European Journal of Neuroscience, Vol. 18, pp. 2374–2390, 2003

Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex

Gustavo Deco[1] and Edmund T. Rolls[2]
[1]Institució Catalana de Recerca i Estudis Avançats (ICREA), Universitat Pompeu Fabra, Department of Technology, Computational Neuroscience, Passeig de Circumval.lació, 08003 Barcelona, Spain
[2]University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, UK

Keywords: attractor network, autoassociation network, dopamine, executive function, prefrontal cortex, short-term memory, task switching

Abstract

Cognitive behaviour requires complex context-dependent mapping between sensory stimuli and actions. The same stimulus can lead to different behaviours depending on the situation, or the same behaviour may be elicited by different cueing stimuli. Neurons in the primate prefrontal cortex show task-specific firing activity during working memory delay periods. These neurons provide a neural substrate for mapping stimulus and response in a flexible, context- or rule-dependent, fashion. We describe here an integrate-and-fire network model to explain and investigate the different types of working-memory-related neuronal activity observed. The model contains different populations (or pools) of neurons (as found neurophysiologically) in attractor networks which respond in the delay period to the stimulus object, the stimulus position ('sensory pools'), to combinations of the stimulus sensory properties (e.g. the object identity or object location) and the response ('intermediate pools'), and to the response required (left or right) ('premotor pools'). The pools are arranged hierarchically, are linked by associative synaptic connections, and have global inhibition through inhibitory interneurons to implement competition. It is shown that a biasing attentional input to define the current rule applied to the intermediate pools enables the system to select the correct response in what is a biased competition model of attention. The integrate-and-fire model not only produces realistic spiking dynamicals very similar to the neuronal data but also shows how dopamine could weaken and shorten the persistent neuronal activity in the delay period; and allows us to predict more response errors when dopamine is elevated because there is less different activity in the different pools of competing neurons, resulting in more conflict.

Introduction

There is much evidence that the prefrontal cortex (PFC) is involved in at least some types of working memory and related processes (Jacobsen, 1935; Goldman-Rakic, 1995; Goldman-Rakic, 1996; Fuster, 2000). Working memory refers to a system for maintaining and manipulating information in mind, held during a short period, usually of seconds (Baddeley, 1986). Neuronal recording studies indicate continuing firing of prefrontal neurons during the delay period of working memory tasks (Fuster & Alexander, 1971; Kubota & Niki, 1971; Funahashi *et al*., 1989; Funahashi *et al*., 1993), and imaging studies have confirmed activation of the PFC (Leung *et al*., 2002; Ungerleider *et al*., 1998; Adcock *et al*., 2000; Zarahn *et al*., 2000). Prefrontal lesions in humans (Milner, 1963; Goldman-Rakic *et al*., 1987) and monkeys (Butters & Pandya, 1969; Levy & Goldman-Rakic, 1999) produce severe deficits in tasks requiring short-term memory processing.

The memory-related persistent prefrontal neuronal activity during the delay period of short-term memory tasks could be maintained by assuming recurrent collateral excitatory loops (Hebb, 1949; Goldman-Rakic, 1995) which can be formally modelled and analysed by autoassociation networks. These networks store a set of memory patterns in the recurrent synaptic connections between the excitatory neurons (pyramidal cells) in the network, and when triggered with any one of the memory cue patterns, maintain that pattern of neuronal firing even when the cue is removed in a stable 'attractor' state (see Amit, 1995; Rolls & Treves, 1998; Rolls & Deco, 2002).

The neurophysiological investigations of the functions of the PFC in working memory have been extended recently by analysing neuronal activity when the monkey performs two different working memory tasks using the same stimuli and responses (Asaad *et al*., 1998; Hoshi *et al*., 1998; White & Wise, 1999; Asaad *et al*., 2000). The aim of the present work is to model, and therefore help to understand, the underlying mechanisms which implement the working-memory-related activity observed in PFC neurons in the context-dependent stimulus–response (associative) and delayed spatial response tasks investigated by Asaad *et al*. (2000) and Asaad *et al*. (1998). The model builds on the integrate-and-fire attractor network treatment of Brunel & Wang (2001) (which was discussed in the context of a simple object working memory with no arbitrary stimulus to response mapping, reversal, or rule change from associative to delayed spatial response). The new model introduced in this paper greatly extends this approach by introducing a hierarchically organised set of different attractor networks each with a different population of neurons (corresponding to the different types of neuron recorded during the performance of these tasks), and by introducing biasing inputs to the intermediate layer

Correspondence: Professor Edmund T Rolls, as above.
E-mail: Edmund.Rolls@psy.ox.ac.uk
Web: www.cn5.ox.ac.uk

Received 28 January 2003, revised 25 July 2003, accepted 29 July 2003

doi:10.1046/j.1460-9568.2003.02956.x

attractor networks which switch the whole system from performing a delayed response to a conditional object-response task. The biasing function is analogous (see also Miller & Cohen, 2001) to that employed in biased competition models of attention (Moran & Desimone, 1985; Spitzer *et al*., 1988; Chelazzi *et al*., 1993; Miller *et al*., 1993; Motter, 1993; Chelazzi, 1998; Reynolds & Desimone, 1999; Rolls & Deco, 2002). Advantages of the implementation level for modelling complex PFC context-dependent processing introduced in this paper, the integrate-and-fire level, are that (i) realistic spiking dynamicals are produced by the model for direct comparison with the neurophysiological data, (ii) effects on global processes such as memory of transmitters altering the dynamicals of different types of synapse in the brain can be investigated as described here for dopamine, and (iii) predictions can be made about how different manipulations of the system affect behaviour.

Materials and methods

The short-term memory tasks modelled

The modelling was of situations in which the monkey was required to perform two different working memory tasks using the same stimuli and responses (Asaad *et al*., 2000; Asaad *et al*., 1998; Hoshi *et al*., 1998; White & Wise, 1999). In a 'conditional object-response (associative) task' with a delay the monkey was shown one of two stimuli and, after a delay, had to make either a rightward or leftward oculomotor saccade response depending on which stimulus was shown (Asaad *et al*., 1998). The task could be performed directly or with a reversal. In another experiment, recordings were made both during the object-response task and during a 'delayed spatial response task', in which the same stimuli were used but the rule required was different, namely to respond towards the location where the stimulus had been shown (Asaad *et al*., 2000). The main motivation for such studies was the fact that, for real-world behaviour, the mapping between a stimulus and a response is typically more complicated than a one-to-one mapping. The same stimulus can lead to different behaviours depending on the situation, or the same behaviour may be elicited by different cueing stimuli. In the performance of these tasks, neurons have been described which respond in the delay period to the stimulus object, the stimulus position ('sensory pools'), to combinations of the response and the stimulus object or position ('intermediate pools'), and to the response required (left or right) ('premature pools') (Asaad *et al*., 2000, 1998). Further details of the implementation of the tasks for the neurophysiology and for the simulations, and a comparison of the data from the model with that from the neurophysiology, are described in Results. In addition, Wallis *et al*. (2001) have described neurons in the primate PFC which reflect the explicit coding of abstract rules (independently of the identity of the cue and of the response).

Implementation of the integrate-and-fire model

The model simulated is designed to enable investigation of the dynamicals of a set of attractor networks, connected to form a hierarchy, and in which selection of which neuronal pools win the competition implemented by global inhibition is determined by a biasing input which reflects the rule, context or attention. Because the aims include understanding the dynamicals at the level of how the spiking of neurons evolves on each type of trial for comparison with neuronal recording data, and the contribution of different types of synaptic connection (using AMPA or *N*-methyl-D-aspartate (NMDA) receptors for excitatory input and γ-aminobutyric acid (GABA) receptors for inhibitory input), the simulation is at the level of integrate-and-fire neurons. We followed the theoretical framework introduced and studied by Brunel & Wang (2001) and the biased competition-based neurodynamical framework introduced by the authors (Deco & Zihl, 2001; Corchs & Deco, 2002; Deco & Lee, 2002; Rolls & Deco, 2002). We incorporated the shunting inhibition (Battaglia & Treves, 1998; Rolls & Treves, 1998) and inhibitory to inhibitory cell synaptic connections (Battaglia & Treves, 1998) which are useful in maintaining stability of the dynamical system, and incorporated appropriate currents to achieve low firing rates (Amit & Brunel, 1997; Brunel & Wang, 2001). The existence of different types of neuronal response was assumed to be what has been found neurophysiologically (Asaad *et al*., 2000, 1998), and we show that associative connections (which could be set up by long-term potentiation) between these neuronal pools are sufficient for operation of the model. In this section we describe the architecture and operation of the model and in the Appendix we provide a full mathematical specification of the model and the neuronal parameters used.

Both the excitatory pyramidal cells and the inhibitory interneurons are modelled by leaky integrate-and-fire neurons. Figure 1 shows graphically the synaptic and membrane processes. The basic circuit of an integrate-and-fire model consists of the cell membrane capacitance C_m in parallel with the cell membrane resistance R_m driven by a synaptic current (excitatory or inhibitory postsynaptic potential, EPSP or IPSP, respectively). If the voltage across the capacitor reaches a threshold θ the circuit is shunted and a δ-pulse (spike) is generated and transmitted to other neurons. The incoming presynaptic δ-pulse from other neurons is basically first low-pass filtered by the synaptic and membrane time constants before it is utilised as an EPSP or IPSP in the one-compartment neuronal model. We used biologically realistic parameters (McCormick *et al*., 1985). We assumed for both kinds of neuron a resting potential $V_L = -70$ mV, a firing threshold $\theta = -50$ mV and a reset potential $V_{reset} = -55$ mV. The membrane capacitance C_m was 0.5 nF for the pyramidal neurons and 0.2 nF for the interneurons. The membrane leak conductance g_m was 25 nS for pyramidal cells and 20 nS for interneurons. The refractory period τ_{ref} was 2 ms for pyramidal cells and 1 ms for interneurons. Consequently, the membrane time constant $\tau_m = C_m/g_m$ was 20 ms for pyramidal cells and 10 ms for interneurons.

The synaptic current flows into the cells are mediated by four different families of receptors. The recurrent excitatory postsynaptic currents are given by two different types of EPSP mediated, respectively, by AMPA and NMDA receptors. These two glutamatergic excitatory synapses are on the pyramidal cells and interneurons. The external inputs (background, sensory input, or external top–down interaction from other areas) are mediated by AMPA synapses on pyramidal cells and interneurons. Inhibitory GABAergic synapses on pyramidal cells and interneurons yield the corresponding IPSPs. The mathematical descriptions of each synaptic channel are thoroughly described in the Appendix, and the corresponding parameters are also specified there. We consider that the NMDA currents have a voltage dependence which was controlled by the extracellular magnesium concentration (Jahr & Stevens, 1990); for our model, $[Mg^{++}] = 1$ mM. We neglect the rise times of both AMPA and GABA synaptic currents because they are typically extremely short (<1 ms). The rise time for NMDA synapses is $\tau_{NMDA,rise} = 2$ ms (Hestrin *et al*., 1990; Spruston *et al*., 1995). The decay time for AMPA synapses is $\tau_{AMPA} = 2$ ms (Hestrin *et al*., 1990; Spruston *et al*., 1995), for NMDA synapses $\tau_{NMDA,decay} = 100$ ms (Hestrin *et al*., 1990; Spruston *et al*., 1995) and for GABA synapses $\tau_{GABA} = 10$ ms (Salin & Prince, 1996; Xiang *et al*., 1998). The synaptic conductivities for each receptor type, shown in the Appendix, were similar to those used by Brunel & Wang (2001) and are consistent with experimentally observed values (Destexhe *et al*., 1998). As was noted by Brunel & Wang (2001) and Wang (1999) following Lisman *et al*. (1998), the recurrent excitation was assumed to

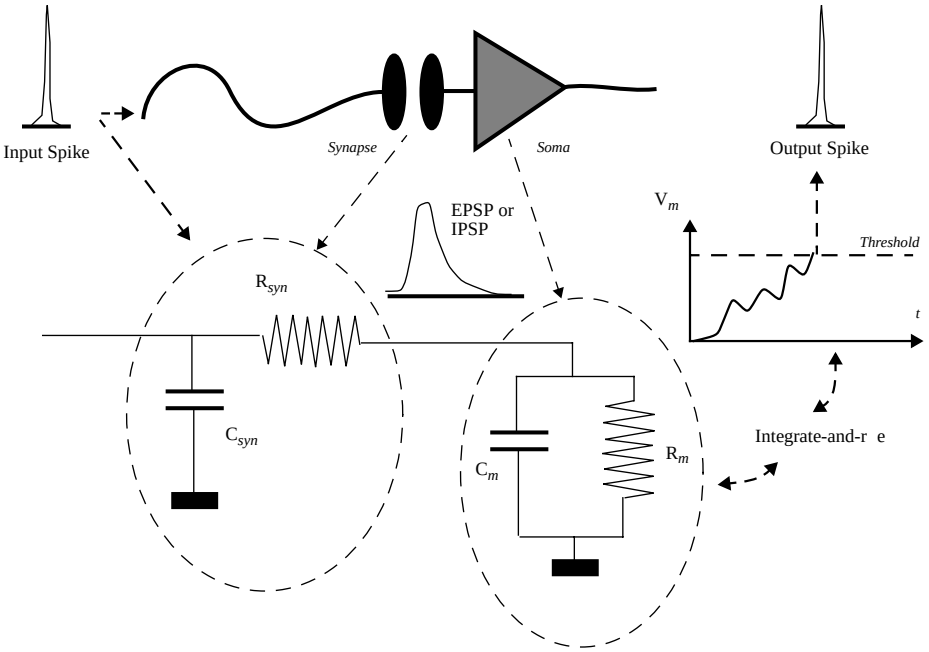

FIG. 1. Integrate-and-fire neuron. The basic circuit of an integrate-and-fire model consists of the neuron's membrane capacitance C_m in parallel with the membrane's resistance R_m driven by a synaptic current with a conductance and time constant determined by the synaptic resistance R_{syn} and capacitance C_{syn} shown in the Figure. These effects produce excitatory or inhibitory postsynaptic potentials, EPSPs or IPSPs. These potentials are integrated by the cell, and if a threshold θ is reached a δ-pulse (spike) is fired and transmitted to other neurons, and the membrane potential is reset.

be largely mediated by the NMDA receptors in order to provide more robust persistent activity during the short-term memory-related delay period, and the amplitude of recurrent excitation was smaller than that of inhibition; therefore the net recurrent input to a neuron was hyperpolarizing during spontaneous activity (i.e. without external inputs) (Amit & Brunel, 1997; Brunel & Wang, 2001).

All neuronal and synaptic equations were integrated using the second-order Runge–Kutta method, with an integration step of $dt = 0.1$ ms. Checks were performed to show that this was sufficiently small. For the neural membrane potential equations, interpolation of the spike times and their use in the synaptic currents and potentials were taken into account following the prescription of Hansel *et al.* (1998), in order to avoid numerical problems due to the discontinuity of the membrane potential and its derivative at the spike firing time. The external trains of Poisson spikes were generated randomly and independently.

The network architecture implemented

Figure 2 shows schematically the synaptic structure assumed in the prefrontal cortical network. The network was composed of N_E (excitatory) pyramidal cells and N_I inhibitory interneurons. In our simulations, we used $N_E = 1600$ and $N_I = 400$. The relationship between excitatory and inhibitory neurons was therefore consistent with the neurophysiologically observed proportion of 80% pyramidal cells vs. 20% interneurons (Abeles, 1991). The neurons were fully connected. There were different populations or pools of neurons in the prefrontal cortical network, as shown in Fig. 2. Each pool of excitatory cells contained fN_E neurons, where f, the fraction of the neurons in any one pool, was set to be 0.05. There were two different types of pool: excitatory and inhibitory. There were four excitatory pools, namely: sensory, task or rule-specific, premotor, and nonselective. The sensory pools encoded information about objects, or spatial location. Object- or feature-based sensory pools were feature-specific, encoding for example the identity of an object (e.g. form, colour, etc.). The spatial sensory pools were location-specific and encoded the spatial position of a stimulus. The premotor pools encoded the motor response (in our case the leftward or rightward oculomotor saccade). The intermediate pools (in that they are between the sensory and premotor pools) were task-specific and performed the mapping between the sensory stimuli and the required motor response. The intermediate pools responded to combinations of the sensory stimuli and the response required, e.g. to object 1 requiring a left oculomotor saccade. The intermediate pools received an external biasing input which reflected the current rule (e.g. on this trial when object 1 was shown make the left response after the delay period). The remaining excitatory neurons did not have specific sensory, response or biasing inputs, and were in a nonselective pool. [They had some spontaneous firing, and helped to introduce some noise into the simulation, to help to produce the almost Poisson spike firing patterns of neurons in the simulation which are a property of many neurons recorded in the brain (Brunel & Wang, 2001)]. All the inhibitory neurons were clustered into a common inhibitory pool so that there was global competition throughout the network.

We assumed that the synaptic coupling strengths between any two neurons in the network are established by Hebbian learning, i.e. the coupling would be strong if the pair of neurons had correlated activity and weak if they were activated in an uncorrelated way. As a consequence of this, neurons within a specific excitatory pool were mutually coupled with a strong weight $w_s = 2.1$. Neurons in the inhibitory pool were mutually connected with an intermediate weight $w = 1$ (forming the inhibitory to inhibitory connections which are useful in achieving nonoscillatory firing). They were also connected with all excitatory neurons with the same intermediate weight $w = 1$. The connection strength between two neurons in two different specific excitatory pools was weak and given by $w_w = 1 - 2f(w_s - 1)/(1 - 2f)$ $(= 0.8778)$ unless otherwise specified (see next paragraph). Neurons in a specific excitatory pool were connected to neurons in the nonselective pool with a feedforward synaptic weight $w = 1$ and a feedback synaptic connection of weight w_w.

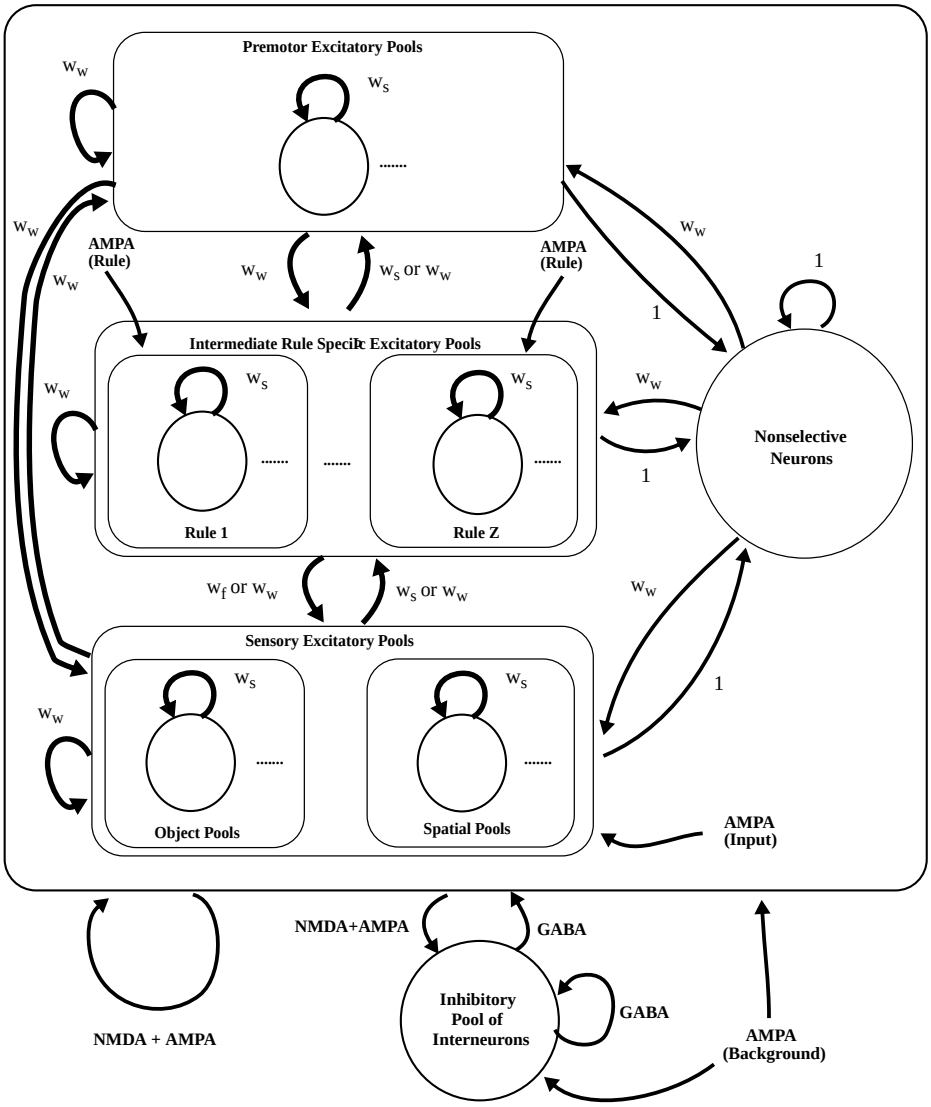

FIG. 2. Prefrontal cortical module. The synaptic current flows into the cell are mediated by four different families of receptors. The recurrent excitatory postsynaptic currents are given by two different types of EPSP, mediated respectively by AMPA and NMDA receptors. These two glutamatergic excitatory synapses are on the pyramidal cells and interneurons. The external inputs (background, sensory input, or external top-down interactions from other areas) are mediated by AMPA synapses on pyramidal cells and interneurons. Inhibitory GABAergic synapses on pyramidal cells and interneurons yield corresponding IPSPs. The connections are set up assuming Hebbian learning during the training of the monkey while performing the different tasks (e.g. conditional object-response and delayed spatial response, or direct and reversed conditional object-response (sometimes called associative). Each neuron also receives N_{ext} excitatory AMPA synaptic connections from outside the network. These connections provide three different type of external interactions: (i) a background noise due to the external spontaneous firing activity; (ii) a sensory input; and (iii) a biasing context task specification.

The connections between the different pools were set up to achieve the required mapping from the sensory input pools through the intermediate pools to the premotor pools, assuming Hebbian learning based on the activity of individual pools while the different tasks are being performed. Unless otherwise specified, the forward connections (input to intermediate to output pools) were $w_{ff} = 2.1$. The corresponding feedback synaptic connections were slightly weaker (w_{fb} for the feedback synapses between rule-specific and sensory pools, and w_w for the feedback synapses between the premotor and rule-specific pools).

Each neuron (pyramidal cells and interneurons) received $N_{ext} = 800$ excitatory AMPA synaptic connections from outside the network. These connections provided three different type of external interactions: (i) a background noise due to the spontaneous firing activity of neurons outside the network; (ii) a sensory-related input; and (iii) a rule or context-related bias input which specifies the task. The external inputs were given by a Poisson train of spikes. In order to model the background spontaneous activity of neurons in the network (Brunel & Wang, 2001), we assumed that Poisson spikes arrive at each external synapse with a rate of 3 Hz, consistent with the spontaneous activity observed in the cerebral cortex (Wilson *et al*., 1994; Rolls & Treves, 1998). In other words, the effective external spontaneous background input rate of spikes to each cell was $\nu_{ext} = N_{ext} \times 3\,\text{Hz} = 2.4\,\text{kHz}$. The sensory input was encoded by increasing the external input Poisson rate ν_{ext} to $\nu_{ext} + \lambda_{input}$ to the neurons in the appropriate specific sensory pools (Brunel & Wang, 2001). For example, if the stimulus was defined as an object with a feature characteristic F_i and at a spatial location S_i, then the neurons in the sensory object pool F_i and in the spatial sensory pool S_i would receive the increased external Poisson input just defined. We used $\lambda_{input} = 120\,\text{Hz}$. Finally, the biasing specification of the context, i.e. which rule is active, was modelled

by assuming that each neuron in each of the pools in the group of intermediate pools associated with the active task received external Poisson spikes with an increased rate from ν_{ext} to $\nu_{ext} + \lambda_{rule}$ throughout the trial. We used $\lambda_{rule} + 100$ Hz. This external top-down rule-specific input probably comes from the external prefrontal neurons which directly encode abstract rules (Wallis *et al*., 2001), which in turn are influenced by the reward system (in the orbitofrontal cortex and amygdala) to enable the correct rule to be selected during for example reversal (Thorpe *et al*., 1983). During the last 100 ms of the response period, the external rate to all neurons was increased by a factor of 1.5 in order to take into account the increase in afferent inputs due to behavioural responses and reward signals (Brunel & Wang, 2001).

The cortical architecture introduced above presents the characteristic that its different global attractors corresponding to the different sensory cue-response context situations are each composed of a set of single pool attractors, where the single pools which are active represent a particular combination of sensory, intermediate and premotor pools. The cue stimulus and the biasing top-down rule or context information drive the system into the corresponding attractor. In fact, the system is dynamically driven according to the biased competition hypothesis (Moran & Desimone, 1985; Spitzer *et al*., 1988; Chelazzi *et al*., 1993; Miller *et al*., 1993; Motter, 1993; Chelazzi, 1998; Reynolds & Desimone, 1999). Multiple excitatory pools of neurons activated by the sensory cue stimulus engage in competitive interactions using the interneurons to implement the global competition. The external top–down interactions bias this competition in favour of specific pools, resulting in the build-up of the global attractor which corresponds to the context-specific cue-response mapping required. In this way, irrelevant sensory information will be suppressed by the underlying neurodynamicals, implementing a form of internal prefrontal attentional system which is the basis of the attentional top-down bias transmitted to posterior perceptual areas (Rolls & Deco, 2002).

The cortical architecture can also be considered to have the structure of a single attractor network, with some associative excitatory connections between all neurons in the network. Within the network, one set of neurons is trained associatively to represent one cue stimulus (e.g. object A), another set to represent object B, another set to represent a left location of the stimulus, etc. The spatial and object sensory neuronal pools referred to above represent the effects of this associative training. Similarly, other neuronal pools are trained associatively to represent different responses, e.g. leftward eye movement and rightward eye movement. Further pools, referred to as intermediate or rule-based, are trained to respond to particular combinations of sensory cues with responses. The strengths of these associative connections are indicated in Tables 1 and 2. The departure from what could be thought of as different patterns trained into a single attractor, and which leads to the description that the network is hierarchical, is that the sensory pools then in general have stronger forward connections to the intermediate neurons than the backprojections, and the intermediate neurons in turn have stronger forward connections to the response pools than the backprojections. Although these forward and backward connections are associative, the strength (in general) is stronger in the forward than the backward direction. These connection strengths are also included in Tables 1 and 2. Another departure from a single attractor architecture is that the intermediate neuronal pools receive a task-specific bias, as described above. The external information which drives the dynamicals of the system towards different attractors is composed of two components, first the sensory input information (i.e. the spatial location and the object features) and second the external rule bias defining the context applied to the intermediate neuronal pools. Consequently, for the same sensory information, different contexts (i.e. different valid rules) will drive the system dynamicalally to different attractors corresponding to different behavioural sensory-response mappings. The whole system operates as an attractor in that the associatively trained recurrent connections between the neurons (shown in Fig. 2) enable the system to maintain its activity in the delay period after the sensory inputs applied during the cue period have been removed.

TABLE 1. Neuronal connectivity between different neuronal pools in Experiment 1, reversal of delayed object–response associations

Pools	A	B	AL	BR	AR	BL	L	R	Nsp.	Inh.
A	w_s	w_w	w_{ff}	w_w	w_{ff}	w_w	w_w	w_w	1	1
B	w_w	w_s	w_w	w_{ff}	w_w	w_{ff}	w_w	w_w	1	1
AL	w_{fb}	w_w	w_s	w_w	w_w	w_w	w_s	w_w	1	1
BR	w_w	w_{fb}	w_w	w_s	w_w	w_w	w_w	w_s	1	1
AR	w_{fb}	w_w	w_w	w_w	w_s	w_w	w_w	w_s	1	1
BL	w_w	w_{fb}	w_w	w_w	w_w	w_s	w_s	w_w	1	1
L	w_w	w_w	w_w	w_w	w_w	w_w	w_s	w_w	1	1
R	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_s	1	1
Nsp.	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_w	1	1
Inh.	1	1	1	1	1	1	1	1	1	1

A, B, objects A and B; AL, intermediate pool for object A on the left, etc.; L, R, left and right response pools; Nsp, non-specific neuronal pool; Inh, inhibitory neuron pool. w_w, weak synaptic strength (0.878); w_s, strong synaptic strength (2.1); w_{ff}, feedforward synaptic strength (1.8); w_{fb}, feedback synaptic strength (1.6).

TABLE 2. Neuronal connectivity between different neuronal pools in Experiment 2, delayed object–response association vs. delayed spatial response

Pools	O1	O2	S1	S2	R1	R2	R3	R4	L	R	Nsp.	Inh.
O1	w_s	w_w	w_w	w_w	w_{ff}	w_w	w_w	w_w	w_w	w_w	1	1
O2	w_w	w_s	w_w	w_w	w_w	w_{ff}	w_w	w_w	w_w	w_w	1	1
S1	w_w	w_w	w_s	w_w	w_w	w_w	w_{ff}	w_w	w_w	w_w	1	1
S2	w_w	w_w	w_w	w_s	w_w	w_w	w_w	w_{ff}	w_w	w_w	1	1
R1	w_{fb}	w_w	w_w	w_w	w_s	w_w	w_w	w_w	w_s	w_w	1	1
R2	w_w	w_{fb}	w_w	w_w	w_w	w_s	w_w	w_w	w_w	w_s	1	1
R3	w_w	w_w	w_{fb}	w_w	w_w	w_w	w_s	w_w	w_s	w_w	1	1
R4	w_w	w_w	w_w	w_{fb}	w_w	w_w	w_w	w_s	w_w	w_s	1	1
L	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_s	w_w	1	1
R	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_s	1	1
Nsp.	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_w	1	1
Inh.	1	1	1	1	1	1	1	1	1	1	1	1

O1, O2, objects 1 and 2; S1, S2, spatial locations 1 and 2 of the stimuli; R1–R4, rule-dependent intermediate pools; L, R, left and right response pools; Nsp, Non-specific neuronal pool; Inh, inhibitory neuron pool. w_w, weak synaptic strength (0.878); w_s, strong synaptic strength (2.1); w_{ff}, feedforward synaptic strength (2.1); w_{fb}, feedback synaptic strength (1.7).

Results

Associative object-response learning and its reversal in the primate PFC

In this subsection, we present a theoretical analysis of neuronal activity in the primate PFC underlying the execution of an associative object-response task (Asaad *et al*., 1998). Particularly interesting in the single-cell experiments of Asaad *et al*. (1998) was the discovery of individual PFC neurons which represent combinations of the stimulus cues and the associated responses, providing a neural substrate for a task-specific association of particular sensory cues with particular behavioural responses.

The experiment of Asaad *et al*. (1998) aimed to explore the role of the PFC in arbitrary cue-response learning by studying the neural activity of lateral PFC neurons during performance of a conditional

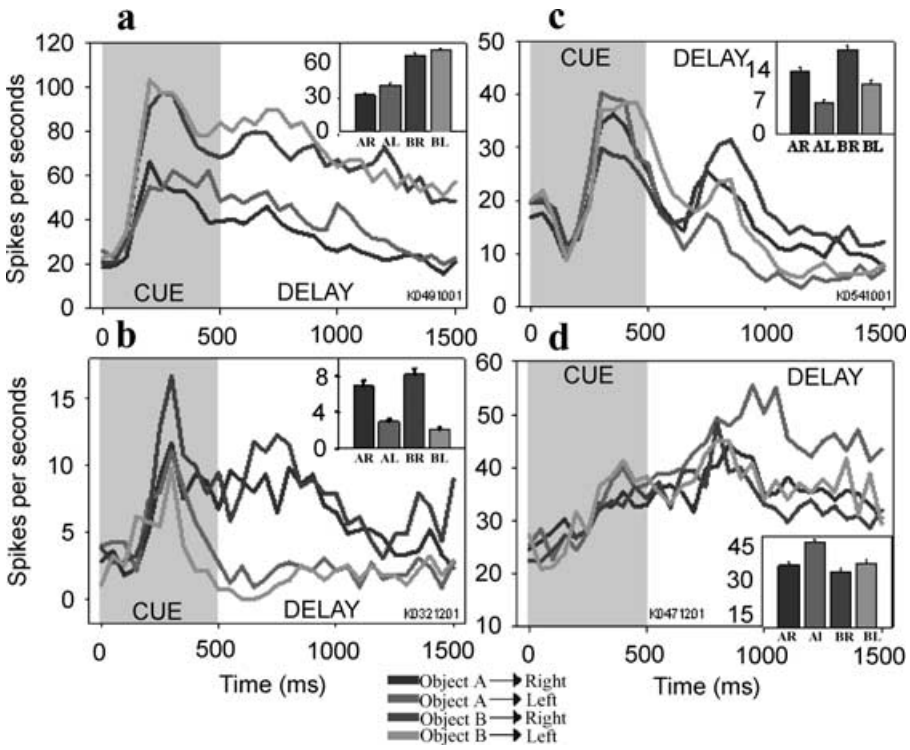

FIG. 3. Experimental recordings of Asaad *et al*. (1998). The figure shows the firing activity of four neurons to each of the four possible cue–saccade associations. The shaded area represents the time of cue presentation and the bar plot in the inset shows the mean level of delay firing activity for each cue–response association: AR (Object A → rightward saccade), AL (Object A → leftward saccade), BR (Object B → rightward saccade), and BL (Object B → leftward saccade). The colours used in the inset match those of the histograms and are keyed by the legend. The top left plot corresponds to an object-selective neuron (selective to object B), the bottom left plot corresponds to a response-selective neuron (selective to a rightward-response), and the top and bottom right plots correspond to neurons which are selective to a task-specific association (top BR, bottom AL). (Reproduced with permission from Asaad *et al*., 1998.)

visuomotor task. The task required the monkeys to associate a foveally presented cue object (Object A or B), after a delay period, with a response consisting of a leftward or rightward saccadic eye movement. The cue period was 500 ms and the short-term memory delay period separating the cue and response was 1000 ms. They trained the monkeys under two different conditions, namely: (i) direct association, and (ii) reverse association. The direct condition corresponded to the association of one object, for example A, with a leftward eye motor response, and the other object, for example B, with a rightward eye motor response. The reverse condition corresponded to the reversed association of cue and responses, i.e. object A was now associated with a rightward eye motor response and object B with a leftward eye motor response. Figure 3 illustrates the experimental recordings of Asaad *et al*. (1998) after learning. The figure shows the firing activity of four neurons to each of the four possible cue–saccade associations. The shaded area represents the time of cue presentation, and the bar plot in the inset shows the mean level of delay firing activity for each cue–response association: AR (Object A → rightward saccade), AL (Object A → leftward saccade), BR (Object B → rightward saccade), and BL (Object B → leftward saccade). The colours used in the inset match those of the histograms and are keyed by the legend. The top left plot corresponds to an object-selective neuron (selective to object B), the bottom left plot corresponds to a response-selective neuron (selective to rightward response), and the top and bottom right plots correspond to neurons which are selective for a task-specific combination of a particular cue with a particular response (top BR, bottom AL).

In order to analyse theoretically the neurodynamical substrate underlying conditional object-response (associative) learning, we performed numerical simulations of the experiment of Asaad *et al*. (1998)

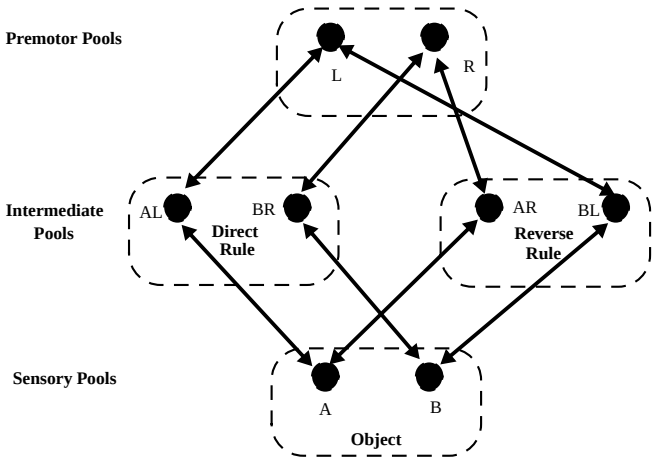

FIG. 4. Sketch of the most relevant recurrent excitatory NMDA and AMPA synaptic connectivity in the prefrontal cortical architecture utilized for the numerical simulations of the experiment of Asaad *et al*. (1998) on conditional object-response learning. All four possible sensory cue–response associations under both task conditions (direct and reversed) are included. In order to simplify the picture, only the connections with increased synaptic strength are shown and we omit the inhibitory and nonselective pools and external influences. The architecture includes: two premotor pools of response neurons, one corresponding with leftward saccade responses (L) and the other corresponding to rightward saccade responses (R); two sensory pools of object-selective neurons, one corresponding to the object A and the other corresponding to the object B; and four intermediate task-specific pools of associative neurons, one for each of the four possible stimulus–response associations. We group these intermediate pools into two groups, one corresponding to the direct condition and the other corresponding to the reverse condition.

utilising the same values for the task parameters and the same measure of neuronal activity (the firing rate). We used the prefrontal cortical architecture introduced in Materials and Methods and indicated for this experiment in Fig. 4. There were two premotor pools of response neurons, one corresponding with leftward saccade responses (L) and the other corresponding to rightward saccade responses (R). There were two sensory pools of object-selective neurons, one corresponding to object A and the other corresponding to object B. Finally, in order to consider all possible cue–response associations, we considered four intermediate task-specific pools of associative neurons, one for each of the four possible combinations of stimuli with responses. We considered these intermediate pools as two groups, one corresponding to the direct object-response condition or rule and the other corresponding to the reverse condition. We assumed that, after training the monkeys, Hebbian learning had formed the connections within the sensory, intermediate and response pools of neurons, and the connections between the different pools, as shown in Fig. 4 for both the direct and reversed conditions, and as specified in Table 1. In order to simplify Fig. 4, only the connections with increased synaptic strength are shown. (As described in Materials and Methods, all neurons were fully connected. Figure 4 does not include the inhibitory and non-selective pools, and the external inputs are not shown. There were no spatial pools for this simulation, because there was no spatial condition in the original neurophysiological experiment.) The excitatory recurrent connections are modelled by NMDA and AMPA receptors. (For the simulation shown in Fig. 4 the forward connections from sensory pools to the intermediate pools were set to have the same value as $w_f = 1.7$ to facilitate global attractors.)

The simulation started with a precue period of 500 ms, in which the network exhibited spontaneous activity [3 Hz for the excitatory pools and 9 Hz for the inhibitory pools as in the experimental recordings of Wilson *et al*. (1994)]. An object stimulus (A or B) was presented next during the cue period of 500 ms (i.e. during this period one of the sensory pools, A or B, received external Poisson spikes with an increased rate from ν_{ext} to $\nu_{ext} + \lambda_{input}$). After the cue period the stimulus was removed and a delay period of 1000 ms followed, until the response period. We assumed that the monkey was performing correctly and therefore that he knew the context, i.e. which rule was active. We modelled this by assuming that all intermediate pools in the group corresponding to the active rule (direct or reversed) received Poisson spikes with an increased rate ($\nu_{ext} + \lambda_{rule}$).

Figures 5 and 6 show the results of the simulations. We emphasise that, during a run of the simulation, the only external inputs to the network are the sensory cues (which activate the correct object and spatial sensory pools) and the bias input to one set of intermediate neurons if the task is being run direct, and to a different set if the task in being run reversed.) Fig. 5 plots the temporal evolution of the averaged population activity for four neural pools. Selective cue-, response- and context-related intermediate neuron activity was maintained by the simulation during the short-term memory-related delay period. As in

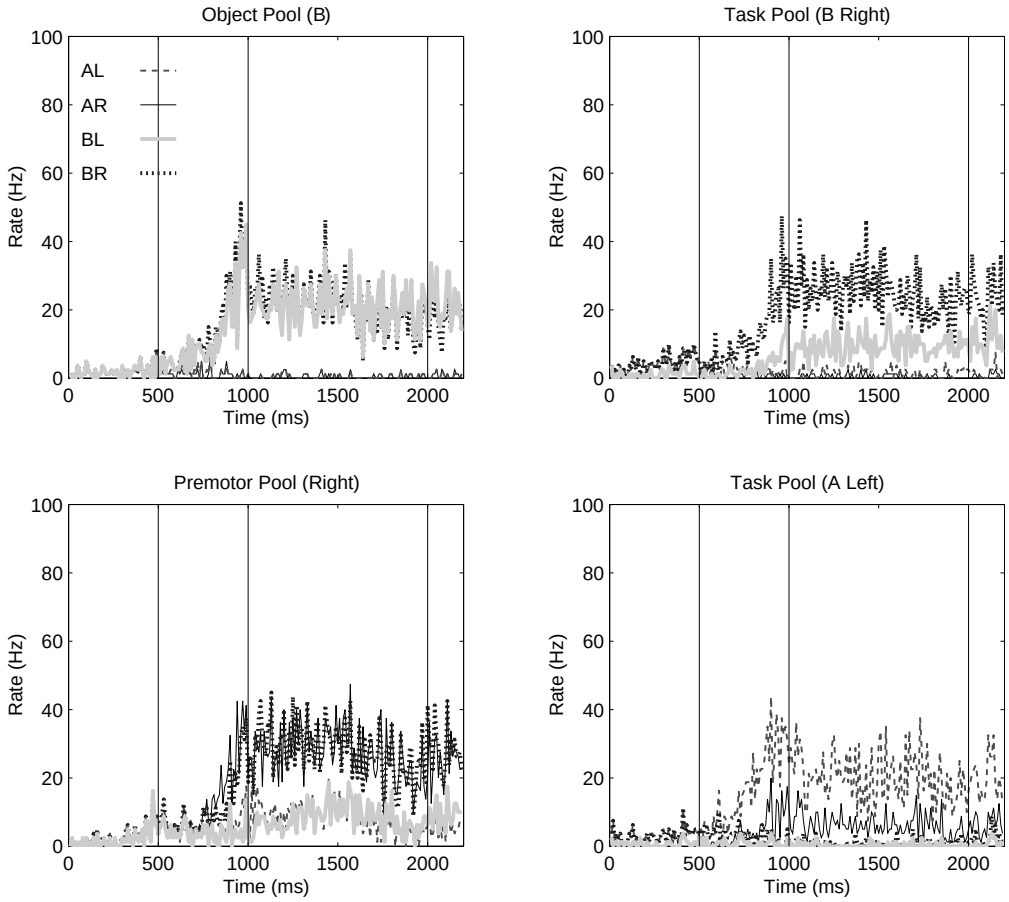

FIG. 5. Temporal evolution of the averaged population activity for four neural pools during the execution of conditional object-responses (associative) learning after the experimental paradigms of Asaad *et al*. (1998). A key to each curve plotted is shown in the top left plot. The top left plot corresponds to an object-selective pool (sensory pool for object B); the bottom left plot corresponds to a response-selective pool (premotor pool R); and the top and bottom right plots show pools which are selective for a given task (top, intermediate pool BR; bottom, intermediate pool AL). All four types of neurons found experimentally by Asaad *et al*. (1998) can be identified with pools in our prefrontal network.

Fig. 3, the top left plot corresponds to an object-selective pool (sensory pool object B), the bottom left plot corresponds to a response-selective pool (premotor pool R), and the top right and bottom right plots correspond to pools which are selective for a task-specific association (top-right intermediate pool BR, bottom-right intermediate pool AL). Thus the types of neuron found experimentally by Asaad *et al*. (1998) can be identified with neuronal pools in our prefrontal network. The exact time at which the bifurcation appeared in the neuronal firing between the two conditions (see Fig. 5, bottom right) depended on the level of inhibition.

Figure 6 plots the rastergrams of randomly selected neurons for each pool in the network (five for each sensory, intermediate and premotor pool, 20 for the nonselective excitatory pool and 10 for the inhibitory pool). The spatio-temporal spiking activity shows that during the short-term memory delay period only the neurons representing the sensory cue, the associated future oculomotor response and the associated intermediate neurons maintained persistent activity and built up a stable global attractor in the network which maintained the firing during the delay period. This specific global attractor, corresponding to a specific cue-response-context condition, incorporates several single pool attractors, with one for the stimulus input, one for the intermediate neurons which represent the relevant combination and one for the response. The cue stimulus, and the biasing context top-down rule information applied to the intermediate neurons, drive the system into the corresponding global attractor utilising biased competition mechanisms.

Rule-dependent object-response or delayed spatial response tasks

We consider now the theoretical analyses of the single-cell recordings of PFC neurons of Asaad *et al*. (2000), in which monkeys were trained to perform either an object-response task with a delay (as in the preceding subsection), or a delayed spatial response task.

The conditional object-response task with a delay was, as in the preceding subsection, defined by the association of the identity of an object (O1 or O2) with a saccade response (L or R), independently of the location of the cue object. On the other hand, the spatial delayed response task required the monkey to make a saccade (L or R) response after a delay towards the location at which the cue object was presented (S1 or S2). Under this second condition the monkey had to ignore the

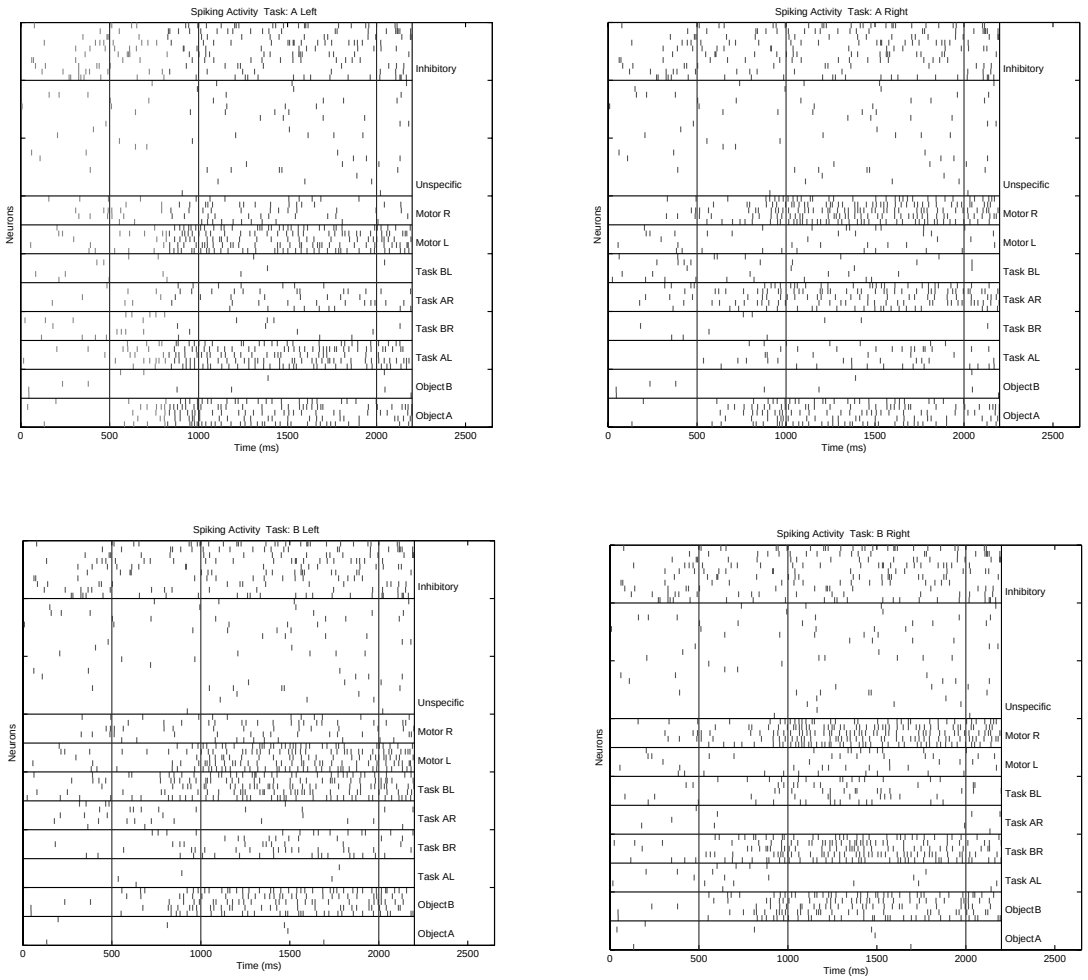

FIG. 6. Rastergrams of randomly selected neurons for each pool in the PFC network for all task conditions in the simulation of the experimental paradigm of Asaad *et al*. (1998), which involved object-response association learning and its reversal. There were five neurons selected randomly for each sensory pool representing Object A or Object B; five for the intermediate pools representing Object A Left Response, etc.; five for the premotor pools representing Right or Left response; 20 for the nonselective excitatory pool; and 10 neurons for the inhibitory pool. Upper left and bottom right panels, Object A requires a Left response, and Object B a Right Response. Lower left and upper right panels, the same task run reversed. The spatio-temporal spiking activity shows that during the short-term memory delay period only the sensory cue, associated future oculomotor response and intermediate neurons maintain persistent activity and build up the stable global attractor of the network. The underlying biased competition mechanisms are evident in the neuronal firing.

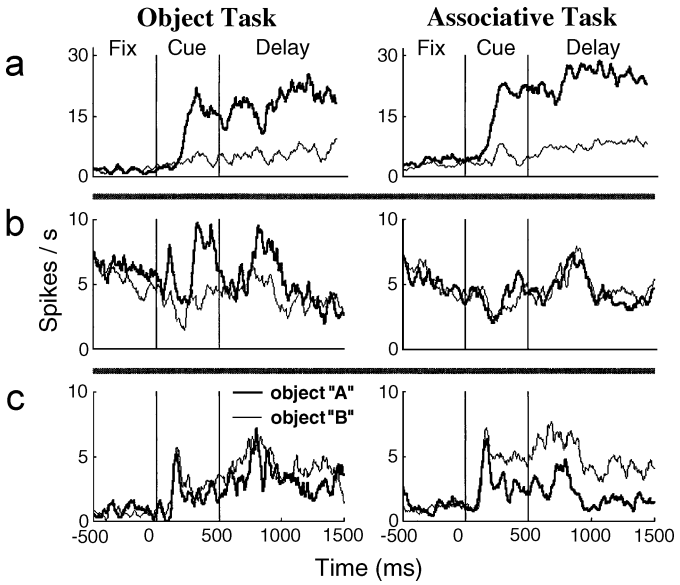

FIG. 7. Experimental results of Asaad *et al*. (2000). The curves plot the averaged firing activity over time for three neurons beginning 500 ms before the cue presentation (500 ms) and during the delay period of 1000 ms. The left panel corresponds to the delayed spatial response task condition and the right panel to the conditional object-response task condition. Each picture plots two curves corresponding to the two possible responses, ipsilateral or contralateral. The first row, a, shows a neuron which was direction-selective in both tasks. The second row, b, shows a neuron which was direction-selective in only the spatial task. The third row, c, shows a neuron which was direction-selective in only the conditional object-response task. (Reproduced with permission from Asaad *et al*., 2000.)

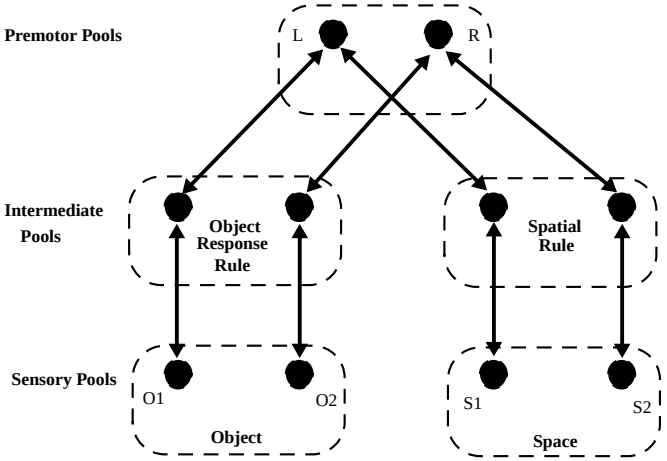

FIG. 8. Sketch of the most relevant recurrent excitatory NMDA and AMPA synaptic connectivity in the prefrontal cortical architecture utilized for the numerical simulations of the experiment of Asaad *et al*. (2000). The figure shows the connectivity after learning corresponding to cue–response associations under both object-response and delayed spatial response task conditions. Only the connections with increased synaptic strength are shown. The PFC architecture includes two premotor pools of response neurons, one corresponding with leftward saccade responses (L) and the other corresponding to rightward saccade responses (R); four sensory pools (two object-selective neuronal pools, one corresponding to object O1 and the other corresponding to object O2; and two spatially selective sensory neuronal pools, one corresponding to location S1 and the other corresponding to location S2); and four intermediate task-specific pools of associative neurons, one for each of the four possible combinations. The intermediate pools are grouped in two groups, one corresponding to the object-response (associative task) and the other corresponding to the delayed spatial response task.

feature characteristics of the object and allocate its attention and memory to process the spatial location of the stimulus. Figure 7 presents the experimental results of Asaad *et al*. (2000). The curves plot the averaged firing activity over time for three neurons beginning 500 ms before the cue presentation (which lasted for 500 ms) and during a delay period of 1000 ms. The left panel corresponds to the delayed spatial response task condition and the right panel to the conditional object-response condition. Each picture plots two curves corresponding to the two possible response directions, ipsilateral or contralateral. Row (a) shows a neuron which was response-direction-selective in both tasks. Row (b) shows a neuron which was direction-selective in only the spatial task. Row (c) shows a neuron which was direction-selective in only the object-response task. These results demonstrate that the information represented by neurons in the lateral PFC of primates was not limited to discrete sensory events or motor responses, but instead that the behavioural context in which the animals were engaged had a decisive influence on the activity of some of the neurons. In particular, some populations of neurons responded to combinations of the task being performed and the response which was required. In this way, PFC neurons provided a neural substrate for responding appropriately on the basis of an abstract rule or context. (The monkey determined which rule was operating in a block of trials by discovering whether reward was obtained by performing according to the object-response or delayed spatial-response rule. It should be emphasised that the cue stimuli presented, and the responses made, were identical in both tasks.)

These phenomena and their underlying neurodynamical mechanisms can be thoroughly studied and explained in our theoretical framework. We performed numerical simulations of the experiment of Asaad *et al*. (2000) by means of the prefrontal cortical architecture shown in Fig. 8. Again, we assumed that, during the training of the monkeys, Hebbian learning built up the associative synaptic connections between the sensory, intermediate and response pools which were coactive during individual trials of the task. Figure 8 shows the connectivity after learning in both tasks. Only the connections with increased synaptic strength are shown. These excitatory recurrent connections are modelled by NMDA and AMPA receptors. The PFC architecture includes two premotor pools of response neurons, one corresponding to leftward saccade responses (L) and the other corresponding to rightward saccade responses (R); four sensory pools (two object-selective neuronal pools, one corresponding to object O1 and the other corresponding to object O2, and two pools with selectivity for the spatial location of the stimulus, one corresponding to location S1 and the other corresponding to location S2), and four intermediate neuronal pools, one for each of the four possible stimulus–response combinations. The intermediate pools are considered as being in two groups, one for the object-response associative task and the other for the delayed spatial response task. The intermediate neurons receive a biasing signal, which is stronger if the relevant task for that intermediate pool is being performed. The values for all the synaptic connections are shown in Table 2. The forward synapses (cue-intermediate-response) are stronger than the backprojection synapses.

The simulation used similar parameters to those used in the original experiment. After a precue period of 500 ms a cue object (O1 or O2) was presented at a particular location (S1 or S2) during the cue period of 500 ms (During this period the corresponding object sensory pool, O1 or O2, and spatial sensory pool, S1 or S2, received external Poisson spikes with an increased rate ν_{ext} to $\nu_{ext} + \lambda_{input}$). During the delay period of 1000 ms the stimulus was removed. We assume that the monkey was performing correctly and therefore that he knew the context, i.e. which rule was active. The active rule was encoded by the external rule-specific input ($\nu_{ext} + \lambda_{rule}$) received by the group of

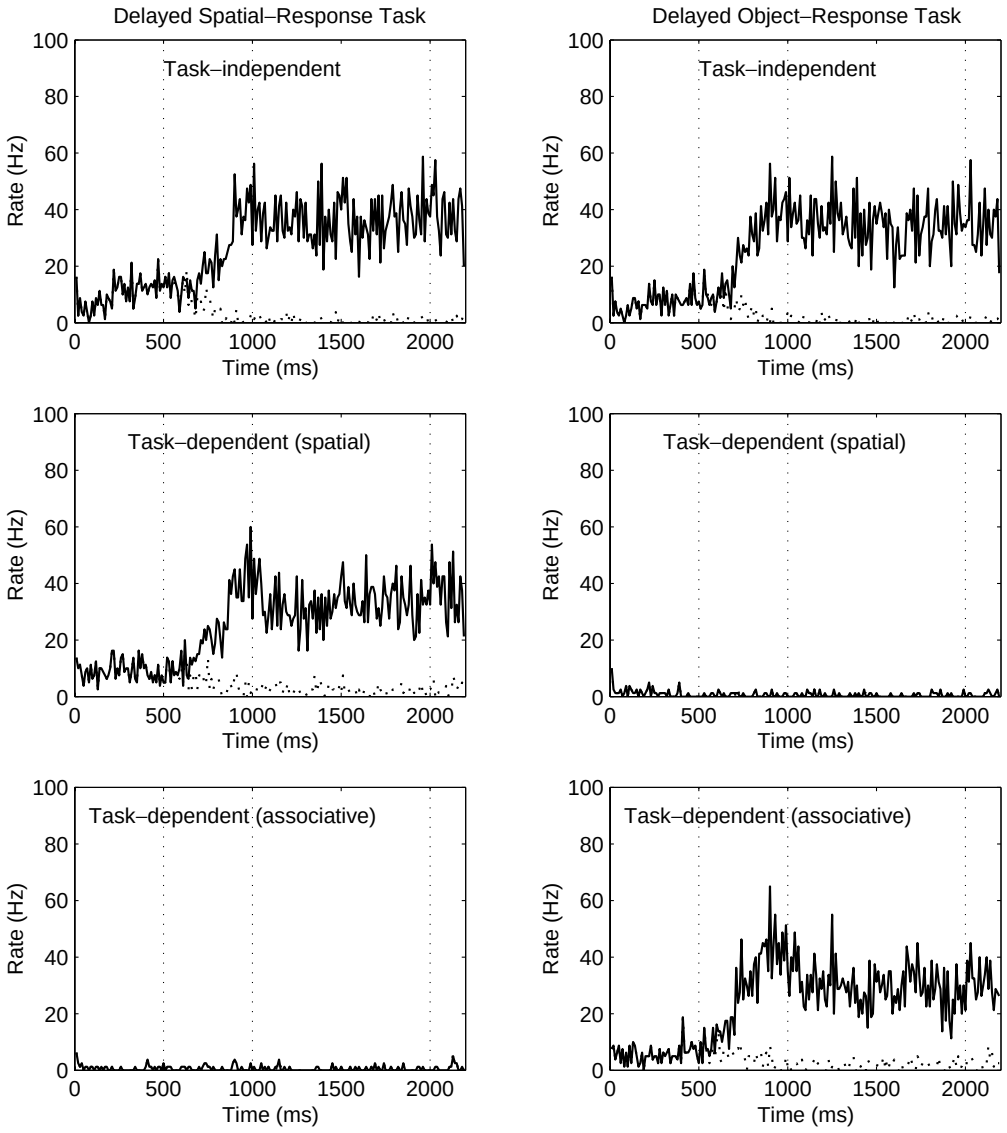

FIG. 9. Temporal evolution of the averaged population activity for three neural pools during the execution of the object-response (associative) and delayed spatial response task after the experimental paradigms of Asaad *et al*. (2000). The figure shows three neural pools, namely the premotor pool 'L', the intermediate spatial pool S1-L, and the intermediate associative pool O1-L. Cue, response and selective context-specific associative activity is explicitly maintained during the short-term memory related delay period. The left panel corresponds to the delayed spatial response task condition and the right panel to the object-response (associative) task condition. Each picture plots two curves corresponding to the two possible responses (dark corresponds to L and light to R). The first row shows the premotor pool 'L' which was response-direction-selective in both tasks. The second row shows the intermediate spatial pool S1-L which was direction-selective (to the L, dark curve) in only the delayed spatial response task. The third row shows the intermediate associative pool O1-L which was direction-selective in only the object-response (associative) task. All three types of neurons found experimentally by Asaad *et al*. (2000) can be identified with pools in our prefrontal network.

intermediate pools which corresponds to the rule for the task being performed (object-response associative, or delayed spatial response).

Figures 9 and 10 show the results of the simulations. Figure 9 plots the temporal evolution of the averaged population activity for three neural pools, namely the premotor pool 'L', the intermediate spatial pool 'S1-L', and the intermediate associative pool 'O1-L'. Cue, response and selective context-specific associative activity was explicitly maintained during the short-term memory-related delay period by the recurrent connections. As in Fig. 7, the left panel corresponds to the delayed spatial response condition and the right panel to the conditional object-response associative task condition. Each graph shows two curves corresponding to the two possible response directions (dark corresponds to L and light to R). The first row shows activity in the premotor pool 'L' which was response-direction-selective in both tasks. The second row shows activity in the intermediate spatial pool 'S1-L' which was response-direction-selective (to the L, dark curve) in only the delayed spatial response task. The third row shows activity in the intermediate associative pool 'O1-L' which was direction-selective in only the conditional object-place associative task. All the three types of neurons found experimentally by Asaad *et al*. (2000) can be identified with pools in our prefrontal network.

Figure 10 plots the rastergrams of randomly selected neurons for each pool in the network (five for each sensory, intermediate and premotor pool, 20 for the nonselective excitatory pool and 10 for the inhibitory pool). The spatio-temporal spiking activity shows that during the short-term memory delay period only the relevant sensory cue, associated future oculomotor response and intermediate neurons maintained persistent activity and built up a stable global attractor in

the network. The underlying biased competition mechanisms were very explicit in this experiment. Note that neurons in pools for the irrelevant input sensory dimension (location for the object-response associative task, and object for the delayed spatial response task), were inhibited during the sensory cue period and were not sustained during the delay short-term memory period. Only the relevant single pool attractors, given the rule context, which were suitable for the cue-response mapping survived the competition and were persistently maintained with high firing activity during the short-term memory delay period. This suppression effect has been recently observed by Everling *et al*. (2002) by recording the activity of prefrontal neurons in monkeys carrying out a focused attention task. In their spatial cueing task, they observed strong filtering of the PFC response to unattended targets (which is similar to that observed in our simulations for irrelevant sensory dimensions). These attentional modulation effects (relative enhancement of neuronal response to an attended stimulus, and a relative suppression of the neuronal response to an unattended stimulus) are well-known in posterior areas of the visual system, including the striate and prestriate cortex (Reynolds *et al*., 1999), parietal cortex (Bushnell *et al*., 1981) and inferotemporal cortex (Chelazzi, 1998). Our previous (Rolls & Deco, 2002) and present computational simulations suggest that, in the PFC, filtering of ignored inputs may reach a level commensurate with the strong, global effects of selective attention in human behaviour, and that this selection in the PFC is the basis of the attentional modulation found in more posterior sensory cortical areas, implemented through backprojections from the PFC to the more posterior cortical areas. In other words, we see in the same set of networks in the PFC both a kind of internal attentional mechanism which selects the relevant input dimensions for the present behavioural or task condition using competition implemented through the inhibitory neurons in an attractor network, and also a mechanism which maintains this information in short-term memory using the recurrent connections between the neurons in the attractor networks. Our neurodynamical architecture of the PFC therefore unifies attentionally biased competition and short-term memory mechanisms implemented by attractor networks with recurrent connections.

The set of parameters found for this architecture shows that if the system works in a biased competition regime, the experimental results

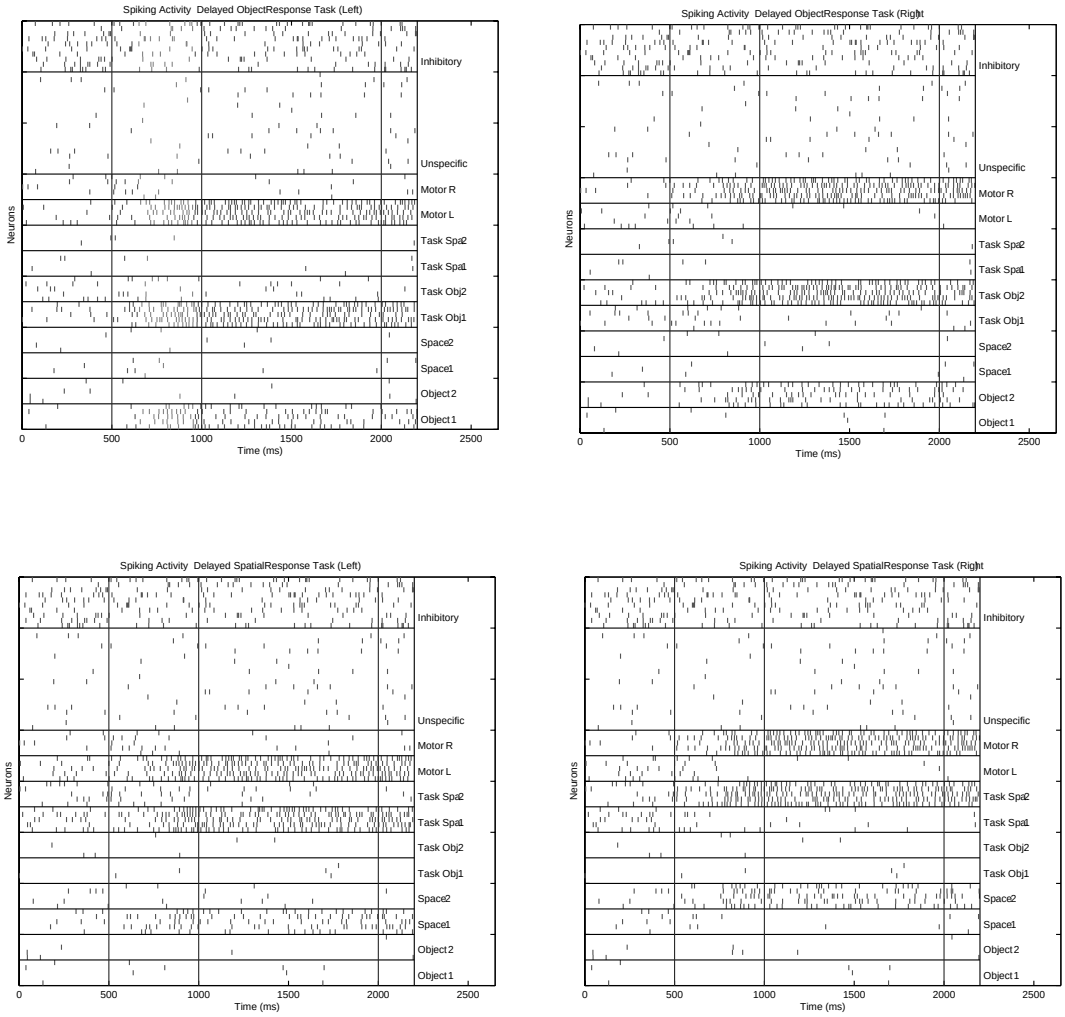

FIG. 10. Rastergrams of randomly selected neurons for each pool in the PFC network (five for each sensory, intermediate and premotor pool, 20 for the nonselective excitatory pool and 10 for the inhibitory pool) and for all task conditions after the experimental paradigms of Asaad *et al*. (2000). The spatio-temporal spiking activity shows that during the short-term memory delay period only the sensory cue, associated future oculomotor response and intermediate neurons maintain persistent activity and build up the stable global attractor of the network. The underlying biased competition mechanisms are explicit. We note that there are our four combinations of stimuli (Object 1–Space 1, Object 1–Space 2, Object 2–Space 1 and Object 2–Space 2) and two context rules, leading to eight conditions. All eight conditions were run with random combinations, and we show in the Figure just four of the possible combinations as rastergrams in order to simplify the Figure, and because consistent results were obtained in the other conditions and do not provide additional information.

can be explained. The simulations also predict the type of activity expected in the inhibitory neurons in the network (see Fig. 10) which was not studied in the neurophysiological investigations. The activity of the inhibitory neurons illustrates the point that they implement just general processes such as feedback inhibition, and do not reflect in their firing any factor such as the particular rule being used. Even more, the rastergrams show the interaction between the biasing context/input/response selection process and attention. Because the context/input/response selection results from an attractor of the system, automatically the other competing and also excited pools (such as the spatial pool in the object task) show suppression of activity (because they do not form part of the attractor), which can be interpreted as an attentional effect. In other words, by analysing computationally the underlying dynamicals, we discover the strong relationship between short-term memory, the influence of context, and attention, as the different aspects of the same process, namely the dynamicals working at a biased competition point (i.e. at a stage or point in the dynamical evolution of the system where the bias is influencing the competition to show a bifurcation in a particular direction).

Dopamine-based neural modulation of task-specific neural activity: experimental predictions

The effect of neurotransmitters on short-term memory-related prefrontal neuronal activity has been studied intensively recently, because of its clinical relevance. Dysfunctions of dopamine receptors have been related to working memory deficits in schizophrenia (Goldman-Rakic, 1994; Egan & Weinberg, 1997; Okubo *et al*., 1997). In addition, amphetamine administration increases dopamine levels and is correlated with improved working memory performance of schizophrenic patients (Daniel *et al*., 1991). Experiments with behaving monkeys have found that delay neuronal activity has a bell-shaped curve in response to activation of dopamine receptors (Williams & Goldman-Rakic, 1995). Brunel & Wang (2001) have explained this inverted U-shape dependence of persistent neuronal activity in delay periods of dopamine by considering a detailed model of the influence of dopamine D1 receptor activation on NMDA receptor-mediated EPSPs on pyramidal cells vs. inhibitory interneurons in the PFC. In the PFC Zheng *et al*. (1999) showed that low concentrations of dopamine act through D1 receptors to increase NMDA receptor-activated excitatory currents in pyramidal cells, whereas higher concentrations of dopamine act through D2 receptors to decrease NMDA receptor-activated excitatory currents.

Following the dopamine D1 models utilised by Brunel & Wang (2001), we extended the simulations of the experiments of Asaad *et al*. (2000) described in the last subsection by manipulating the level of dopamine in our PFC architecture. We aimed to study the influence of dopamine levels not just in the context of short-term memory-related neuronal activity in the delay period but also at the more global cognitive level. In particular, we investigated by simulation how dopamine may influence context or rule-dependent stimulus cue-to-response mapping.

Two different models of dopamine influence were simulated, following the approach specified by Brunel & Wang (2001) and the sources cited there and next. The first one consisted of a simultaneous modulation of the NMDA and GABA conductances (g_{NMDA} and g_{GABA}; see synapse equations in the appendix) (Law-Tho *et al*., 1994; Zheng *et al*., 1999). We ran the simulations by scaling down both conductances by a factor of 0.6, which models the effect of increasing D2 receptor activation described by Zheng *et al*. (1999). The second model considered a differential dopamine D1 modulation of NMDA conductances in pyramidal cells and inhibitory interneurons following Muly *et al*. (1998). In this second model which mimics the effect of decreasing D1 receptor activation, the NMDA conductances on pyramidal cells were multiplied by a factor:

$$c_E(1 + 0.2/\{1 + \exp[(0.8 - D_1)/0.25]\}) = 0.91 \text{ for } D_1 = 0.8,$$

and those on inhibitory interneurons are multiplied by:

$$c_I(1 + 0.2/\{1 + \exp[(1.2 - D_1)/0.25]\}) = 0.97 \text{ for } D_1 = 0.8,$$

where D_1 is the relative change of simulated D1 activation, and the constants c_E and c_I are chosen so that both factors are equal to 1 when $D_1 = 1$.

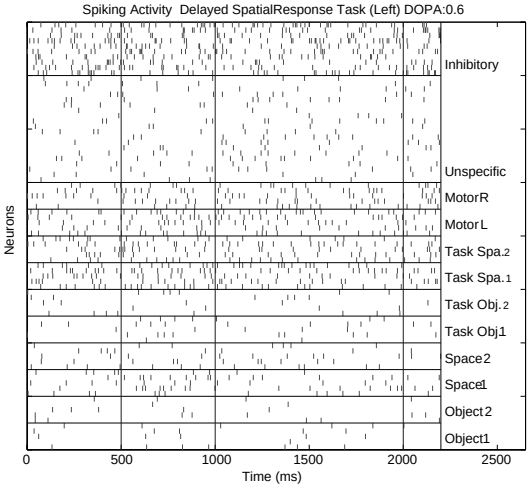

FIG. 11. Rastergrams of simulations of the delayed spatial response task after the experimental paradigm of Asaad *et al*. (2000), by manipulating the level of dopamine in our PFC architecture. The dopamine effect consists of a simultaneous modulation of the NMDA and GABA conductances by a factor 0.6. Note the similar level of activity for the intermediate spatial and both premotor pools for the dopamine condition. These would cause more oculomotor response errors.

Both models, and a third model in which both types of modulation were present simultaneously, yielded the same qualitative effect. Figure 11 shows the results for the first model of increased dopamine D2 receptor activation in which the NMDA-mediated excitatory conductances are decreased. We show a simulation corresponding to the delayed spatial response task. The effects found are not only a decrease in the delay-related short-term memory-related neuronal activity in all task conditions, but also an increase in the firing of the neurons which should not be firing in the particular task. (The neurons which should have a high firing rate in the delayed Left spatial response condition illustrated are the Space 1, Task Space 1, and Motor Left.) This decreased competition arises primarily because the NMDA receptor-activated conductances are decreased, so that the relevant neuronal pools do not enter into an attractor state reflected in high firing. Thus the difference between the neurons which should be firing in the task and those which should not is decreased. Consistent with this, in the rastergrams in Fig. 11, the level of activity of the spatial intermediate neurons and of both premotor neurons is similar, meaning that the external sensory and context biasing were not able to drive the competition in the correct direction to produce a strong attractor. In fact, because the two premotor neurons show similar activity, the monkey will not be able to decide which is the right response in that task and cue situation, and we predict will make more behavioural errors. Figure 12 shows the results for all other conditions in the form of histograms. The same conclusions holds. The behavioural prediction is again that more saccade errors should be made when dopamine is elevated.

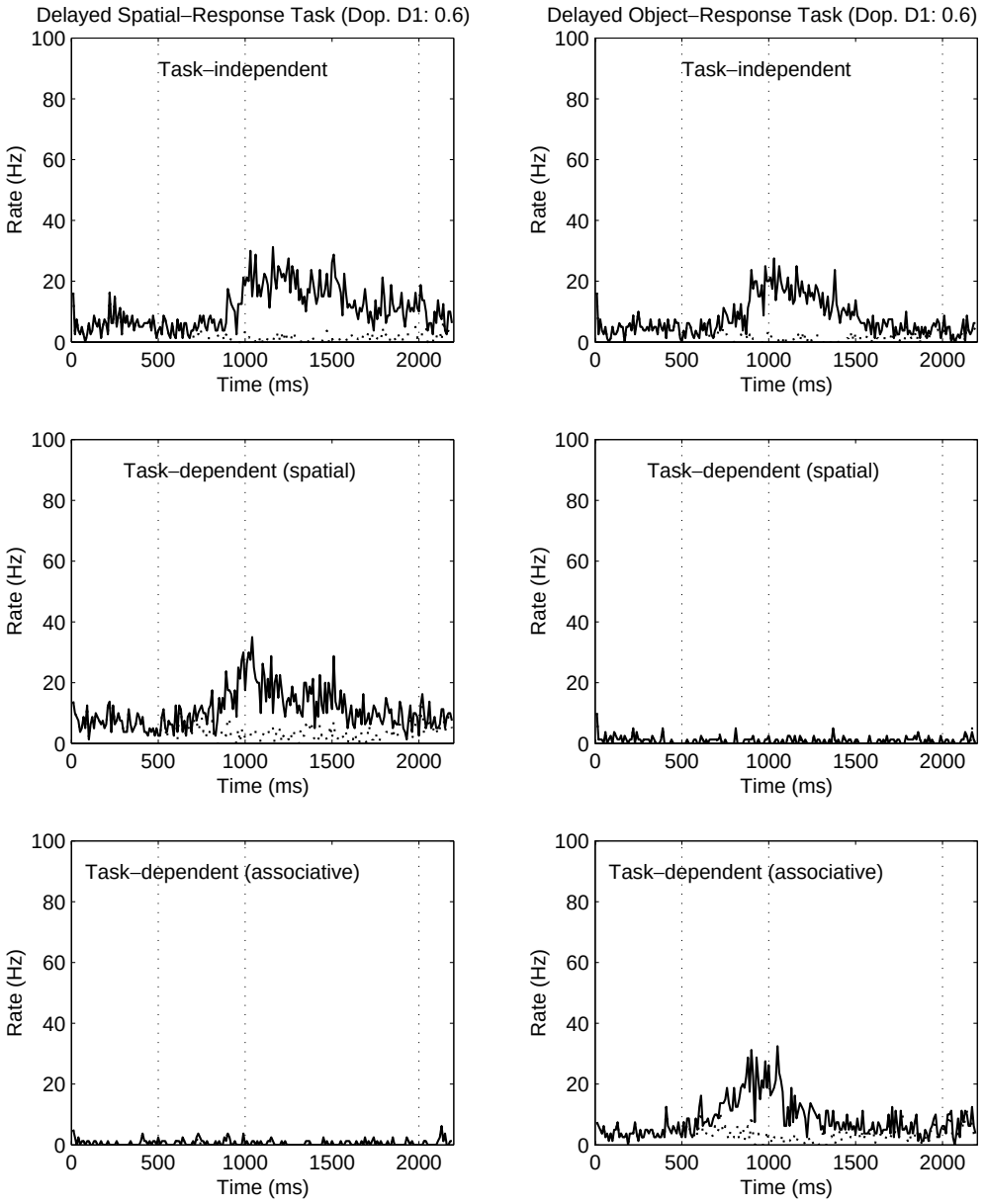

FIG. 12. Temporal evolution of the averaged population activity for four neural pools during the execution of conditional object-response (associative) and delayed spatial response task after the experimental paradigms of Asaad *et al*. (2000), produced by manipulating the level of dopamine in the PFC architecture. The dopamine effect consists of a simultaneous modulation of the NMDA and GABA conductances with a factor 0.6. Note the effect of reduced competition, conventions and arrangement as in Fig. 9.

Discussion

We have presented a detailed theoretical neurodynamical analysis of the spiking and synaptic mechanisms underlying behaviour which requires complex context-dependent mapping between sensory stimuli and actions. Depending on the current rule (or context, or attentional state), the same stimulus can lead to different behaviours, or the same behaviour may be elicited by different cueing stimuli. Neurophysiological investigations have revealed the existence of neurons in the primate PFC showing task-specific firing activity, providing a neural substrate for rule-dependent associations between specific cues and responses, i.e. for mapping stimulus and response in a flexible context-dependent fashion. Our neurodynamical architecture of the PFC unifies attentionally biased competitive mechanisms and recurrent excitatory mechanisms which support short-term memory-related neuronal activity.

We apply the model to explaining in particular the firing of neurons in conditional object-place and delayed spatial response tasks (Asaad *et al*., 2000, 1998), in which different neurons have activity which reflects in the delay period the sensory properties of the stimuli (with tuning to the object or to its location), the responses about to be made or in a rule-dependent way combinations of the stimuli and the motor responses. The whole architecture is implemented with integrate-and-fire neurons (in the theoretical framework of Brunel & Wang, 2001), so that the details of the spiking and synaptic mechanisms involved can be understood and so that predictions can be made about the effects of for example neurotransmitters and pharmacological agents which have particular effects on synaptic transmission. However, the architecture

described is an extension beyond attractor architectures in that different neuronal pools or populations, connected hierarchically, are simulated, and in that an attentional or rule bias is applied which allows the network to select the correct response given the current stimulus and attentional bias.

One of the most important aspects of the architecture is that it incorporates both short-term memory properties and a biasing effect of attention in a single architecture. The short-term memory properties are bestowed on the system by virtue of the recurrent collaterals. This attractor architecture, by now well known (Hopfield, 1982; Amit, 1989; Kohonen, 1995) and implemented with integrate-and-fire neurons which allow the attractors to work with the relatively low neuronal firing rates found neurophysiologically (Brunel & Wang, 2001), enables the short-term memory properties of the network to be produced. (A separate short-term memory attractor network is also needed in order to maintain the activity of the neurons that represent the rule which is currently in operation in a given set of trials.) The short-term memory properties are needed to bridge the delay between the end of the stimuli and the time when the response can be made. More generally, for the short-term memory to be maintained selective with just those neurons which are part of the pattern to be kept firing and the others to remain below the threshold for firing, inhibition (implemented through inhibitory neurons) is of course required. The biasing effect of attention (or the current rule in the conditional tasks) is realised by allowing the bias to affect selected groups of neurons. Because there is global inhibition in the network (implemented through the inhibitory neurons which are required for the attractor properties, and are a feature of cortical architecture (Rolls & Treves, 1998; Rolls & Deco, 2002), the bias applied to some neurons produces competition within the network and the other neurons become less active. (The effects of the bias on these neurons can then affect other neurons in the network by backprojections and forward projections.) In this particular network, the bias is applied to the intermediate neurons [which receive strong connections from the sensory pools, and which have strong connections to the response (output) neurons], and the effect of this rule-dependent bias signal is to enable the correct mapping to be made between the sensory inputs and the motor responses. It is thus, it is suggested, a general feature of the networks which control attention that they utilise both competition and short-term memory. The important role of the PFC in attention, it may therefore be suggested, is that it enables a short-term memory to be kept active despite changing perceptual stimuli. This short-term memory could not be implemented in posterior perceptual networks (in for example the inferior temporal visual cortex or the parietal cortex), because these networks cannot both hold a pattern of neuronal activity in the memory implemented by attractor networks (which can in general maintain active just one pattern of neuronal activity) and also respond to a new incoming perceptual stimulus (Rolls & Deco, 2002, Chapter 12).

The stimulus-to-response mapping implemented by the network described here has two useful features. One is a set of (intermediate) neurons which respond to a combination of the current stimulus and the current rule. In experiment 2 with either object-response or delayed spatial response mapping required, one such intermediate pool responds to 'the rule is to use spatial mapping and left spatial response'. In experiment 1 where the rule is to use either direct object-response mapping or reversed object-response mapping, the intermediate neurons allow the reversal to be implemented by applying a tonic attentional or rule-dependent bias to two intermediate pools of neurons (e.g. to the pools which respond to 'Object A and Left response' and 'Object B and Right Response'). The second property which is useful is a stronger mapping in the forward direction through the network, e.g. from intermediate neurons to response neurons (as is shown in Tables 1 and 2).

The backprojections in the network do serve useful functions. First, and formally, they encourage the network to enter a global attractor state. This state has one function of enabling the activity in the network to be sustained, as is shown in runs not illustrated in which the backprojection connection strengths are reduced. A second, and important, function is that they enable early (e.g. sensory) neuronal pools in the network to show selective attentional effects. For example, in experiment 1 (Object-response task performed directly vs. reversed) as shown in Fig. 6, when the rule is to perform the object-response mapping task reversed (with Object B requiring a Left Response), the effects of the attentional bias applied to the intermediate neurons facilitate the BL and the AR intermediate neurons, and when object B is shown, there is more selective activity in the Object B vs. the Object A neurons (shown at the bottom left of Fig. 6) than would be the case without the feedback from the intermediate to the sensory neurons. Moreover, the modulation by attentional factors of PFC neurons representing sensory stimuli is found experimentally (Everling *et al.*, 2002), and these backprojection connections show how this could be implemented.

Overall, the network has the architecture of a single attractor network with multiple activated populations or pools of neurons. These different pools engage in competitive interactions, are organised with some hierarchy imposed by the asymmetrically strong forward and backward connections, and receive biasing inputs to influence the relative activity of the different pools, thus implementing attention-based or rule-based mapping from sensory inputs to motor outputs.

Another issue is how the connectivity between the different pools is set up during the learning phase. One process, which works for many of the connections, is Hebbian associative learning. However, even in the cases when the forward and backward connections are not identical (evident in Tables 1 and 2 by entries in the table on opposite sides of the diagonal which are not equal), the pattern of the synaptic weights required can still be set up during training by simple associative learning. All that is required even in these cases is some asymmetry in the gain of the forward and backward connections. This could be implemented by the forward and backward connections terminating on different parts of the dendrite, as occurs for connections between cortical areas (see Rolls & Treves, 1998 and Rolls & Deco, 2002). The implication of this would be that there would be a trend through prefrontal cortical areas, from those closer to the sensory input, through areas between the sensory and motor-related areas, to areas with response-related neuronal activity. With the connections through the networks in this direction, there would be stronger connections in the correct direction, because the forward connections are more likely to end on the main part of the dendrites of pyramidal cells and the backprojections are more likely to end on the apical dendrites of cortical pyramidal cells (Rolls & Treves, 1998; Rolls & Deco, 2002). Such a trend, from prefrontal cortical areas which receive from posterior perceptual areas, through regions which are intermediate, through to regions closer to motor output, could in fact be one of the principles of prefrontal cortical connectivity, which would not be inconsistent with what is known about prefrontal connectivity. For example, the orbitofrontal cortex has mainly sensory inputs (with little response-related neuronal activity) (Rolls, 1999), and so does the ventrolateral PFC. The dorsolateral PFC is more of a mixed area, with neurons which respond to combinations of sensory inputs and responses, and where effects of biasing attentional signals are evident (Asaad *et al.*, 2000, 1998). Finally the more dorsal and posterior prefrontal cortical areas may be more closely related to the responses being made, including oculomotor responses (Kandel *et al.*, 2000).

Synaptic plasticity is dependent on the timing of the spikes in the pre- and postsynaptic neuron (Markram *et al*., 1997; Bi & Poo, 2001; Senn *et al*., 2001), and a theoretical and computational analysis of these effects in the context of working memory formation in the PFC has been performed by Fusi and colleagues (Fusi & Mattia, 1999; Fusi *et al*., 2000; Fusi, 2002, 2003). They have shown how Hebbian dynamical learning can cope with both stability of the network states and stability of the learning process. They have shown that a spike time-based learning rule can result in a rate dependent long-term synaptic modification, and that a working memory prefrontal architecture similar to ours (i.e. excitatory pools of neurons strongly connected within a pool, and weakly connected between other excitatory pools, and with a common inhibitory pool) can indeed be formed by this kind of spike time-based learning. Further, there is accumulating evidence (Sjöström *et al*., 2001) that a more realistic description of the protocols for inducing long-term potentiating and long-term depression probably requires a combination of the dependence on spike-timing (to take into account the effects of the backpropagating action potential) and of the dependence on the subthreshold depolarization of the postsynaptic neuron.

Another issue is how the attractor which holds the current rule active is reset to represent the alternative rule during reversal. The resetting must reflect the fact that expected rewards are no longer being obtained. This is detected and represented by a population of neurons in the primate orbitofrontal cortex (Thorpe *et al*., 1983). How this nonreward or error signal could be used to reverse the neurons that represent the current rule will be the topic of a future paper.

The integrate-and-fire implementation of the network enables us to make explicit predictions of the effect of neuromodulation by manipulation of the dopamine level on the conditional object-response and delayed spatial response tasks. In addition to the effect we describe of a decrease in NMDA-related conductances produced by an increase in D2 receptor activation or a decrease in D1 receptor activation which weakens and shortens the persistent neuronal activity in short-term memory periods, we predict more response errors in the object-response task as a consequence of the more similar level of neuronal firing in the competing neuronal pools. The fact that attention and context effects are less able to produce selective and discriminating firing in different neuronal populations involved in working memory in the PFC when dopamine receptor activation is modulated helps to provide a model based on neuronal population response properties which provides insight into why schizophrenic patients have difficulties in working memory tasks, in paying and maintaining attention selectively, and thus more generally in executive functions (Cohen & Servan-Schreiber, 1992).

The architecture for attention and short-term memory described here and implemented in the PFC may have implications for understanding how attention operates in other brain regions. As noted above, the PFC may be especially important in attention because it can implement the short-term memory functions which are characteristic of most attentional tasks, because it receives inputs from posterior cortical regions of the sensory and other inputs which need to compete with each other in the attentional process, because it has the short-term memories which can maintain the rule active, and because it can be influenced by the reward (and nonreward) signals generated in the orbitofrontal cortex. However, once having performed the computations necessary to implement attention, the prefrontal cortical areas are then in an ideal position to provide the attentional biasing signals needed by posterior cortical areas, through the backprojections from the prefrontal cortical areas to the posterior perceptual areas. It is exactly this type of influence which is examined by Rolls & Deco (2002) in their model of interactions between the dorsal and ventral visual systems. In that extensive model (which can account for many types of attention, including visual search guided by an attentional cue specifying either the location or the identity of the to-be-searched-for target visual stimulus), the source of the biasing attentional signal is the PFC. Putting together that model with the present model of how the PFC implements attentional mechanisms is leading towards a unified theory of attentional processes in the brain which can be fully specified mathematically and which performs attentional processes without mystery (Rolls & Deco, 2002).

We wish to emphasize that the type of modelling described here enables one to define precisely a particular functional neuronal architecture with particular sets of inputs and to perform a thorough study of the parameter space. This enables one to show whether the hypothesized architecture can account for the neurophysiological results and, if so, what the parameters are. In the case described here a global attractor network with hierarchically connected sets of sensory, intermediate and motor networks, each composed of spiking neurons, and with a biasing input applied to the intermediate networks, was able to account for many of the details of what was found in neuronal recording. This provides an example of the value of computational neuroscience where a tool has been provided for a thorough analysis of the assumed hypothesis. It may be possible in words to formulate the hypothesis, but it is impossible to analyse the dynamical evolution of the system and consequently its attractors just with words, and instead simulations and theoretical analysis is required. The hypothesis and architecture described here are interesting and sufficiently complex to be worth testing, in that the biased competition operates by biasing some subpopulations of neurons which are part of an attractor network. The use of an attractor network, which requires feedback inhibition which is a crucial part of the network (Amit, 1989; Rolls & Treves, 1998), thus neatly solves both the short-term memory functions which are an important property of the PFC (Goldman-Rakic, 1996; Rolls & Deco, 2002) and are required in the task being performed (for the continuing firing in the delay period), and the competition between the different neuronal populations so that the biased competition can select the correct mapping through the network from input to output. As far as we are aware, this paper introduces for the first time this new hypothesis, and the use of hierarchically organised subpopulations of neurons within a larger global attractor (implemented by asymmetric connections between the different attractor subpopulations), and the implementation of both at the spiking neuron level so that the details of operation of this model can be directly compared with neurophysiological results. We also emphasize that in this network the bias is just a very abstract kind of information, namely the rule which is valid, and not the specific path (combination of sensory-response pools). The specific path is the result of the biased competition dynamicals which combine in a hierarchical attractor network the input, context and motor response.

In conclusion, we show that the same set of networks in the PFC implement both a kind of internal attentional mechanism which selects the relevant input dimensions for the present behavioural or task condition using competition implemented through the inhibitory neurons in an attractor network, and also a mechanism which maintains this information in short-term memory using the recurrent connections between the neurons in the attractor networks. Our neurodynamical architecture of the PFC therefore unifies attentionally biased competition and short-term memory mechanisms implemented by attractor networks with recurrent connections. A feature of the work described here is that several different integrate-and-fire attractor networks are linked hierarchically to implement the tasks being performed. Advantages of the integrate-and-fire implementation level for modelling complex PFC context-dependent processing introduced in this paper

are that (i) realistic spiking dynamicals are produced by the model for direct comparison with the neurophysiological data, (ii) effects on global processes such as memory of transmitters altering the dynamicals of different types of synapse in the brain can be investigated as described here for dopamine and (iii) predictions can be made about how different manipulations of the system affect behaviour.

Abbreviations

AMPA, α-amino-3-hydroxy-5-methylisoxazde-4-propionic acid; C_m, membrane capacitance; δ-pulse, spike; EPSP, excitatory postsynaptic potential; f, fraction of the neurons in any one pool; GABA, γ-aminobutyric acid; IPSP, inhibitory postsynaptic potential; $I_{syn}(t)$, total synaptic current flow to the cell; NMDA, *N*-methyl-D-aspartate; PFC, prefrontal cortex; θ, threshold; $V(t)$, membrane potential; w, synaptic weight.

References

Abeles, A. (1991) *Corticonics*. Cambridge University Press, New York.

Adcock, R., Constable, R., Gore, J. & Goldman-Rakic, P. (2000) Functional neuroanatomy of executive processes involved in dual-task performance. *Proc. Natl Acad. Sci. USA*, **97**, 3567–3572.

Amit, D.J. (1989) *Modeling Brain Function. The World of Attractor Neural Networks*. Cambridge University Press, Cambridge.

Amit, D. (1995) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Behav. Brain Sci.*, **18**, 617.

Amit, D. & Brunel, N. (1997) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, **7**, 237–252.

Asaad, W.F., Rainer, G. & Miller, E.K. (1998) Neural activity in the primate prefrontal cortex during associative learning. *Neuron*, **21**, 1399–1407.

Asaad, W.F., Rainer, G. & Miller, E. (2000) Task-specific neural activity in the primate prefrontal cortex. *J. Neurophysiol.*, **84**, 451–459.

Baddeley, A. (1986) *Working Memory*. Oxford University Press, New York.

Battaglia, F. & Treves, A. (1998) Stable and rapid recurrent processing in realistic autoassociative memories. *Neural Computation*, **10**, 431–450.

Bi, G. & Poo, M.-M. (2001) Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.*, **24**, 139–166.

Brunel, N. & Wang, X. (2001) Effects of neuromodulation in a cortical networks model of object working memory dominated by recurrent inhibition. *J. Computational Neurosci.*, **11**, 63–85.

Bushnell, C., Goldberg, M. & Robinson, D. (1981) Behavioral enhancement of visual responses in monkey cerebral cortex. I. Modulation in posterior parietal cortex related to selective visual attention. *J. Neurophysiol.*, **46**, 755–772.

Butters, N. & Pandya, D. (1969) Retention of delayed-alternation: effect of selective lesions of sulcus principalis. *Science*, **165**, 1271–1273.

Chelazzi, L. (1998) Serial attention mechanisms in visual search: a critical look at the evidence. *Psychol. Res.*, **62**, 195–219.

Chelazzi, L., Miller, E., Duncan, J. & Desimone, R. (1993) A neural basis for visual search in inferior temporal cortex. *Nature (Lond.)*, **363**, 345–347.

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

Corchs, S. & Deco, G. (2002) Large-scale neural model for visual attention: integration of experimental single cell and fMRI data. *Cerebral Cortex*, **12**, 339–348.

Daniel, D., Weinberger, D., Jones, D., Zigun, J., Coppola, R., Handel, S., Bigelow, L., Goldberg, T., Berman, K. & Kleinman, J. (1991) The effect of amphetamine on regional cerebral blood flow during cognitive activation in schizophrenia. *J. Neurosci.*, **11**, 1907–1917.

Deco, G. & Lee, T.S. (2002) A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing*, **44–46**, 769–774.

Deco, G. & Zihl, J. (2001) Top-down selective visual attention: a neurodynamical approach. *Visual Cognition*, **8**, 119–140.

Destexhe, A., Mainen, Z. & Sejnowski, T. (1998) Kinetic models of synaptic transmission. In Koch, C. & Segev, I. (eds), *Methods in Neural Modeling: from Ions to Networks*, 2nd edn. MIT Press, Cambridge, MA, pp. 1–25.

Egan, M. & Weinberg, D. (1997) Neurobiology of schizophrenia. *Curr. Opin. Neurobiol.*, **7**, 701–707.

Everling, S., Tinsley, C., Gaffan, D. & Duncan, J. (2002) Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature Neurosci.*, **5**, 671–676.

Funahashi, S., Bruce, C. & Goldman-Rakic, P. (1989) Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.*, **61**, 331–349.

Funahashi, S., Chafee, M. & Goldman-Rakic, P. (1993) Prefrontal neural activity in rhesus monkeys performing a delayed anti-saccade task. *Nature*, **365**, 753–756.

Fusi, S. (2002) Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. *Biol. Cybernetics*, **87**, 459–470.

Fusi, S. (2003) Spike-driven synaptic plasticity for learning correlated patterns of mean firing rates. *Rev. Neurosciences*, **14**, 73–84.

Fusi, S., Annunziato, M., Badoni, D., Salamon, A. & Amit, D.J. (2000) Spike-driven synaptic plasticity: theory, simulation, VLSI implementation. *Neural Computation*, **12**, 2227–2258.

Fusi, S. & Mattia, M. (1999) Collective behavior of networks with linear (VLSI) integrate and fire neuron. *Neural Computation*, **11**, 633–652.

Fuster, J. (2000) Executive frontal functions. *Exp. Brain Res.*, **133**, 66–70.

Fuster, J. & Alexander, G. (1971) Neuron activity related to short-term memory. *Science*, **173**, 652–654.

Goldman-Rakic, P. (1987) Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In Plum, F. & Mountcastle, V. (eds), *Handbook of Physiology – The Nervous System.* American Physiological Society, Bethesda, Maryland, pp. 373–417.

Goldman-Rakic, P. (1994) Working memory dysfunction in schizophrenia. *J. Neuropsychol. Clin. Neurosci.*, **6**, 348–357.

Goldman-Rakic, P. (1995) Cellular basis of working memory. *Neuron*, **14**, 477–485.

Goldman-Rakic, P. (1996) Regional and cellular fractionation of working memory. *Proc. Natl Acad. Sci. USA*, **93**, 13473–13480.

Hansel, D., Mato, G., Meunier, C. & Neltner, L. (1998) On numerical simulations of integrate-and-fire neural networks. *Neural Computation*, **10**, 467–483.

Hebb, D. (1949) *The Organization of Behavior – a Neurophysiological Theory.* John Wiley, New York.

Hestrin, S., Sah, P. & Nicoll, R. (1990) Mechanisms generating the time course of dual component excitatory synaptic currents recorded in hippocampal slices. *Neuron*, **5**, 247–253.

Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Nat. Acad. Sci. USA*, **79**, 2554–2558.

Hoshi, E., Shima, K. & Tanji, J. (1998) Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. *J. Neurophysiol.*, **80**, 3392–3397.

Jacobsen, C.F. (1935) Functions of the frontal association areas in primates. *Arch. Neurol. Psychiatry*, **33**, 558–569.

Jahr, C. & Stevens, C. (1990) Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *J. Neurosci.*, **10**, 3178–3182.

Kandel, E.R., Schwartz, J.H. & Jessell, T.M. , (2000) *Principles of Neural Science*, 4th edn. McGraw-Hill, New York.

Kohonen, T. (1995) *Self-Organizing Maps*. Springer-Verlag, Berlin.

Kubota, K. & Niki, H. (1971) Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J. Neurophysiol.*, **34**, 337–347.

Law-Tho, D., Hirsch, J. & Crepel, F. (1994) Dopamine modulation of synaptic transmission in rat prefrontal cortex: An in vitro electrophysiological study. *Neurosci. Res.*, **21**, 151–160.

Leung, H., Gore, J. & Goldman-Rakic, P. (2002) Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *J. Cogn. Neurosci.*, **14**, 659–671.

Levy, R. & Goldman-Rakic, P. (1999) Executive frontal functions. *J. Neurosci.*, **19**, 5149–5158.

Lisman, J.E., Fellous, J.M. & Wang, X.J. (1998) A role for NMDA-receptor channels in working memory. *Nature Neurosci.*, **1**, 273–275.

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

McCormick, D., Connors, B., Lighthall, J. & Prince, D. (1985) Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons in the neocortex. *J. Neurophysiol.*, **54**, 782–806.

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

Miller, E., Gochin, P. & Gross, C. (1993) Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Res.*, **616**, 25–29.

Milner, B. (1963) Effects of different brain lesions on card sorting. *Arch. Neurol.*, **9**, 100–110.
Moran, J. & Desimone, R. (1985) Selective attention gates visual processing in the extrastriate cortex. *Science*, **229**, 782–784.
Motter, B. (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.*, **70**, 909–919.
Muly, E., Szigeti, K. & Goldman-Rakic, P. (1998) D1 receptor in interneurons of macaque prefrontal cortex: Distribution and subcellular localization. *J. Neurosci.*, **18**, 10553–10565.
Okubo, Y., Suhara, T., Suzuki, K., Kobayashi, K., Inoue, O., Terasaki, O., Someya, Y., Sassa, T., Sudo, Y., Matsushima, E., Ito, M., Tateno, Y. & Toru, M. (1997). Decreased prefrontal dopamine D1 receptors in schizophrenia revealed by PET. *Nature*, **385**, 634–636.
Reynolds, J., Chelazzi, L. & Desimone, R. (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J. Neurosci.*, **19**, 1736–1753.
Reynolds, J. & Desimone, R. (1999) The role of neural mechanisms of attention in solving the binding problem. *Neuron*, **24**, 19–29.
Rolls, E.T. (1999) *The Brain and Emotion*. Oxford University Press, Oxford.
Rolls, E.T. & Deco, G. (2002) *Computational Neuroscience of Vision*. Oxford University Press, Oxford.
Rolls, E.T. & Treves, A. (1998) *Neural Networks and Brain Function*. Oxford University Press, Oxford.
Salin, P. & Prince, D. (1996) Spontaneous GABA-A receptor mediated inhibitory currents in adult rat somatosensory cortex. *J. Neurophysiol.*, **75**, 1573–1588.
Senn, W., Markram, H. & Tsodyks, M. (2001) An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Computation*, **13**, 35–67.
Sjöström, P.J., Turrigiano, G.G. & Nelson, S.B. (2001) Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, **32**, 1149–1164.
Spitzer, H., Desimone, R. & Moran, J. (1988) Increased attention enhances both behavioral and neuronal performance. *Science*, **240**, 338–340.
Spruston, N., Jonas, P. & Sakmann, B. (1995) Dendritic glutamate receptor channel in rat hippocampal CA3 and CA1 pyramidal neurons. *J. Physiol. (Lond.)*, **482**, 325–352.
Thorpe, S.J., Rolls, E.T. & Maddison, S. (1983) Neuronal activity in the orbitofrontal cortex of the behaving monkey. *Exp. Brain Res.*, **49**, 93–115.
Ungerleider, L., Courtney, S. & Haxby, J. (1998) A neural system for human visual working memory. *Proc. Natl Acad. Sci. USA*, **95**, 883–890.
Wallis, J., Anderson, K. & Miller, E. (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature*, **411**, 953–956.
Wang, X. (1999) Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *J. Neurosci.*, **19**, 9587–9603.
White, I. & Wise, S. (1999) Rule-dependent neuronal activity in the prefrontal cortex. *Exp. Brain Res.*, **126**, 315–335.
Williams, G. & Goldman-Rakic, P. (1995) Modulation of memory fields by dopamine D1 receptors in the prefrontal cortex. *Nature*, **376**, 572–575.
Wilson, F.A.W., O'Scalaidhe, S.P. & Goldman-Rakic, P. (1994) Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. *Proc. Natl Acad. Sci. USA*, **91**, 4009–4013.
Xiang, Z., Huguenard, J. & Prince, D. (1998) GABA-A receptor mediated currents in interneurons and pyramidal cells of rat visual cortex. *J. Physiol. (Lond.)*, **506**, 715–730.
Zarahn, E., Aguirre, G. & D'Esposito, M. (2000) Replication and further studies of neural mechanisms of spatial mnemonic processing in humans. *Cogn. Brain Res.*, **9**, 1–17.
Zheng, P., Zhang, X.-X., Bunney, B.S. & Shi, W.-X. (1999) Opposite modulation of cortical N-methyl-D-aspartate receptor-mediated responses by low and high concentrations of dopamine. *Neuroscience*, **91**, 527–535.

Appendix

In this appendix we give the mathematical equations which describe the spiking activity and synapse dynamics in the network, following in general the formulation described by Brunel & Wang (2001). Each neuron is described by an integrate-and-fire model. The subthreshold membrane potential $V_{(t)}$ of each neuron evolves according to the following the equation:

$$C_m \frac{dV(t)}{dt} = -g_m(V(t) - V_L) - I_{syn}(t) \quad (1)$$

where $I_{syn}(t)$ is the total synaptic current flow into the cell. When the membrane potential $V_{(t)}$ reaches the threshold θ a spike is generated and the membrane potential is reset to V_{reset}. The neuron is unable to spike during a period τ_{ref} which is the absolute refractory period.

The total synaptic current is given by the sum of glutamatergic excitatory components (NMDA and AMPA) and inhibitory components (GABA). As we described above, we consider that external excitatory contributions are produced through AMPA receptors ($I_{AMPA,ext}$) while the excitatory recurrent synapses are produced through AMPA and NMDA receptors ($I_{AMPA,rec}$ and $I_{NMDA,rec}$). The total synaptic current is therefore given by:

$$I_{syn}(t) = I_{AMPA,ext}(t) + I_{AMPA,rec}(t) + I_{NMDA,rec}(t) + I_{GABA}(t) \quad (2)$$

where

$$I_{AMPA,ext}(t) = g_{AMPA,ext}(V(t) - V_E) \sum_{j=1}^{N_{ext}} s_j^{AMPA,ext}(t) \quad (3)$$

$$I_{AMPA,rec}(t) = g_{AMPA,rec}(V(t) - V_E) \sum_{j=1}^{N_E} w_j s_j^{AMPA,rec}(t) \quad (4)$$

$$I_{NMDA,rec}(t) = \frac{g_{NMDA}(V(t) - V_E)}{(1 + C_{Mg^{++}} \exp(-0.062V(t)/3.57))} \sum_{j=1}^{N_E} w_j s_j^{NMDA}(t) \quad (5)$$

$$I_{GABA}(t) = g_{GABA}(V(t) - V_I) \sum_{j=1}^{N_I} s_j^{GABA}(t) \quad (6)$$

In the preceding equations $V_E = 0$ mV and $V_I = -70$ mV. The synaptic strengths w_j are specified in Materials and methods and in Tables 1 and 2. The fractions of open channels s are given by:

$$\frac{ds_j^{AMPA,ext}(t)}{dt} = -\frac{s_j^{AMPA,ext}(t)}{\tau_{AMPA}} + \sum_k \delta(t - t_j^k) \quad (7)$$

$$\frac{ds_j^{AMPA,rec}(t)}{dt} = \frac{s_j^{AMPA,rec}(t)}{\tau_{AMPA}} + \sum_k \delta(t - t_j^k) \quad (8)$$

$$\frac{ds_j^{NMDA}(t)}{dt} = \frac{s_j^{NMDA}(t)}{\tau_{NMDA,decay}} + \alpha x_j(t)(1 - s_j^{NMDA}(t)) \quad (9)$$

$$\frac{dx_j(t)}{dt} = \frac{x_j(t)}{\tau_{NMDA,rise}} + \sum_k \delta(t - t_j^k) \quad (10)$$

$$\frac{ds_j^{GABA}(t)}{dt} = \frac{s_j^{GABA}(t)}{\tau_{GABA}} + \sum_k \delta(t - t_j^k) \quad (11)$$

where the sums over k represent a sum over spikes emitted by presynaptic neuron j at time t_j^k. The value of α is $0.5\,\text{ms}^{-1}$.

The values of the conductances (in nS) for pyramidal neurons were $g_{AMPA,ext} = 2.08$, $g_{AMPA,rec} = 0.052$, $g_{NMDA} = 0.164$ and $g_{GABA} = 0.65$, and for interneurons $g_{AMPA,ext} = 1.62$, $g_{AMPA,rec} = 0.0405$, $g_{NMDA} = 0.129$ and $g_{GABA} = 0.49$.