and Stewart, 1983; Markram *et al.*, 1997;

[4]The buffer holds two items so that the buffered activity $\mathbf{c}_i(n)$ can replace $\mathbf{c}_i(n-1)$ as the memory of preceding activity in $\mathbf{c}_i$ when the next association with minicolumn $n+1$ is encoded.

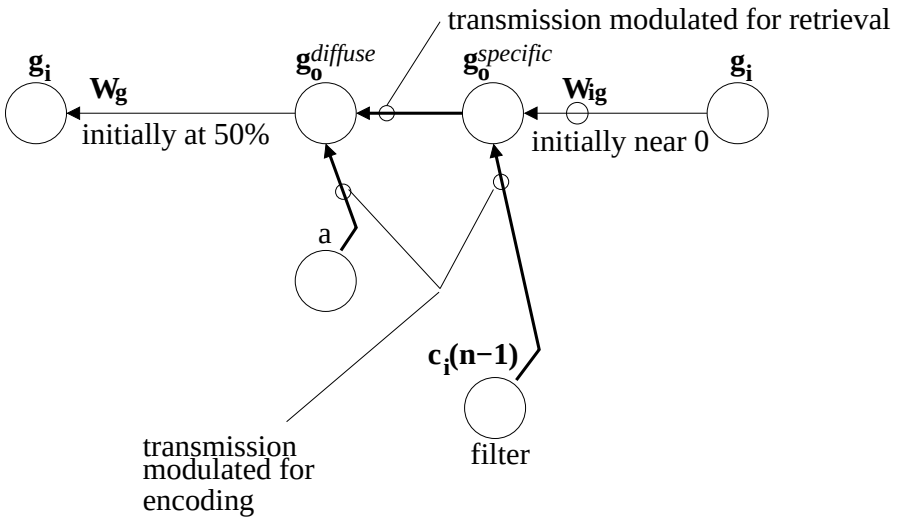


Figure 5: Subdivision of the $\mathbf{g}_o$ population into functional $\mathbf{g}_o^{diffuse}$ and $\mathbf{g}_o^{specific}$ neuron populations. Neurons in $\mathbf{g}_o^{diffuse}$ all spike in response to activity in $\mathbf{a}$, while the spiking of neurons in $\mathbf{g}_o^{specific}$ reflects the specific patterns of spikes received through one-to-one connections from $\mathbf{c}_i(n-1)$. Spiking in the filter population relies on rhythmic depolarization, so that only $\mathbf{c}_i(n-1)$ activity in the short-term memory buffer of $\mathbf{c}_i$ drives $\mathbf{g}_o^{specific}$ during encoding. This way, the strength of unique connections in $\mathbf{W}_g$ to other minicolumns is encoded separately from the encoding of $\mathbf{W}_{ig}$ in accordance with the mapping of a pattern of spikes in $\mathbf{g}_i$ to a pattern of spikes in $\mathbf{g}_o^{specific}$. During retrieval, strong one-to-one connections from $\mathbf{g}_o^{specific}$ to $\mathbf{g}_o^{diffuse}$ drive the entire $\mathbf{g}_o$ population as one.

Bi and Poo, 1998), while input to PFC may arrive with arbitrary large time intervals. As mentioned previously, we therefore presuppose that firing patterns may be reactivated in a persistent manner by intrinsic neuronal mechanisms, such as after-depolarization (ADP) of membrane potential (fig. 6A), caused by calcium sensitive cation currents that are induced by muscarinic receptor activation (Andrade, 1991; Klink and Alonso, 1997a). We also presuppose that a common brain rhythm may produce oscillatory modulation in different regions that provides synchronization of activity. The reactivation of firing patterns by ADP in one population of neurons at specific phases of the brain rhythm can thereby reliably provide input to other populations in PFC where STDP can occur in an encoding mode (Fig. 6B). Using rhythmic modulation and ADP, we provide short-term memory (STM) in a manner similar to the STM model first proposed by Lisman and Idiart (Lisman and Idiart, 1995; Jensen and Lisman, 1996). Recurrent inhibition within such a buffer separates the reactivation of sequential items to maintain their order. The STM may reside in PFC or may be provided by input from entorhinal cortex.

The membrane potentials of three neurons of a STM buffer are plotted in figure 6B. In the hippocampus, regular activity originating in the septum (Brazhnik and Fox, 1999) is believed to cause 8 Hz oscillations of the membrane potential by modulating the GABAergic inhibition of pyramidal cells via networks of interneurons (Alonso *et al.*, 1987; Stewart and Fox, 1990). A similar mechanism appears to cause theta rhythm oscillations in limbic cortices due to rhythmic activity of basal forebrain neurons Manns *et al.* (2000). Those oscillations define two functional phases of the buffer neurons. We call the phase interval of greatest rhythmic depolarization the reactivation phase of STM and the remaining interval the input phase of STM. The plots show that spiking produced by afferent activity during the input phase of the buffer is reactivated by the ADP during subsequent repetition phases. The duration of the rise of ADP matches the period of oscillation. This means that the ADP of the earliest neuron to spike in one cycle allows that neuron to reach threshold first in the following cycle. The order of spikes is maintained during reactivation in STM. As spikes caused by the buffer occur in pre- and postsynaptic neurons of modifiable connections in PFC, an asymmetric function of spike-timing dependent potentiation takes into account the order of spikes. This ensures that STDP is elicited in specific connections so that a direction of causality is inferred during rule learning. Furthermore, the separation of consecutive spikes is maintained in STM by recurrent inhibition that is caused by the activation of an interneuronal network (Bragin *et al.*, 1995) each time a buffer neuron spikes.

In the absence of input, the contents of a STM buffer decay gradually, due to noise and a slow-AHP. But when a full buffer receives new input, such as when rule learning involves a long sequence of states and actions, the earliest item in the buffer needs to retire so that the new item is maintained. The item replacement must also avoid changing the order of items. To achieve this, we propose that the appearance of a new item leads to inhibition at a specific phase of the rhythmic oscillation (see dashed box in figure 6C). Inhibition at that specific phase suppresses the reactivation of the first item (Koene *et al.*, 2003)

until its ADP has subsided, as shown in figure 6C. The new item, represented by action potentials in the plot of the membrane potential of the third cell, assumes the last position in the sequence of reactivation.

Each neuron in a STM buffer projects output to a corresponding target neuron in $\mathbf{a}$ or $\mathbf{c}_i$. Current and preceding activity are therefore available for encoding, as shown in figure 7 for the membrane potential of $\mathbf{a}$ neurons throughout the network. The activity in $\mathbf{a}$ corresponds to current and preceding input, as pairs of state and action spikes are received in PFC during the seven simulated encoding trials of rule learning (fig. 2).

2 RESULTS

The network described above effectively encoded the different rules of the task and showed effective behavioral performance when tested with different stimuli, generating a Go response to Srm, a NoGo response to Srnm and a Go response to Surm stimuli. This behavior was guided by spiking activity that matches the data obtained by Schultz *et al.* (Schultz *et al.*, 2000).

In the seven training trials (fig. 2), the necessary associations for stimulus gated selection of action were encoded with strengthening of connections using STDP at synapses in $\mathbf{W}_g$, $\mathbf{W}_c$, $\mathbf{W}_{ig}$ and $\mathbf{W}_{ic}$. Six trials were used to test performance with all possible initial stimuli. For these trials, the spike trains that represent the sensation of the initial stimulus were provided as input and the model generated motor commands that lead to behavioral responses and the sensation of reward received were observed. The network showed the correct behavior in the task. The correct action followed each initial state during tests of task performance. Inspection of individual neuronal responses reveals that the three main types of responses observed by Schultz *et al.* were also found in the present simulations: (1) neurons that respond selectively to a trial-specific initial stimulus, (2) neurons that respond prior to reward in a specific trial and may indicate a chosen course of action, and (3) neurons that respond selectively to predicted and obtained reward. In addition to these, several more specialized responses were observed, providing predictions of the model.

During performance of the operant task, a desire for reward begins at the onset of every trial in the form of regular suprathreshold input to all neurons of the $\mathbf{g}_o$ population of the minicolumn that represents the goal. When trial input stimuli appear in different trials they are maintained as persistent spikes of buffer neurons that cause the spiking of $\mathbf{a}(Srm)$, $\mathbf{a}(Srnm)$ and $\mathbf{a}(Surm)$ in fig. 8). These input stimuli also provide subthreshold input to the $\mathbf{c}_o$ population of the minicolumn that represents the current state. Converging with the spread of activity from the goal minicolumn, spiking $\mathbf{c}_o$ neurons drive goal-directed behavior, resulting in the generation of output which in turn causes proprioceptive feedback of the correct action in each sequence in figure 8, as well as the perception of reward received.

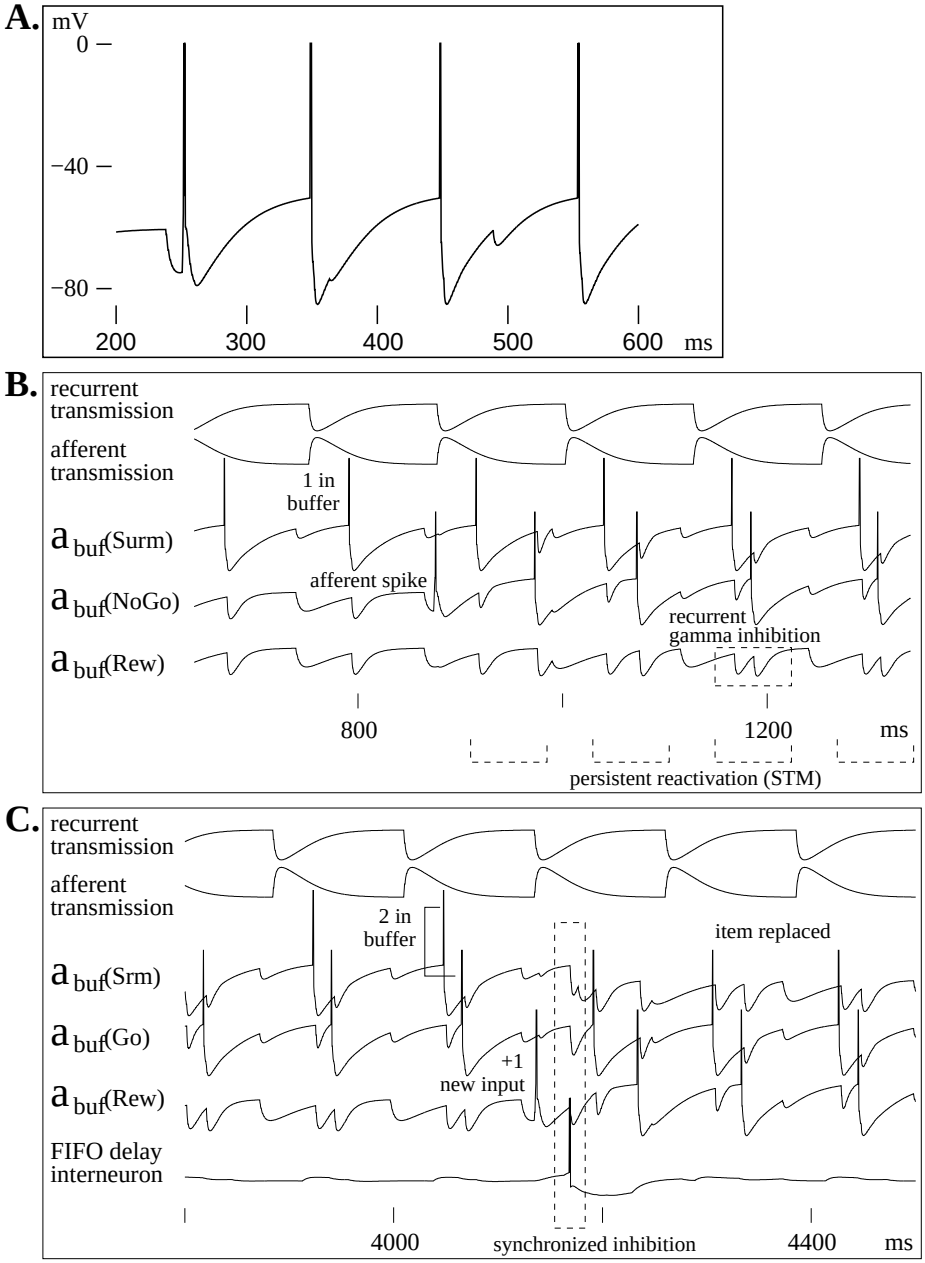


Figure 6: **A.** After-depolarization (ADP) supports persistent firing. Each spike causes initial after-hyperpolarization (AHP) of the membrane, which is followed by a slow ADP. That depolarization can ultimately lead to another spike. **B.** A buffer based on persistent firing receives afferent input during one phase of its rhythmic cycle and reactivates items (separated by competitive inhibition) in order in each cycle. **C.** First-in-first-out (FIFO) item replacement. In a full buffer, afferent input plus retrieval activity elicit inhibition synchronized to suppress reactivation of the first item$_{19}$The input is added at the end of the sequence.

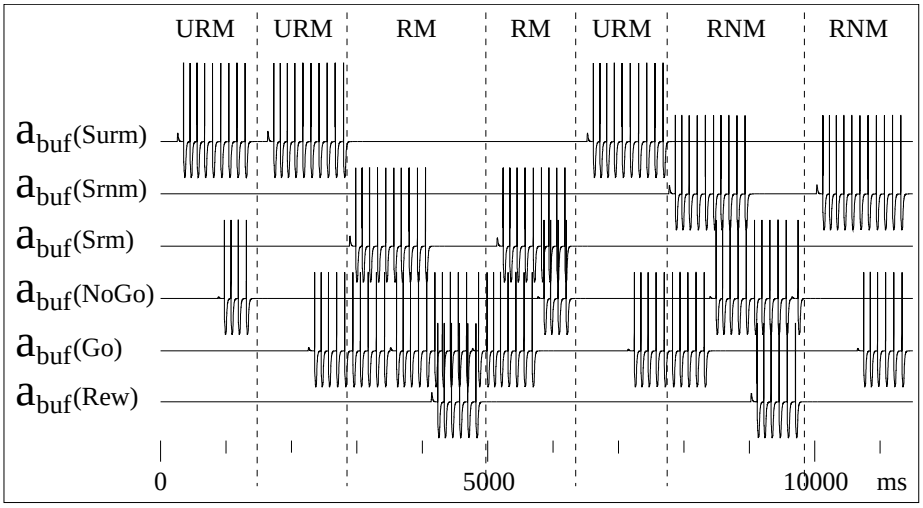


Figure 7: The membrane potential of neurons in the **a** population, responding to input from the short-term memory buffer during the training stage (encoding) of the visual discrimination task in a sequence of six trials.

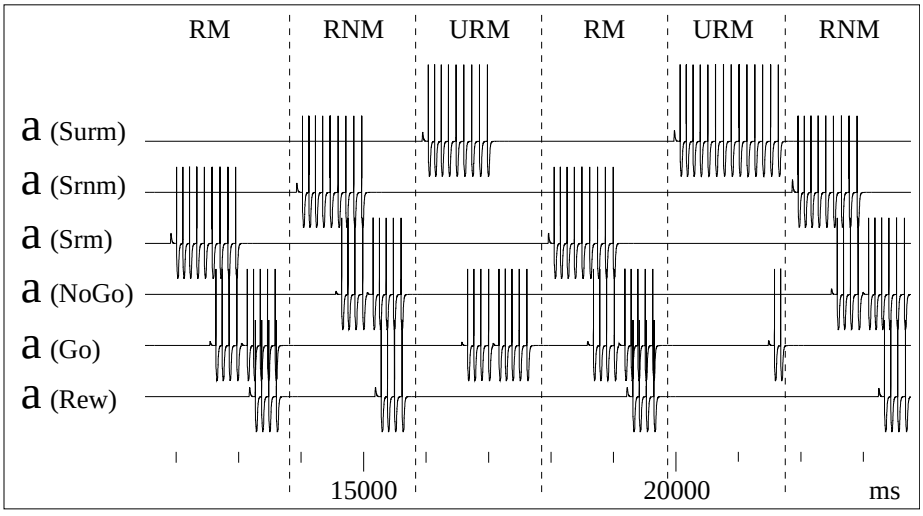


Figure 8: The membrane potential of neurons in the **a** population, responding to input from the short-term memory (STM) buffer during behavioral performance (retrieval) of the visual discrimination task in a sequence of six trials. A change of context between rewarded movement (RM), rewarded non-movement (RNM) and unrewarded movement (URM) trials causes the STM buffer to clear during the those intervals.

2.1 Activity underlying selective responses in the model

Membrane potentials of those neurons within a minicolumn that are involved in the choice of action demonstrate the decision process that is based on a forward spread of activity that is gated by the spread of activity from the goal. This is shown in figure 9, in which membrane potentials of relevant $\mathbf{a}$, $\mathbf{g}_i$ and $\mathbf{c}_o$ neurons in the minicolumn that represent the Surm instruction state are plotted during an interval within an Surm trial (the convergence looks the same for the Srm example in fig. 3). The plots show that neurons in the $\mathbf{c}_o$ population of that minicolumn experience subthreshold depolarization due to current state input from $\mathbf{a}$. This contribution is joined by converging input from a specific neuron in the $\mathbf{g}_i$ population that spikes due to the spread of activity from the minicolumn that represents the goal (dashed arrows in figure. 3). When the inputs converge a neuron of the $\mathbf{c}_o$ population fires (bottom of fig 9). Activity in $\mathbf{c}_o$ was gated by activity in $\mathbf{g}_i$, and recurrent inhibition assured that only the first spike in $\mathbf{c}_o$ led to a behavioral response. The chosen behavior was determined by the minicolumn that was targeted by that spike, in this example a Go motor command for the simulated task environment.

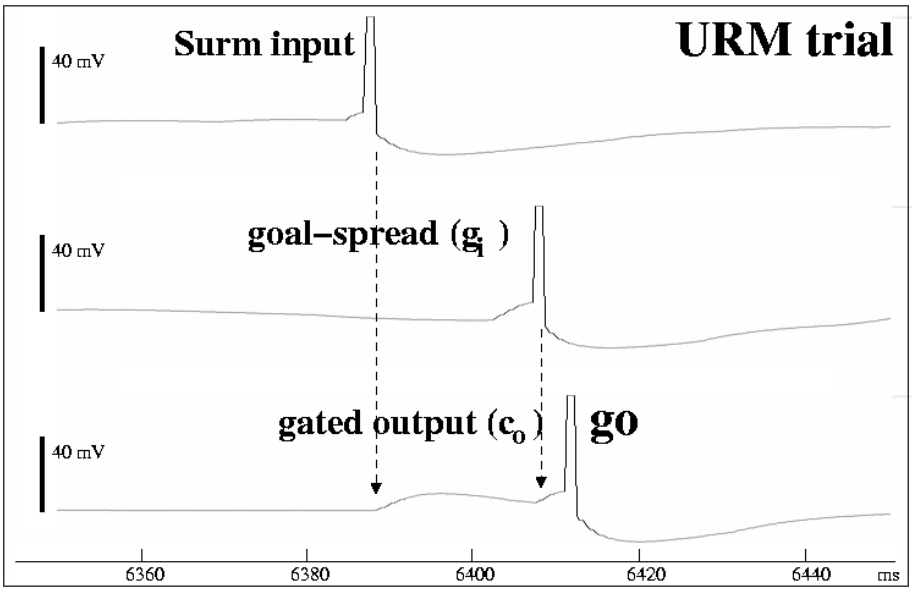


Figure 9: Selected membrane potentials during converging forward spread and spread from the goal in a unrewarded move (Surm) trial. The forward spread is initiated in state Surm, as represented by the action potential of the $\mathbf{a}$ neuron (top). The spread of activity from the goal reaches the Surm minicolumn when an action potential appears at a specific $\mathbf{g}_i$ neuron (middle). A resulting action potential that directs Go action appears at a specific $\mathbf{c}_o$ neuron of the same minicolumn (bottom).

For the six test trials, the spike trains that represent the sensation of the initial stimulus, motor commands that lead to behavioral responses and the sensation of reward received are shown in figure 10. The spike trains show that Srm stimuli were followed by Go responses and reward, Srnm was followed by NoGo responses and reward and that Go action responses followed Surm stimuli and led to subsequent rewarded trials. The network can perform correctly regardless of the order of presented test stimuli.

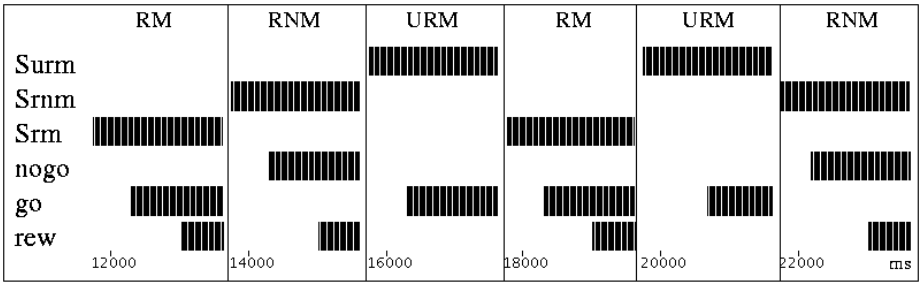


Figure 10: The guided output (Go or NoGo) of the model in response to test stimuli (Srm, Srnm and Surm). Spike trains produced in the motor circuitry of the simulated operant task environment during trials that test task performance (testing retrieval). The spike trains represent sensory stimuli received during the trials (separated by vertical lines), as well as behavioral Go and NoGo motor responses and the sensation of reward received.

Schultz *et al.* plotted the recorded spikes of three orbitofrontal neurons during many rewarded move (Srm) and unrewarded move (Surm) trials. We compare our simulation results with those of the experiment by Schultz *et al.* by displaying results for the three main categories of neuronal responses described by Schultz *et al.* side by side in figure 11. These plots show spikes in individual trials (short vertical lines) aligned to specific parts of the task.

As in the W.Schultz *et al.* results, our results showed that individual neurons activate specifically when one of the three cue stimuli is perceived. In our model, this is caused by the current state response of the **a** population (fig. 11A,D). We also found individual neurons that activate for a chosen behavioral response. This activity results when neurons of the $\mathbf{c}_o$ population in the current state minicolumn receive gating activity from $\mathbf{g}_i$ neurons due to the spread of activity from the goal minicolumn (fig. 11B,E). We also found neurons that activate specifically when reward is received. This activity is caused by the current state activation of the **a** neuron in the goal minicolumn in our model (fig. 11C,F).

As in the Schultz *et al.* data, there is spiking in E during Srm and Surm trials, but the spike rate is higher during the Go action in Srm. Both the data and the output of our model show a quantitative difference in the amount of firing between Srm and Surm trials before reward is received. In our model, this is explained because $\mathbf{c}_o(Srm{\rightarrow}Go)$ is activated in encoding phases in both trials when $\mathbf{a}(Go)$ is maintained by the STM buffer, since strengthened connec-

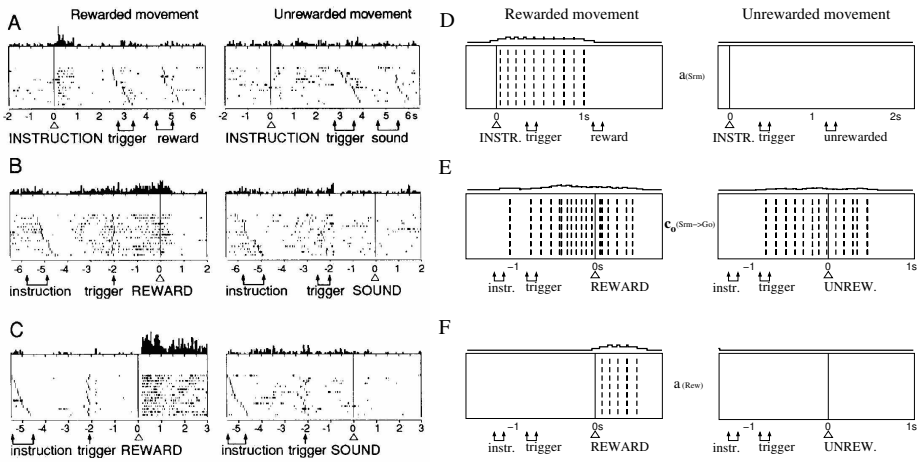


Figure 11: A side-by-side comparison of neuronal activity recorded by W.Schultz *et al.* (A-C, figure reproduced from W.Schultz *et al.*, Cerebral Cortex, 2000) and that produced by our simulation of PFC minicolumns (D-F). Figures in A to C display spikes of three different orbitofrontal neurons. For each, the activity in rewarded and unrewarded movement trials is shown side by side. And every row within the borders of a graph represents the activity of that neuron during a separate trial. The time course of the data and of the model output are aligned to specific task events. Labels below a horizontal time axis indicate stages of the operant task: instruction stimulus, action trigger, reward. Above each graph, a histogram shows the sum of spikes in each bin of time, i.e. in a corresponding column over all trials. Figure D shows the spike responses of an **a** population neuron in the RM (rewarded move) minicolumn aligned to instruction. Figure E shows a $\mathbf{c}_o$ population neuron in the RM minicolumn with output connections to the GO minicolumn aligned to reward. Figure F shows an **a** population neuron in the REW (Reward) minicolumn. Again, the spikes (short vertical lines) of each neuron are shown side by side in both rewarded and unrewarded movement trials. Rows within each figure show the results of separate simulation runs, while the cumulative spike rate is plotted above each figure by counting the number of spikes within an interval around t. The three neurons in D-F replicate the experimental results by Schultz *et al.* in the corresponding categories A-C.

tions from $\mathbf{g}_o(Go{\rightarrow}Srm)$ to $\mathbf{g}_i(Srm{\leftarrow}Go)$ propagate the activity. Additionally, $\mathbf{c}_o(Srm{\rightarrow}Go)$ is activated specifically in the Srm trial when the goal spread causes spiking in the gating $\mathbf{g}_i(Srm{\leftarrow}Go)$ neuron, while current state input depolarizes the $\mathbf{c}_o(Srm)$ population. The appearance of similar activity at the trigger time during URM trials in B suggests that the activity is not merely background noise and supports the possible explanation provided by our model.

A smaller temporal overlap of activity similar to that in the Schultz *et al.* results is achieved if the intervals between instruction stimulus, action trigger and reward delivery are increased in the model to match the data, for a trial length of six to eight seconds instead of 1500 ms in the simulation. The shorter intervals in the model significantly reduced the time needed to compute each simulation run without affecting resulting behavior.

2.2 Some neurons in prefrontal cortex are active in multiple behaviors

In addition to the results above, we found that some neurons in the simulation activate selectively for a specific phase of two different trials. As shown in figure 12A, the $\mathbf{a}(Go)$ neuron in the minicolumn that represents a movement response spikes in rewarded movement and unrewarded movement trials. Similarly, the $\mathbf{a}(Rew)$ neuron in the minicolumn that represents the perception of reward spikes in rewarded movement and rewarded non-movement trials.

In figure 12B, we show that specific neurons in the $\mathbf{g}_i$ and $\mathbf{g}_o$ populations of minicolumns that are involved in the retrieval of associations with a goal generated a spike in every trial of that specific task. The neurons that activate throughout each trial correspond to those involved in the learned associations for the spread of activity from the goal during retrieval, as shown in figure 3. Thus, even neurons with very extensive response properties are important for performance of this task. Activity of the $\mathbf{a}$ population in the current state produces end-stopping of activity through the $\mathbf{g}_o$ population in the same minicolumn. Therefore, the onset of a rewarded move (RM) trial produces end-stopping at $\mathbf{g}_o(Srm)$ cells, but due to the associations from Reward to Srnm via NoGo and from Srnm to Surm via Go a neuron in $\mathbf{g}_i(Surm)$ also spikes during that trial. Similarly, $\mathbf{g}_i(Surm)$ spikes during rewarded non-movement trials due to the alternate path for the spread of retrieval activity from the goal via the Srm minicolumn. Thus, we predict a correlation of neuronal firing during Surm and Srm trials (strong Go involvement in both), and a lesser correlation of neuronal firing during Surm and Srnm trials, as shown in rows 1 and 3 of figure 12B.

Activity in figure 12C demonstrates the end-stopping function proposed in the minicolumn model. During rewarded movement trials, the neuron $\mathbf{g}_o^{diffuse}(Srm{\leftarrow}Go)$ is active until reward is received. As soon as the perception of reward becomes the current state of the PFC network, the neuron is no longer active. This is not the case in rewarded non-movement and unrewarded movement trials. In rewarded movement (RM) trials, end-stopping prevents the spread of activity from the goal to the $\mathbf{g}_o$ population of the Srm minicolumn. During these trials, the $\mathbf{g}_o^{diffuse}(Srm{\leftarrow}Go)$ neuron is active in encoding modes of each rhythmic

cycle while maintained in the STM buffer. When reward is perceived, the Go-Reward pair replaces the Srm-Go pair in the buffer, as seen in the bottom two rows of figure 12B. End-stopping appears in Srm (RM) trials and Srnm (RNM) trials, but not Surm (URM) trials, since two associative paths can be taken from the goal minicolumn to the Surm minicolumn.

Schultz *et al.* point out that some neurons activated less selectively, namely in a manner that was selective for the instruction cue regardless of trial type and expected reward. Similarly, our simulation shows that a neuron of the $\mathbf{c}_i(Srm{\rightarrow}Go)$ population in the Go minicolumn that receives input from the Srm minicolumn exhibits retrieval spikes in both Srm and Surm trials during instruction activity in the Srm or Surm minicolumns. Those retrieval spikes disappear once the Go minicolumn receives proprioceptive input about a key press movement in the environment and spikes begin to occur in the encoding phase of the theta modulated network. This produces a 180 degree phase shift of firing at the time of the movement generation. The Go minicolumn $\mathbf{c}_i(Surm{\rightarrow}Go)$ neuron that receives input from the Surm minicolumn exhibits the same transition of spiking from the retrieval to the encoding phase, but its retrieval spiking is more selective and appears only during an Surm trial, since no sequence exists that involves the Surm minicolumn in other trials.

Schultz et al. provide a quantitative assessment of the trial and phase selective responses recorded. Of 505 neural responses identified at recording sites, 188 exhibited task related activity: 99 responses showed selective activity at the instruction phase of trials. Of those, 63 reflected the type of reinforcer or trial (38 active during RM, RNM or both trial types, 22 active only during URM trials and 3 active during RM and URM trials). 51 responses showed selective activity at the trial phase preceding reward (41 during both RM and RNM trials, 6 during RM or RNM trials, and 4 during URM trials). 67 responses showed selective activity at the reinforcer delivery phase of trials (62 during both RM and RNM trials, 2 during only RM trials and 3 during URM trials).

Before comparison of these numbers with the model, some caveats should be raised. The small sample sizes in terms of the number of sites recorded by Schultz et al. and the number of neurons simulated in the model is too small to allow statistical comparison. Also, the number of selective model responses in a specific category depends on the arbitrary number of neurons chosen as a cell assembly within a population of neurons in each minicolumn. When the model is minimized so that individual functions of the minicolumn are performed by the smallest number of neurons then the following quantitative assessment of responses was obtained.

In the simulation, the neural circuitry of the model prefrontal minicolumns consisted of 328 neurons (excluding neurons that form short-term buffers and circuitry to process prefrontal input and output). From those neurons, 169 task related responses were recorded: 37 responses showed selective activity at the instruction phase of trials. Of those, 34 reflected the type of reinforcer or trial (21 active during RM or RNM trials, 10 active only during URM trials and 3

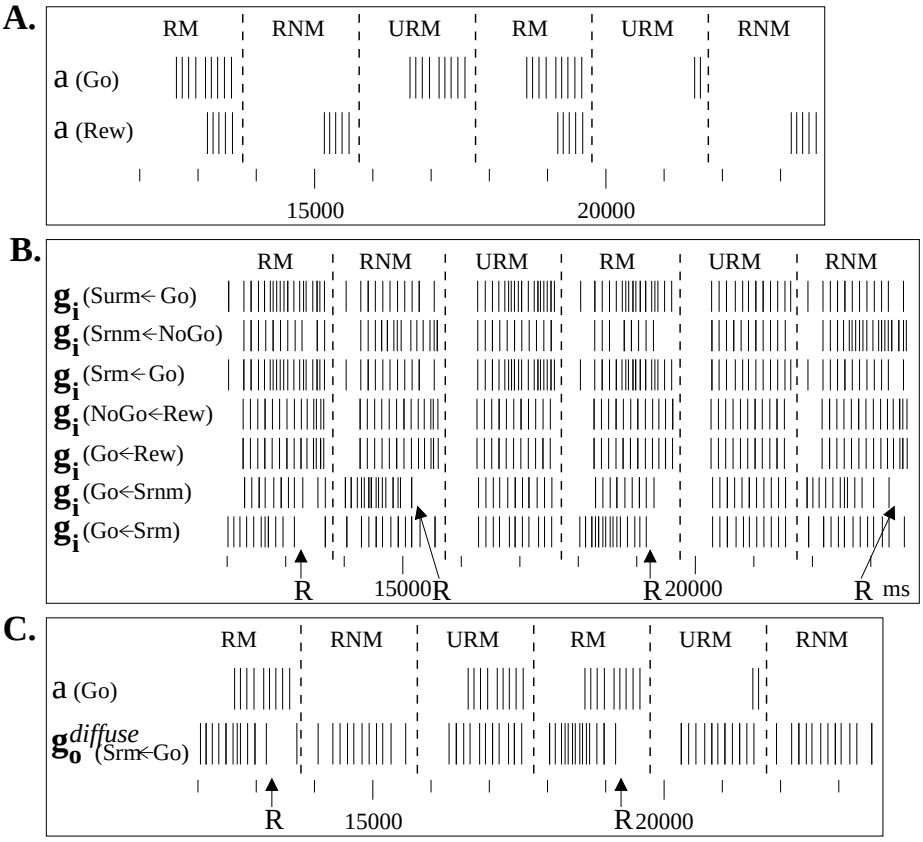


Figure 12: **A.** Spike activity of **a** neurons in Go and Reward minicolumns during performance trials. Neurons $\mathbf{a}(Go)$ that predict a movement response are active in rewarded movement and unrewarded movement trials. Neurons $\mathbf{a}(Rew)$ that spike when reward is received are active in rewarded movement and rewarded non-movement trials **B.** Retrieval activity in the simulation shows that specific $\mathbf{g}_i$ and $\mathbf{g}_o$ population neurons spike in all trials. Regular spike trains span trials for each of the neurons shown. In our example, 24 neurons in $\mathbf{g}_i$ and $\mathbf{g}_o$ populations were found to spike regularly during retrieval in all trials of the performance stage of the specific task, about 7% of the total of 328 neurons involved in retrieval functions. **C.** The $\mathbf{g}_o^{diffuse}(Srm{\leftarrow}Go)$ neuron of the Srm minicolumn shows the end-stopping function. The neuron spikes throughout trials due to the spread of activity from the goal minicolumn, but in rewarded movement (RM) trials spiking stops as soon as reward is received (indicated by arrows with the label "R"). The overlap between spike trains of the $\mathbf{a}(Go)$ neuron and the $\mathbf{g}_o^{diffuse}(Srm{\leftarrow}Go)$ neuron shows the period during which both Srm and Go minicolumn activity are maintained in a STM buffer for encoding.

active during RM and URM trials). 75 responses showed selective activity at the trial phase preceding reward (40 during RM or RNM or both trial types, 11 during only URM trials, 17 during RM and URM trials, and 7 unselective for trial type). 57 responses showed selective activity at the reinforcer delivery phase of trials (20 during both RM and RNM trials, 14 during only RM trials, 21 during RNM trials, and 2 unselective for trial type).

These results support a correlation during the instruction phase between RM and RNM trials seen in both data and model. The absence of a correlation between URM and RNM during the trial phase preceding reward is also consistent with the data. The number of responses for both RM and URM trials is rather higher than the data, as is the response activity for only RNM trials. Both differences may reflect a difference in the model or merely statistical variability.

3 DISCUSSION

Our model replicates goal-directed behavior in a visual discrimination task based on a hypothesis about the functional connectivity of prefrontal cortex circuits (Hasselmo, 2005). Behavioral responses and reward associations to visual cues are encoded in synaptic strengths between neuronal networks representing cortical minicolumns. The goal-directed behavior is retrieved by means of a converging spread of activity from a representation of desired reward and the spread of activity from the current state. Our results specifically replicate the qualitative findings by Schultz *et al.* (Schultz *et al.*, 2000) in terms of individual neuronal responses, while suggesting a possible neural mechanism for learning and retrieval. We use the model to propose explanations for the selective responses of individual neurons in orbitofrontal cortex during goal-directed behavior.

The model provides a framework for the context/stimulus dependent change in action selection, as proposed by Miller and Cohen (Miller and Cohen, 2001). In particular, it provides a spiking neuron implementation of context effects similar to those of Cohen and Servan-Schreiber (Cohen and Servan-Schreiber, 1992). We show how populations of spiking neurons could interact to allow selection of specific actions based on the context of specific sensory input (states) and the desire for reward. Because activity in a specific minicolumn (Fuster, 2000) that represents such a state or action may play a role in different contexts that require its association with different state-action-state transitions, we presuppose separate populations of neurons within a minicolumn for input from and output to other minicolumns (Hasselmo, 2005). For example, the Go and Reward minicolumns in the experimental task fulfill such multiple roles, as shown in figures 3 and 12A.

We show what functional role the individual neurons in these populations could play in the performance of the task by replicating essential features of the Schultz *et al.* experiment. We used similar learning and retrieval protocols and

replicated individual neuronal responses that are selective for a specific state in a specific trial (see figure 11). These selective responses may be understood in the context of a neuron's function in the minicolumn model.

In addition to these explanations, the model generates predictions for this task about what other types of responses should appear in the prefrontal cortex, including neuronal responses which would look rather complex and might therefore not normally be classified. One set of complex responses is shown in figure 12B. The model predicts that some neurons will spike throughout all trials of a goal-directed task, not just for a specific state, due to the spreading activity from a goal representation. And if encoding and retrieval alternate continuously as modeled then such responses that are indicative of spreading activity should be recorded during stages of novel learning as well as task performance.

Our results also propose that end-stopping implemented in the retrieval function of the model may be detected as shown in figure 12C. Evidence that supports possible end-stopping of spreading activity is provided by the termination of recorded spikes in Schultz *et al.* (Schultz *et al.*, 2000), where neuronal activity that is selective for Srm or Srnm instruction stimuli and for action preceding reward terminates as soon as reward is received.

Predictions of the model suggest experiments that test the validity of two of its central tenets: Convergence of activity through representations that may be associated in multiple ways (Sutton and Barto, 1981), and the need for a short-term buffer.

The structure of the model uses a progressive backward spread of activity from the goal. This suggests an experiment that could test this feature, in which associations are formed sequentially between states and actions leading to a particular goal. Imagine an operant task, in which specific sequences of lever presses result in reward. For example, pressing levers in the sequence A-B-C should result in reward. If the levers are pressed randomly, eventually the correct sequence will occur, in a learning paradigm analogous to the one used in experiments by Terrace *et al.* (Terrace *et al.*, 2003). In the model, this will initially lead to an association between the final action "press C" and reward (note that this action involves being at a specific state — in front of lever C — and generating the action "press"). Upon further accidental production of the sequence, it will lead to association of "press B, then press C" with reward, and finally "press A, press B, press C" with reward. The activity of the $\mathbf{g}_i$ and $\mathbf{g}_o$ neurons in the model would initially show activity only for reward, then would show a persistent increase when the association is first formed with "press C", followed by increases in separate populations when the association is formed for "press B" and finally for "press A". Thus, the overall population of neurons firing during the task would show a progressive increase as the specific sequence is learned.

During encoding, our model depends on the function of STM buffers, and data by Andrade shows sustained currents that may support such a function in PFC (Andrade, 1991). However, those buffers need not reside in PFC. A plausible alternative source of buffered perceptual spike patterns is in the entorhinal cortex, in which neurons that exhibit intrinsic persistent spiking have

been found (Klink and Alonso, 1997b). In either case, it is possible that STM function may be disrupted without impairing decision making for known tasks. The function of short-term buffers may be blocked by pharmacological agents. For example, the muscarinic antagonist scopolamine will block the ADP which provides one mechanism for sustained spiking of cortical neurons (Andrade, 1991; Klink and Alonso, 1997b; Fransen *et al.*, 2002). Without working short-term buffers in prefrontal cortex, the model predicts correct retrieval function for learned tasks, but an inability or impairment to learn new tasks. This may underlie the impairment of task rule shifting seen with cholinergic lesions (McGaughy *et al.*, 2004, 2005). Cholinergic blockade does cause impairment of short-term delayed matching function (Bartus and Johnson, 1976; Penetar and McDonough Jr., 1977).

Critical variables of the simulation

The successful results obtained with the simulation depend on several critical variables. Within the model of a prefrontal minicolumn, a specific set of connections must have conductances that lead to subthreshold excitation of postsynaptic neurons and another set must have conductances that lead to suprathreshold excitation and therefore drive spiking in postsynaptic neurons. The set of subthreshold connection consists of the connections from $\mathbf{a}$ to $\mathbf{g}_i$ and the connections from $\mathbf{g}_o$ to $\mathbf{c}_i$. The set of suprathreshold connections consists of the connections from $\mathbf{a}$ to $\mathbf{g}_o$, from $\mathbf{g}_i$ to $\mathbf{c}_o$, and from $\mathbf{c}_i$ to $\mathbf{g}_o$ (as shown in figure 4). For goal-directed prefrontal output, it is necessary that current state input to a $\mathbf{c}_o$ neuron population does not achieve spiking, except at those neurons that also receive gating input from neurons activated in the $\mathbf{g}_i$ population by the spread from the goal representation. Synapses at modifiable connections $\mathbf{W}_g$, $\mathbf{W}_{ig}$, $\mathbf{W}_c$ and $\mathbf{W}_{ic}$ are initialized with small subthreshold conductances. There is no need to adjust the learning rate during encoding, since a specific maximum conductance is achieved in strengthened connections. That maximum is set to provide suprathreshold excitation through the goal-spread connections $\mathbf{W}_g$ and $\mathbf{W}_{ig}$, and subthreshold excitation through $\mathbf{W}_c$ and $\mathbf{W}_{ic}$ (where the spiking of neurons in $\mathbf{c}_i$ is gated by $\mathbf{g}_o$ the spiking of neurons in $\mathbf{c}_o$ is gated by $\mathbf{g}_i$ during retrieval). The excitation of a neuron in $\mathbf{c}_i$ by individual input from $\mathbf{g}_o$ or through $\mathbf{W}_c$ and the excitation of a neuron in $\mathbf{c}_o$ by individual input from $\mathbf{g}_i$ or through $\mathbf{W}_{ic}$ is insufficient to elicit a spike. When two subthreshold inputs combine at a neuron in $\mathbf{c}_i$ (one from $\mathbf{g}_o$ and one through $\mathbf{W}_c$), or when two subthreshold inputs combine at a neuron in $\mathbf{c}_o$ (one from $\mathbf{g}_i$ and one through $\mathbf{W}_{ic}$) then a spike is elicited.

Another critical variable is the modulation of specific connection strengths in the minicolumn model by theta input (Hasselmo *et al.*, 2002). Theta modulation allows $\mathbf{g}_o^{specific}$ to drive $\mathbf{g}_o^{diffuse}$ through suprathreshold excitation and act as one population $\mathbf{g}_o$ for the spread of activity from a goal representation during retrieval phases. During encoding phases the connection is weakened and the two populations are treated separately as shown in figure 5. Differential modulation of excitatory input from $\mathbf{a}$ to $\mathbf{g}_o^{diffuse}$ (see fig. 5) and of input

from **a** to an interneuron population that sends inhibitory input to $\mathbf{g}_o^{diffuse}$ switches from suprathreshold excitatory input from **a** during encoding phases to providing the end-stopping function during retrieval phases. During encoding phases, the theta modulation enables input from buffered activity in $\mathbf{c}_i(n-1)$ to $\mathbf{g}_o^{specific}$ (fig. 5). Input from $\mathbf{g}_i$ to $\mathbf{c}_o$ is modulated so that suprathreshold excitation drives $\mathbf{c}_o$ during encoding phases, but subthreshold excitation performs the gating function during retrieval phases.

Lastly, critical variables are involved in the timing of short-term buffers (Lisman and Idiart, 1995; Jensen *et al.*, 1996; Koene *et al.*, 2003). A working buffer requires that the rise time of ADP matches the period of a theta cycle (Fransen *et al.*, 2002) and that recurrent inhibition separates consecutive spikes sufficiently to retain their order, but within a time interval that enables STDP between neurons that spike in response to the buffer output. For the first-in-first-out replacement of spikes maintained in a buffer, inhibitory input presented due to the combination of new input to the buffer and the last spike in the buffer must cause hyperpolarization at the phase of first spike reactivation (see fig. 6C). Theta oscillations achieve the necessary synchronization of reactivation cycles in the STM buffers and encoding and retrieval phases in the minicolumns.

Correspondence of simulation results and data

As mentioned in the results, the present study does not attempt to attribute meaning to the quantitative assessment of numbers of responses that belong to any specific category of responses that are selective for a trial type and a phase of that trial. For a quantitative comparison of that sort, an experimental study would have to record from a larger sample of neurons and the simulation would have to include a rationale for the number of cells in assemblies that correspond to each functional unit of the prefrontal model.

The model effectively matches the data in many ways, in addition to successfully learning the goal-directed behavior for the visual discrimination task. Our results show that the simulations replicate trial and phase-of-trial selective activity in individual neurons. A direct comparison between the selective activity recorded by Schultz *et al.* and that produced in the simulation (Fig. 11) demonstrates the correspondence between the two sets of results. Both the Schultz *et al.* data and our simulation results show individual neurons that are selective for the presentation of a visual cue, the period preceding potential reward in which a decision for motor action may be made, or the receipt of reward. That selectivity is specific to a particular trial type: rewarded movement, rewarded non-movement or unrewarded movement.

Significantly, both the data and the simulation results show that selectivity for exactly one specific trial type (RM, RNM or URM) was typical of responses that showed selective activity during the instruction phase of a trial, and atypical for responses that showed selective activity during a later phase of a trial. This correspondence supports the idea that those minicolumns that represent specific actions or rewards may be associated with multiple trial types. Another significant feature of the model is the absence of neurons that respond in both

RNM and URM trials, which also corresponds with the data.

Some properties of neuronal responses in the model are important for function, but may not be tested by the analysis procedures of the experiment. In particular, the analysis of experimental data did not specifically search for neurons which turned on continuously during task performance without showing specificity, and did not search for neurons which terminated activity at a specific time. The model produced background spiking activity that appears unselective for trial and phase throughout the task in 38 neurons. For the purpose of response categorization, this background spiking rate was subtracted to identify selective spike trains in those responses. The cells with this background activity are those that are involved in the spread of activity from the goal through associated minicolumns. Note that many such cells may have been deemed not task related by Schultz *et al.*, while they clearly perform an important function in the model. One indication of such background activity in the report by Schultz *et al.* comes in the form of neurons with task specific activity that appeared prior to the instruction stimulus. Schultz *et al.* evaluated activity in 188 out of 505 neurons. As specified in Tremblay and Schultz (Tremblay and Schultz, 2000), they did find 14 neurons that activated unselectively for all familiar instruction types in the task. Yet Schultz *et al.* evaluated neurons that activated selectively for one or two phases of specific task trials, since responses demonstrating activity throughout a trial may have been discarded by the one-tailed Wilcoxon test of the evaluation software that they used to assess task related activity.

The simulation results identified significant periods of inactivity in addition to the detection of selective activity. Some of the cells with background spiking throughout the trials of the task exhibit periods of inactivity that correspond directly with their involvement in the retrieval of a known association that determines goal-directed behavior in a specific trial. At such a simulated cell, inhibition (end-stopping) of the spread of activity from the goal representation causes the period of inactivity. Schultz *et al.* did not report a specific evaluation of the times at which the activity of some neurons ends, while other responses with rhythmic background activity during the same trial continue. Schultz *et al.* mention neurons that remain active throughout the instruction-trigger delay, but do not quantify the number of such cases. Cases reported in the data, where neural activity within a trial turns off immediately at the onset of a following phase may be indicative of end-stopping.

The simulation results show some differences compared to the data obtained by Schultz *et al.*. One that is immediately apparent is the precise and reproducible nature of specific intervals of spiking and of silence for each neuron in the model. This is a feature caused by the absence of noise in the simulated physiological functions.

A greater proportion of the responses recorded by Schultz *et al.* showed selective activity prior to reward or during reward in both RM and RNM trials than in only one of those two trial types. The proportions were reversed in the results obtained with the model, where more neurons responded to both, but these differences may not be meaningful due to the sample size issue outlined above.

The model responses contained a larger proportion of cells that respond selectively during both RM and URM trials than that reported by Schultz *et al.* In the trial phase preceding the reinforcer, this was a category not reported by Schultz *et al.* and a prediction of the model that further experiments with recordings at a greater number of sites may verify.

Relation to other physiological studies

This study shows how neuronal responses that guide behavior could reflect a conjunction of forward spread (stimulus dependent spread) and backward spread from goal (goal-dependent spread). The latter relates to responses obtained by Thorpe, Rolls and Maddison (Thorpe *et al.*, 1983), where the change in reward contingency demonstrates evidence for reward dependent response. The Schultz *et al.* experiments replicated here were an extension of the work by Thorpe and Rolls, who recorded single unit activity of orbitofrontal neurons in primates during a Go/NoGo operant task. In that task, monkeys learned to associate reward or an aversive outcome with movement following a specific stimulus. The meaning of a stimulus was reversed during this task. Thorpe and Rolls showed that most neurons responded selectively to specific stimuli and that the responses were also selective to whether the stimulus indicated reward in a specific trial. Simulation of Thorpe *et al.* using our model would require changes in reward contingency in the task, and the use of some mechanism of long-term depression in the model to replicate decrease in response to previously rewarded stimuli.

Tetrode recordings by Jung *et al.* (Jung *et al.*, 1998) showed that the correlation of activity in neurons in PFC does not map directly to sensory information such as location in spatial tasks. Rather, the activity correlates with behavioral requirements that are task specific, as shown with other simulations of a virtual rat in spatial tasks (Hasselmo, 2005). The present experimental results also relate to response data obtained by Schoenbaum *et al.* (Schoenbaum *et al.*, 1998), where changes in reward contingency were also shown to influence neuronal responses in rats. These responses were recorded in brain areas that communicate with orbitofrontal cortex through reciprocal connections, such as the basolateral amygdala which may provide feedback of an error function to avoid an aversive outcome.

In order to encode the specific components of a task and to encode predictive relationships by associating those components, the connections between neurons in networks of minicolumns and connections with the areas that provide input and receive output must be easily modifiable. Experimental evidence has been found for a rapid change in functional connectivity in terms of modifications of the strength of connections in orbitofrontal cortex and between orbitofrontal cortex and related areas such as the basolateral amygdala (Schoenbaum *et al.*, 2000; Mulder *et al.*, 2003). In those experiments, observed changes in odor selectivity were closely matched by changes in correlated firing activity during initial learning that led to accurate performance on a discrimination problem.

Relation to reinforcement learning theory: a biological implementation of reinforcement learning

Rules that govern successful behavior are discovered by learning how a specific action taken in one circumstance is followed by another circumstance. In other words, a causal effect is inferred from the results of a possible action that is explored while in a perceived state. In machine learning, algorithms for this are known as reinforcement learning (Sutton and Barto, 1998). In reinforcement learning, goals are explicit and formally represented by a reward value. The reinforcement learning framework has also been related to cognitive neural processes (Barto, 1995a,b; Montague *et al.*, 1996; Schultz *et al.*, 1997).

Reinforcement learning defines a state signal as any information that is available about the environment at a given time, which may be pre-processed sensory input and may include some memory of preceding states. The state signal has what is known as the Markov property if it contains a representation of all the information about current and preceding states and actions that are relevant to future decisions (White, 1969; Ross, 1983; Bertsekas, 1995). A state signal with the Markov property may be evaluated independent of the states and actions that precede it.

Reinforcement learning algorithms do not provide instruction about correct actions. Instead, an action is given a value by learning its consequences. Yet, reinforcement learning allows a range of different algorithms for learning these values. A popular algorithm for reinforcement learning is temporal difference (TD) learning (Sutton, 1988), which is related to models of conditioning (Konorski, 1948; Rescorla and Wagner, 1972). This algorithm learns from raw experience by updating predictive associations using a reward value at the time of update.

TD learning is useful, since it requires no information prior to exploration about the probabilities of transitions between states in an environment. And TD learning methods with Hebbian mechanisms (Hancock *et al.*, 1991; Montague *et al.*, 1993; Montague and Sejnowski, 1994; Rao and Sejnowski, 2001) have been proposed for the canonical circuit of neocortex (Douglas *et al.*, 1989; Artola *et al.*, 1990). An approach to TD learning, known as SARSA (state-action reward state-action), is notable for learning the value of actions in transitions between state-action pairs instead of the value of a state in transitions from state to state (Sutton, 1996; Sutton and Barto, 1998, chap. 7.5). The learning method in this paper assumes state-action pairs, as in the SARSA approach, although it is not derived from SARSA or TD learning.

The present model focuses on selection of actions on the basis of action-value. It does not require the use of TD learning to create the action value function, because the constrained nature of training ensured that it learned effective action value functions. Further modification will be needed to allow effective learning with random generation of actions during exploration, using a mechanism analogous to TD learning (Hasselmo, 2005). The model nevertheless provides a neural implementation of the action selection process in the reinforcement learning framework that does not depend on lookup tables.

In the model, encoding of behavioral rules requires that PFC contains unique

representations of specific states and actions. Fuster (Fuster, 2000) presented evidence that activity in the prefrontal cortex is representative of two types of perception, one that correlates with the sensory state evoked by past and current stimuli and one related to proprioceptive sensation and prediction of motor actions.

Given the representation of states and actions, the transition from one state to another state via a specific action can be encoded uniquely if there is specific neural activity that occurs only for that action and only when the action is initiated in a particular state. This requirement leads to the presupposition that a functional minicolumn contains populations of input neurons and populations of output neurons that form connections with other minicolumns, and that the neurons in those populations are connected in a structured manner to other minicolumns, in this simulation to exactly one. The internal weight matrices $\mathbf{W}_{ig}$ and $\mathbf{W}_{ic}$ of an action minicolumn act as second order conditional transition matrices from one state to another. A functionally similar pattern of connectivity could be learned by self-organization. Since the combination of activity at a specific input neuron and a specific output neuron of an action minicolumn represents the transition from a preceding state to a following state, that information gives the model the Markov property (e.g. Sutton and Barto, 1998, chap. 3.5). This property means that one-step dynamics enable us to predict the next state and expected reward for a specific action. Our model therefore provides a means of extending principles of reinforcement learning to biological circuits and the spiking responses of neurons.

Relation to anatomical data on minicolumns

The successive neuronal layers in a canonical circuit of the neocortex, as described by Douglas *et al.* (Douglas *et al.*, 1989), can be represented by the individual networks at the branch nodes of a hierarchical network (Felleman and Van Essen, 1991). Categorizing the parts of our model in such a hierarchy, the motor output (by populations $\mathbf{c}_i$ and $\mathbf{c}_o$) corresponds to the activity of the infragranular layer of the neocortex. Since sensory input is received in layer IV, its function may correspond to that of neurons designated $\mathbf{a}$. And the supragranular layer has many extensive and long range excitatory connections with other regions so that it can perform the function of our minicolumn model populations $\mathbf{g}_i$ and $\mathbf{g}_o$. This function that achieves the convergence of goal spread with current state input depends on the lateral connectivity within the neocortex. The lateral connectivity has been associated in studies of the visual cortex (Dayan and Hinton, 1996; Kawato *et al.*, 1993) with a necessary role in the interpretation of input and its translation into a complex hierarchical model. The generation of visual receptive fields that are tuned to recognize different orientations (Somers *et al.*, 1995; Yishai *et al.*, 1995) was related to this proposed role.

Lateral connectivity in the prefrontal region of neocortex includes short- and long-range excitatory connections, as well as short-range inhibitory connections (Barbas and Pandya, 1989; Barbas, 2000). The result is a patchy lateral

layout of cells that are highly interconnected within a column of cortical layers, the so-called neocortical minicolumn. It has been shown that strong local connectivity in a minicolumn can sustain activity during delayed response tasks such as long-term goal directed behavior for which a subject must be able to maintain information regarding a stimulus (Gutkin *et al.*, 2000; Wood and Grafman, 2003).

Local circuits that may exhibit the function of the proposed minicolumns were identified in the lateral connectivity of PFC, and Constantinidis and Goldman-Rakic (Constantinidis and Goldman-Rakic, 2002) showed that the activity of interneurons within such ensembles is strongly correlated. The correlated firing does not extend to distant areas or modules, and the activity of spatially proximate excitatory cells is less correlated than that of interneurons. In fact, spiking of different pyramidal cells responsible for the long-range propagation of activity is largely independent. Lund *et al.* (Lund *et al.*, 1993) proposed means by which such local circuits may arise during development. Analogous connectivity was described for the middle temporal visual area (Maunsell and Van Essen, 1983), and a model for similar local circuit development was proposed by Grossberg and Williamson (Grossberg and Williamson, 2001) for visual cortex areas V1 and V2. While our model resembles interaction of feedback and feedforward used in Grossberg and Williamson (Grossberg and Williamson, 2001), the visual models focus on top-down spread mediating global feature detection rather than reward contingencies. Our model more closely resembles the proposal by Mumford (Mumford, 1992) for bottom-up and top-down interactions.

If goal-directed behavior is to emerge in PFC its neuroanatomy must support activity that interprets sensory and proprioceptive motor input, and it must enable subsequent output that affects behavior. Previous surveys of the neuronal architecture of neocortex show that dual pathways between cortical areas could implement the necessary pathways for the analysis of input and the synthesis of output that guides behavior (Mumford, 1991, 1992, 1994). In the framework presented here, neuronal populations that correspond to cells in layer IV of neocortex are identified as input neurons for bottom-up cortical processing. Their ability to analyze input is represented by consequent activity of input neurons in a specific minicolumn. The associative connections between minicolumns lead to a synthesis of activity that represents goal-directed output.

While the model is intended to be applicable to the function of prefrontal minicolumns in general and not specific to orbitofrontal cortex, the encoding of reward found in orbitofrontal cortex for the Schultz *et al.* task led to a minicolumn representation of "reward state". In other (e.g. spatial) tasks where multiple routes can achieve a goal, a specific reward value may be encoded by differential strengthening of associations between reward and specific goal-directed strategies.

When a task includes multiple goals or strategies with different reward values, a mechanism must exist to select one goal over another and to direct behavior accordingly. The recruitment of distinct regions of orbitofrontal cortex has been observed during incentive judgments and goal selection. Lateral orbitofrontal activity has been observed selectively when a task required that

responses to alternative desirable items must be suppressed (Arana *et al.*, 2003). As implemented in the present model, gating by the spread of activity from one goal would compete with that of another goal at neuronal populations where goal spread and forward spread from current state converge. Successful neuronal firing suppresses the selection of other possibilities through recurrent inhibition.

Acknowledgments

The CATACOMB simulations described here and information about CATACOMB are available on our Computational Neurophysiology web site at http://askja.bu.edu.

Supported by NIH R01 grants DA16454, MH60013 and MH61492 to Hasselmo and by Conte Center Grant MH60450, as part of the NSF/NIH Collaborative Research in Computational Neuroscience Program.

Address correspondence to M.E.Hasselmo, Center for Memory and Brain, Department of Psychology and Program in Neuroscience, Boston University, 64 Cummington Street, Boston, MA 02215, U.S.A. Email: hasselmo@bu.edu.

References

Alonso A, Gaztelu J, Bruno Jr. W, Garcia-Austt E (1987). Cross-correlation analysis of septohippocampal neurons during theta-rhythm. *Brain Research* 413:135–146.

Andrade R (1991). The effect of carbachol which affects muscarinic receptors was investigated in prefrontal layer v neurons. *Brain Research* 541:81–93.

Arana F, Parkinson J, E. H, Holland A, Owen A, Roberts A (2003). Dissociable contributions of the human amygdala and orbitofrontal cortex to incentive motivation and goal selection. *Journal of Neuroscience* 23(29):9632–9638.

Artola A, Brocher S, Singer W (1990). Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature* 347:69–72.

Balleine B, Dickinson A (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407–419.

Barbas H (2000). Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices. *Brain Research Bulletin* 52(5):319–330.

Barbas H, Pandya D (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology* 286(3):353–375.

Barto A (1995a). Adaptive critics and the basal ganglia. In *Models of Information Processing in the Basal Ganglia* (Houk J, Davis J.L.and Beiser DG, eds.), pages 215–232. Cambridge, MA: MIT Press.

Barto A (1995b). Reinforcement learning. In *Handbook of Brain Theory and Neural Networks* (Arbib M, ed.), pages 804–809. Cambridge, MA: MIT Press.

Bartus R, Johnson H (1976). Short-term memory in rhesus monkey: disruption from the anti-cholinergic scopolamine. *Pharmacological Biochemical Behavior* 5:39–46.

Bechara A, Damasio A, Damasio H, Anderson S (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 50:7–15.

Bechara A, Damasio H, Tranel D, Damasio A (1997). Deciding advantageously before knowing the advantageous strategy. *Science* 275:1293–1295.

Bertsekas D (1995). *Dynamic Programming and Optimal Control*. Belmont, MA: Athena.

Bi G, Poo M (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience* 18(24):10464–10472.

Bliss T, Collingridge G (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature* 361:31–39.

Bliss T, Lømo T (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology* 232:331–356.

Bragin A, Jando G, Nadasdy Z, Hetke J (1995). Gamma (40-100 hz) oscillation in the hippocampus of the behaving rat. *Journal of Neuroscience* 15:47–60.

Brazhnik E, Fox S (1999). Action potentials and relations to the theta rhythm of septohippocampal neurons in vivo. *Experimental Brain Research* 127:244–258.

Cannon R, Hasselmo M, Koene R (2003). From biophysics to behaviour: Catacomb2 and the design of biologically plausable models for spatial navigation. *Neuroinformatics* 1:1:3–42.

Cohen J, Servan-Schreiber D (1992). Context, cortex and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review* 99:45–77.

Constantinidis C, Goldman-Rakic P (2002). Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *Journal of Neurophysiology* 88:3487–3497.

Dayan P, Hinton G (1996). Varieties of helmholtz machine. *Neural Networks* 9(8):1385–1403.

Douglas R, Martin K, Whitteridge D (1989). A canonical microcircuit for neocortex. *Neural Computation* 1:480–488.

Elman J (1990). Finding structure in time. *Cognitive Science* 14:179–211.

Elman J (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning* 7:195–224.

Felleman D, Van Essen D (1991). Distributed hierarchical processing in the primate cerebral cortx. *Cerebral Cortex* 1:1–47.

Fransen E, Alonso A, Hasselmo M (2002). Simulations of the role of the muscarinic activated calcium-sensitive nonspecific cation current i_{NCM} in entorhinal neuronal activity during delayed matching tasks. *Journal of Neuroscience* 22(3):1081–1097.

Frey S, Petrides M (1997). Orbitofrontal cortex: A key prefrontal regions for encoding information. *Proceedings of the National Academy of Sciences* 15:8723–8727.

Funahashi S, Bruce C, Goldman-Rakic P (1989). Mnemonic coding of visual space by neurons in the monkey's dorsolateral prefrontal cortex revealed by an oculomotor delayed response task. *Journal of Neurophysiology* 61:331–349.

Fuster J (1973). Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *Journal of Neurophysiology* 36:61–78.

Fuster J (2000). Prefrontal neurons in networks of executive memory. *Brain Research Bulletin* 52(5):331–336.

Fuster J, Bauer R, Jervey J (1982). Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks. *Experimental Neurology* 77:679–694.

Gerstner W (2002). *Integrate-and-Fire Neurons and Networks*. Cambridge, MA: MIT Press.

Gerstner W, Kistler W (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge, UK: Cambridge University Press.

Grossberg S, Williamson J (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual grouping and learning. *Cerebral Cortex* 11(1):37–58.

Gutkin B, Ermentrout G, O'Sullivan J (2000). Layer 3 patchy recurrent excitatory connections may determine the spatial organization of sustained activity in the primate prefrontal cortex. *Neurocomputing* 32-33:391–400.

Hancock P, Smith L, Phillips W (1991). A biologically supported error-correcting learning rule. *Neural Computation* 3:201–212.

Harel D (1987). Statecharts: A visual formalism for complex systems. *Scientific Computer Programming* 8:231–274.

Hasselmo M (2005). A model of prefrontal cortical mechanisms for goal directed behavior. *Journal of Cognitive Neuroscience* In press.

Hasselmo M, Bodelon C, Wyble B (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation* 14(4):793–817.

Izquierdo A, Murray E (2004). Combined unilateral lesions of the amygdala and orbital prefrontal cortex impair affective processing in rhesus monkeys. *Journal of Neurophysiology* 91:2023–2039.

Jensen O, Idiart M, Lisman J (1996). Physiologically realistic formation of autoassociative memory in networks with theta/gamma oscillations: Role of fast NMDA channels. *Learning & Memory* 3:243–256.

Jensen O, Lisman J (1996). Novel lists of 7 ± 2 known items can be reliably stored in an oscillatory short-term memory network: Interaction with long-term memory. *Learning & Memory* 3:257–263.

Jung M, Qin Y, McNaughton B, Barnes C (1998). Firing characteristics of deep layer neurons in prefrontal cortex in rats performaing spatial working memory tasks. *Cerebral Cortex* 8(5):437–450.

Kawato M, Hayakama H, Inui T (1993). A forward-inverse optics model of reciprocal connections between visual cortical areas. *Network* 4:415–422.

Klink R, Alonso A (1997a). Morphological characteristics of layer ii projection neurons in the rat medial entorhinal cortex. *Hippocampus* 7:571–583.

Klink R, Alonso A (1997b). Muscarinic modulation of the oscillatory and repetitive firing properties of entorhinal cortex layer ii neurons. *Journal of Neurophysiology* 77(4):1813–1828.

Koene R, Gorchetchnikov A, Cannon R, Hasselmo M (2003). Modeling goal-directed spatial navigation in the rat based on physiological data from the hippocampal formation. *Neural Networks* 16(5-6):577–584.

Konorski J (1948). *Conditioned reflexes and neuron organization*. Cambridge: Cambridge University Press.

Levy W, Stewart D (1983). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience* 8(4):791–797.

Lisman J, Idiart M (1995). Storage of 7 ± 2 short-term memories in oscillatory subcylces. *Science* 267:1512–1515.

Lübke J, von der Malsburg C (2004). Rapid processing and unsupervised learning in a model of the cortical macrocolumn. *Neural Computation* 16(3):501–533.

Lund J, Yoshioka T, Levitt J (1993). Comparison of intrinsic connectivity in different areas of macaque monkey cerebral cortex. *Cerebral Cortex* 3:148–162.

Manns I, Alonso A, Jones B (2000). Discharge properties of juxtacellularly labeled and immunohistochemically identified cholinergic basal forebrain neurons recorded in association with the electroencephalogram in anesthetized rats. *Journal of Neuroscience* 20(4):1505–1518.

Markram H, Lübke J, Frotscher M, Sakmann B (1997). Regularion of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science* 225:213–215.

Maunsell J, Van Essen D (1983). The connections of the middle temporal visual area (mt) and their relationship to a cortical hierarchy in the macaque monkey. *Journal of Neuroscience* 3(12):2563–2586.

McGaughy J, Koene R, Eichenbaum H, Hasselmo M (2004). Effects of cholinergic deafferentation of prefrontal cortex on working memory: A convergence of behavioral and modeling results. In *Proceedings of the 2004 Annual Meeting of the Society for Neuroscience*. San Diego, CA.

McGaughy J, Koene R, Eichenbaum H, Hasselmo M (2005). Cholinergic deafferentation of the entorhinal cortex impairs encoding of novel but not familiar stimuli. Unpublished data.

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

Montague P, Dayan P, Nowlan S, Pouget A, Sejnowski T (1993). Using aperiodic reinforcement for directed self-organization. In *Advances in Neural Information Processing Systems* (Giles C, Hanson S, Cowan J, eds.), volume 5, pages 969–977. San Mateo, CA: Morgan Kaufmann.

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

Montague P, Sejnowski T (1994). The predictive brain: Temporal coincidence and temporal order in synaptic learning mechanisms. *Learning and Memory* 1:1–33.

Mountcastle V (1997). The columnar organization of the neocortex. *Brain* 120:701–722.

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

Mumford D (1991). On the computational architecture of the neocortex: I. the role of the thalamo-cortical loop. *Biological Cybernetics* 65:135–145.

Mumford D (1992). On the computational architecture of the neocortex. ii. the role of cortico-cortical loops. *Biological Cybernetics* 66(3):241–251.

Mumford D (1994). Neuronal architectures for pattern-theoretic problems. In *Large-Scale Neuronal Theories of the Brain* (Koch C, Davis J, eds.), pages 125–152. Cambridge, MA: MIT Press.

O'Reilly R, Munakata Y (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press.

Pears A, Parkinson A, Hopewell L, Everitt B, Roberts A (2003). Lesions of the orbitofrontal but not medial prefrontal cortex disrupt conditioned reinforcement in primates. *Journal of Neuroscience* 23(35):11189–11201.

Penetar D, McDonough Jr. J (1977). Effects of cholinergic drugs on delayed match-to-sample performance of rhesus monkeys. *Pharmacological Biochemical Behavior* 19:963–967.

Quintana J, Fuster J (1992). Mnemonic and predictive functions of cortical neurons in a memory task. *NeuroReport* 3:721–724.

Rao R, Sejnowski T (2001). Spike-timing-dependent hebbian plasticity as temporal difference learning. *Neural Computation* 13(10):2221–2237.

Rescorla R, Wagner A (1972). A theory of pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In *Classical Conditioning II: Current Research and Theory* (Black A, Prokasy W, eds.), pages 64–69. New York, NY/somers: Appleton-Century-Crofts.

Rolls E (1999). The functions of the orbitofrontal cortex. *Neurocase* 5:301–312.

Ross S (1983). *Introduction to Stochastic Dynamic Programming*. New York: Academic Press.

Schoenbaum G. Setlow B, Ramus S (2003). A systems approach to orbitofrontal cortex function: recordings in rat orbitofrontal cortex reveal interactions with different learning systems. *Behavioral Brain Research* 146:19–29.

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

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

Schoenbaum G, Eichenbaum H (1995a). information coding in the rodent prefrontal cortex. i. single-neuron activity in orbitofrontal cortex compared with that in pyriform cortex. *Journal of Neurophysiology* 74(2):733–750.

Schoenbaum G, Eichenbaum H (1995b). Information coding in the rodent prefrontal cortex. ii. ensemble activity in orbitofrontal cortex. *Journal of Neurophysiology* 74(2):751–762.

Schultz W (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology* 80:1–27.

Schultz W, Dayan P, Montague P (1997). A neural substrate of prediction and reward. *Science* 275:1593–1598.

Schultz W, Dickinson A (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience* 23:473–500.

Schultz W, Tremblay L, Hollerman J (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cerebral Cortex* 10:272–283.

Somers D, Nelson S, Sur M (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *Journal of Neuroscience* 15:5448–5465.

Stein R (1967). Some models of neuronal variability. *Biophysics Journal* 7:37–68.

Stewart M, Fox S (1990). Do septal neurons pace the hippocampal theta rhythm? *Neuron* 13:163–168.

Sutton R (1988). Learning to predict by the methods of temporal difference. *Machine Learning* 3:9–44.

Sutton R (1996). Generalization in reinforcement learning: Successful examples using sparse coarse coding. In *Advances in Neural Information Processing Systems 8*, pages 1038–1044. Cambridge, MA: MIT Press.

Sutton R, Barto A (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review* 88:135–140.

Sutton R, Barto A (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press. A Bradford Book.

Terrace H, Son L, Brannon E (2003). Serial expertise of rhesus macaques. *Psychological Science* 14(1):66–73.

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

Tremblay L, Schultz W (1999). Relative reward preference in primate orbitofrontal cortex. *Nature* 398(6729):704–708.

Tremblay L, Schultz W (2000). Reward related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. *Journal of Neurophysiology* 83(4):1877–1885.

Wallis J, Anderson K, Miller E (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature* 411(6840):953–956.

Wallis J, Miller E (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *European Journal of Neuroscience* 18:2069–2081.

White D (1969). *Dynamic Programming*. San Francisco: Holden-Day.

Wood J, Grafman J (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews Neuroscience* 4:139–147.

Yishai B, Baror R, Sompolinsky H (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences* 92:3844–3848.