

# Synaptic and Spiking Dynamics underlying Reward Reversal in the Orbitofrontal Cortex

Gustavo Deco<sup>1</sup> and Edmund T. Rolls<sup>2</sup>

<sup>1</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA) Universitat Pompeu Fabra Dept. of Technology Computational Neuroscience Passeig de Circumval.lació, 8, 08003 Barcelona, Spain and <sup>2</sup>University of Oxford Department of Experimental Psychology South Parks Road Oxford OX1 3UD, UK

**Cognitive and emotional flexibility involve a coordinated interaction between working memory, attention, reward expectations, and the evaluation of rewards and punishers so that behaviour can be changed if necessary. We describe a model at the integrate-and-fire neuronal level of the synaptic and spiking mechanisms which can hold an expectation of a reward rule in working memory, and can reverse the reward rule if expected rewards are not obtained. An example of a reward rule is that stimulus 1 is currently associated with reward, and stimulus 2 with punishment. The attractor-based reward rule working memory incorporates a spike-frequency synaptic adaptation mechanism which supports the neural switching between rules by being shut down by a general inhibitory input produced by punishment, so that when the attractor starts up again is in the opposite state. The mechanism can implement one-trial reward reversal, which is a property of orbitofrontal cortex neurons. We show how this reward rule input can operate in a biased competition way to influence which one of two stimuli is currently associated with reward and which with punishment, and to map the stimuli correctly to the reward or punishment representations, providing a basis for action selection required to obtain the reinforcer.**

**Keywords:** attention, computational neuroscience, emotion, orbitofrontal cortex, reversal learning set, reward reversal, working memory

## Introduction

Higher brain functions, such as *cognitive flexibility*, require associative cortical areas that mediate the coordination between working memory, attention, whether expected rewards are obtained, and the alteration of behaviour if the reinforcers are not obtained as expected. Brain regions such as the orbitofrontal cortex, amygdala, and anterior cingulate cortex have been implicated in this remarkable ability of primates to learn associations between sensory stimuli and rewarding or punishing reinforcers that can rapidly and flexibly alter the probability of behaviour (Nauta, 1971; Rolls, 1974, 1975, 1990, 1999, 2000a,b). The ability to respond rapidly to changing reinforcement contingencies is fundamental to an understanding of emotion, given that emotion can be at least operationally defined in terms of states elicited by rewards and punishers (Rolls, 1990, 1999). Accordingly, disorders in this ability to respond rapidly to changing reinforcement contingencies produced by damage to brain regions such as the orbitofrontal cortex may be related to the emotional and behavioural changes that follow alteration of the function of these regions (Damasio, 1994; Rolls *et al.*, 1994; Hornak *et al.*, 2003, 2004). A clear view of the processes that give rise to reward-based context evaluation is important for a better understanding of the behaviour of psychiatric patients.

In order to provide a fundamental basis for the understanding of the neural mechanisms underlying cognitive flexibility, reward context evaluation, and emotion, studies have been performed of the learning of associations between stimuli and rewards or punishers, and the reversal of these stimulus-reinforcement associations. It has been shown in primates that both the orbitofrontal cortex and amygdala are involved in the learning of stimulus-reinforcement associations, but that the orbitofrontal cortex is especially important when these associations must be rapidly reversed. The evidence for this comes from lesion studies (Milner, 1964; Iversen and Mishkin, 1970; Mishkin and Manning, 1978; Rolls *et al.*, 1994; Rolls, 1999; Hornak *et al.*, 2004), and from experiments showing that neurons in the amygdala reflect the reward associations of visual stimuli (Rolls, 2000a) and that neurons in the orbitofrontal cortex not only do this, but also rapidly reverse the visual stimulus to which they respond when the stimulus that is rewarded is reversed (Thorpe *et al.*, 1983; Rolls *et al.*, 1996; Rolls, 2000b). The computational issue that we address here is how this rapid reversal of neurons that code for the reward associations of visual stimuli could occur. This is the general issue of stimulus-reward association learning and reversal. Given that the rewards are stimuli (e.g. a sweet taste is a primary or unlearned reward), this is a type of stimulus-stimulus learning.

In addition, rewards can control the way that stimuli are mapped to responses, as shown, for example, in *conditional object-response tasks*, in which when one object is seen, one response (e.g. a right oculomotor saccade) must be made to obtain reward, and when a second object is seen, a different response (e.g. a left oculomotor saccade) must be made to obtain reward. In this type of task, stimulus-reward association learning alone is insufficient, because each stimulus is equally associated with reward. To account for the performance of and the rapid reversal of this task, and the types of neuron recorded in the prefrontal cortex in this task (Asaad *et al.*, 1998), Deco and Rolls (2003) proposed and investigated a network in which the mapping between the stimuli and the responses could be switched by a rule or contextual input operating to bias competition in stimulus-response combination neurons in an intermediate layer between the sensory inputs and the motor outputs. Deco and Rolls (2003) described an analogous network using biased competition produced by a rule or contextual input that could make a hierarchically organised set of integrate-and-fire networks change the mapping from sensory inputs to motor outputs on the basis of either the conditional object-response rule or a delayed spatial response rule (requiring attention to switch from objects to the spatial position of the objects), accounting for the neurophysiological data obtained in this task by Asaad *et al.* (2000). In modelling these two tasks,

Deco and Rolls (2003) postulated but did not explicitly model the rule or contextual input (acting as a bias) that could be reversed when the reinforcement contingencies in the tasks changed. The second computational issue we address here is how the change in the rewards being received could implement a switch from one rule to another for this rule or contextual representation, to produce a full model of how the changing reinforcement contingencies could switch between the different types of stimulus-to-motor response mapping required in these tasks. We note that the issues we address here are part of the large and important area of reinforcement learning, which is how behaviour is altered on the basis of reinforcement (Sutton and Barto, 1998).

We note that stimulus–reinforcement learning can be implemented by a pattern association network (where the unconditioned stimulus forcing the output neurons to respond is the reinforcer, and the conditioned stimulus becomes associated with this by associatively modifiable synapses) (Rolls and Treves, 1998; Rolls, 1999; Rolls and Deco, 2002). Such a pattern association network could in principle unlearn the association by using associative synapses that incorporate long term depression (Rolls and Treves, 1998; Rolls and Deco, 2002). Although reversal might be implemented by having long-term synaptic depression for synapses that represented the reward-associated stimulus before the reversal, and long-term potentiation of the new stimulus that after reversal is associated with reward, this would require one-trial long-term potentiation and one-trial hetero-synaptic long-term depression (LTD) to account for one-trial stimulus–reward reversal (Thorpe *et al.*, 1983; Rolls *et al.*, 1996; Rolls, 2000b). Moreover, this mechanism would not provide a source for the contextual, rule-based input with persistent (continuing) activity required for the biased competition solution to rapid remapping of stimuli to responses (Deco and Rolls, 2003). We therefore investigate here a mechanism that by utilizing an attractor recurrent autoassociative network (Hopfield, 1982; Rolls and Treves, 1998; Rolls and Deco, 2002) can maintain the rule that is current in a continuing active state of firing until the rule is reversed by punishment or the failure to receive an expected reward. To implement this model in a way that can be compared directly with neurophysiology, the processes occurring at the AMPA, NMDA and GABA synapses are dynamically modelled in an integrate-and-fire implementation to produce realistic spiking dynamics (Brunel and Wang, 2001; Deco and Rolls, 2003). This also enables the synaptic adaptation that is part of the rule reversal mechanism to be realistically implemented. The stimulus to reward part of the model contains different populations of neurons in attractor networks that respond selectively to the sensory stimulus, to a combination of a particular sensory stimulus and whether it currently signifies reward or punishment, and reward, all of which are present in the primate orbitofrontal cortex (Thorpe *et al.*, 1983; Rolls *et al.*, 1996) (see below). The neuronal populations or pools are arranged hierarchically, and have global inhibition through inhibitory interneurons to implement competition. The hierarchical structure is organized within the general framework of the biased competition model of attention (Moran and Desimone, 1985; Spitzer *et al.*, 1988; Chelazzi *et al.*, 1993; Miller *et al.*, 1993; Motter, 1993; Chelazzi, 1998; Reynolds and Desimone, 1999; Rolls and Deco, 2002). Rolls and Deco (2002) added to this framework by introducing a neurodynamically theoretical framework for biased competition, which assumes that multiple activated populations of neurons engage in competitive inter-

actions, and that top-down interactions with other cortical modules bias this competition in favour of specific neurons.

## Methods

### *Visual Discrimination Reward Reversal: Experimental Paradigm*

The basic experimental paradigm that we model in this paper, is the visual discrimination task utilized by Thorpe *et al.* (1983) and Rolls *et al.* (1996). They recorded neuronal activity in the orbitofrontal cortex of macaques while they were performing a Go/NoGo visual discrimination task with reversals. In this task, one of two visual stimuli was presented on each trial in a pseudo-random sequence. One stimulus indicated that if the monkey licked a tube positioned in front of his mouth, he would obtain a reward of fruit juice. The other stimulus indicated that if he licked the tube he would obtain a punishment of a taste of aversive hypertonic saline, and the monkeys learned not to lick when they saw this stimulus. The task is prototypical of emotion-related learning, in that this is stimulus–reinforcer association learning (Rolls, 1999). The monkey's response latencies were 300–400 ms from the onset of the visual presentation of the stimulus to the time of tongue contact with the lick tube. Several different pairs of discriminative stimuli were used, such as a triangle versus a square, a red versus a green square, a vertical versus a horizontal grating, and a syringe from which the monkey could be fed glucose or salt (mildly aversive hypertonic saline). Sometimes reversal was tested by feeding the monkey from a syringe filled with glucose, and then filling it instead with saline Thorpe *et al.* (1983). The monkey learned to reject the syringe (by closing its mouth) within one trial of receiving saline when glucose was expected given the preceding trials.

A reversal of the Go/NoGo visual discrimination was frequently performed in which the meaning of the two visual discriminanda in the task was reversed, so that the previously rewarded stimulus was now negative and vice versa. All the macaques learned over a series of reversals to reverse their behavioural responses quickly, so that if they obtained saline for licking to a stimulus which had previously been associated with reward, they subsequently only licked to the previously punished stimulus which now indicated that reward was available. The acquisition of this ability to reverse very rapidly, in one trial, compared to the 20–50 trials taken the first time that the reinforcement contingencies are reversed, is called the acquisition of a reversal learning set.

Figure 2A shows the reversal of visual responses in a single neuron recorded in the orbitofrontal cortex (Thorpe *et al.*, 1983). The significance of the visual stimulus, a syringe from which the monkey was fed, was altered during the trials. On trials 1–5, no response of the neuron occurred to the sight of the syringe from which the monkey had been given glucose solution to drink from the syringe on the preceding trials. On trials 6–9, the neuron responded to the sight of the same syringe from which he had been given aversive hypertonic saline drink on the preceding trial. Two more reversals (trials 10–15 and 16–17) were performed. The reversal of the neuron's response when the significance of the visual stimulus was reversed shows that the responses of the neuron only occurred to the stimulus when it was associated with aversive saline and not when it was associated with glucose reward. Similar neurons are found using an automated visual discrimination task with two stimuli (Thorpe *et al.*, 1983; Rolls *et al.*, 1996), in which this first class of neuron learns to respond to any visual stimulus associated with reward, and not to respond to any visual stimulus associated with punishment (or vice versa, that is responding to any stimuli associated with taste punishment, but no stimuli associated with taste reward). The responses of these neurons represent the preference of the monkey for a stimulus (a measure of its reward value), in that feeding the monkey to satiety with glucose gradually reduces the response of the neuron and the monkey's preference for the sight of glucose to zero, while leaving the response of the neuron, and the monkey's preference, unaltered for the sight of other food reward-related stimuli (Critchley and Rolls, 1996). In the visual discrimination reversal task, a second class of neuron was found that codes for particular stimuli only if they are associated with reward, and not if

they are associated with punishment. Such a neuron might respond to a green stimulus associated with reward; after reversal not respond to the green stimulus when it was associated with punishment; and not respond to a blue stimulus irrespective of whether it was associated with reward or punishment (Thorpe *et al.*, 1983) (see example in Fig. 2C). The neurons described by Asaad *et al.* (1998) and Asaad *et al.* (2000) in the dorsolateral prefrontal cortex that respond to combinations of particular visual stimuli and responses are analogous to these orbitofrontal cortex neurons that respond to combinations of particular visual stimuli and rewards (Thorpe *et al.*, 1983). A third class of neuron described by Thorpe *et al.* (1983) respond when an expected reward is not obtained, signifying, on these error trials, that a reversal should be made. These neurons frequently responded for many seconds after an error trial. Consistent with this neurophysiology, non-reward, signalling reversal, activates a part of the human orbitofrontal cortex (Kringelbach and Rolls, 2003). All three classes of neuron found neurophysiologically (Thorpe *et al.*, 1983; Rolls *et al.*, 1996) are incorporated into the model we describe.

### Orbitofrontal Architecture

In order to investigate the neurodynamics underlying the rapid stimulus-reward reversal in the context of the findings of Thorpe *et al.* (1983) and Rolls *et al.* (1996), we explicitly model the level of processes occurring at the AMPA, NMDA and GABA synapses in the integrate-and-fire implementation to produce realistic spiking dynamics. We follow the biased competition based neurodynamical framework introduced by the authors (Deco and Zihl, 2001; Corchs and Deco, 2002; Deco and Lee, 2002; Rolls and Deco, 2002) and the integrate-and-fire neuronal framework introduced and studied by Brunel and Wang (2001). We incorporate shunting inhibition (Battaglia and Treves, 1998; Rolls and Treves, 1998) and inhibitory-to-inhibitory cell synaptic connections (Brunel and Wang, 2001) which are useful in maintaining stability of the dynamical system, and incorporate appropriate currents to achieve low firing rates (Amit and Brunel, 1997; Brunel and Wang, 2001). In accordance with the neurophysiological evidence of Thorpe *et al.* (1983) and Rolls *et al.* (1996), we assume in the network architecture investigated the existence of different types of neuronal populations or pools. One type shows object-tuned sensory neuronal responses (selective visual responses). A second type of neuron responds to a combination of a particular object and it being associated with reward, or a particular object and it being associated with punishment. (They can be described as object-and-expected-reward-tuned, or object-and-expected-non-reward-tuned.) A third type of neuron is reward-tuned in that it responds whenever a stimulus is decoded as reward regardless of whether the contingencies are reversed or not; or is punishment-tuned in that it responds whenever a stimulus is decoded as punishment-associated regardless of whether the contingencies are reversed or not. We show that local synaptic connections (which could be set up by development and learning) between these neuronal pools are sufficient for operation of the model. In this section we describe the architecture and operation of the model, and provide a full mathematical specification of the model, and the neuronal parameters used, in the Appendix.

A conceptual overview of the architecture, illustrated in Figure 1, is that in the lower module, stimuli are mapped from sensory neurons (level 1, at the bottom), through an intermediate layer of object-reward combination neurons with rule-dependent activity, to layer 3 which contains Reward/Punishment neurons. The mapping through the intermediate layer can be biased by the rule module inputs to perform a direct or reversed mapping. The activity in the rule module can be reversed by the error signal which occurs when an expected reward is not obtained. [Neurophysiologically, it is found that some error neurons in the primate orbitofrontal cortex respond when instead of an expected reward, nothing is obtained, that is, in extinction; other error neurons respond when instead of an expected reward, punishment such as a drop of saline is obtained, that is in reversal or passive avoidance; and other neurons respond to either type of non-reward (Thorpe *et al.*, 1983; Rolls, 1999).] The reversal occurs because the attractor state in the rule module is shut down by inhibition arising from the effects of the error signal, and restarts in the opposite attractor state because of partial synaptic or neuronal adaptation of the previously active rule neurons.

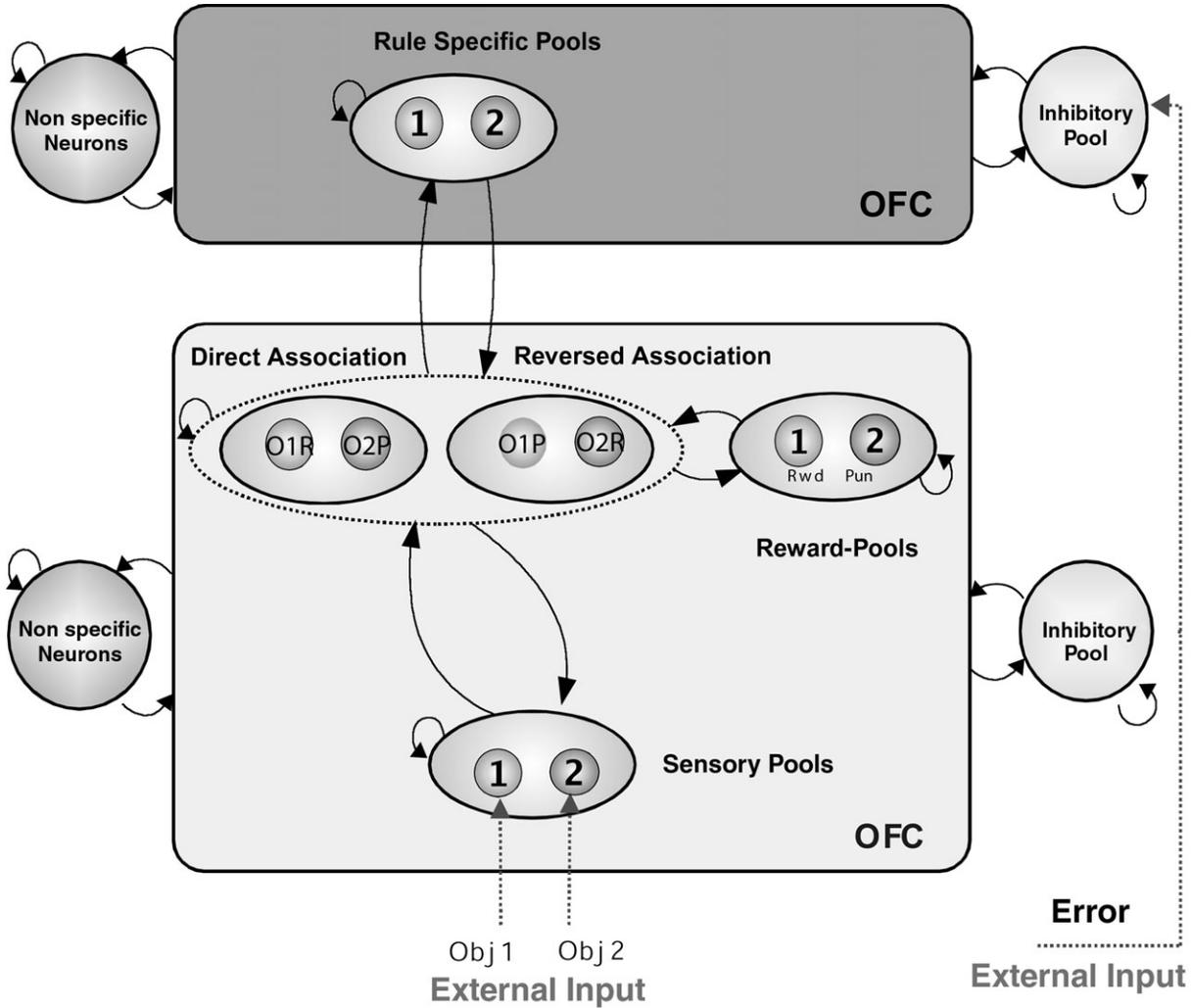
### The Network

The network is composed of two modules: a rule module and a sensory-intermediate neuron-reward module, as shown in Figure 1. Each module contains  $N_E$  (excitatory) pyramidal cells and  $N_I$  inhibitory interneurons. In our simulations, we use  $N_E = 1600$  and  $N_I = 400$  for the sensory-intermediate neuron-reward module, and  $N_E = 1000$  and  $N_I = 200$  for the rule module, consistent with the neurophysiologically observed proportion of 80% pyramidal cells versus 20% interneurons (Abeles, 1991; Rolls and Deco, 2002). In each module, the neurons are fully connected (with synaptic strengths as specified below). Neurons in the orbitofrontal cortical network shown in Figure 1 are clustered into populations or pools. Each pool of selective excitatory cells contains  $fN_E$  neurons. In our simulations  $f = 0.05$  for the associative module (where there are thus 80 neurons in each selective pool); and  $f = 0.1$  for the rule module (where there are thus 100 neurons in each selective pool). There are two different types of pool: excitatory and inhibitory.

In the sensory-intermediate neuron-reward module, there are four subtypes of excitatory pool, namely: object-tuned (visual sensory pools), object-and-expected-reward-tuned (intermediate or associative pools), reward (versus punishment)-tuned pools, and non-selective pools. Object pools are feature-specific, encoding for example the identity of an object (in our case two object-specific pools: syringe and other object, or triangle versus square). The Reward/Punishment pools represent whether the visual stimulus being presented is currently associated with Reward (and for other neurons, with Punishment). Reward neurons are envisaged as naturally leading to an approach response, such as Go, and Punishment neurons as naturally leading to escape or avoidance behaviour, characterized as NoGo behaviour. (In the brain, part of the utility of Reward and Punishment representations is that the animal can learn any action to obtain the reward or avoid the punishment, but for the purposes of the simulation we assume that Rewards lead to Go behaviour such as a lick to obtain glucose taste, and Punishment-association decoding by these neurons to NoGo behaviour such as no lick in order to avoid the taste of aversive saline.) The intermediate or associative pools (in that they are between the sensory and Reward/Punishment association representing pools) are context-specific and perform the mapping between the sensory stimuli to the anticipated reward/punishment pool. (In our case, there are four pools at the intermediate level, two for the direct rewarding context: object 1-rewarding, object 2-punishing, and two for the reversal condition: object 1-punishing, object 2-rewarding.) These intermediate pools respond to combinations of the sensory stimuli and the expected reward, e.g. to object 1 and an expected reward (glucose obtained after licking). The sensory-intermediate neuron-reward module consists of three hierarchically organised levels of attractor network, with stronger synaptic connections in the forward than the backprojection direction. The rule module acts as a biasing input to bias the competition between the object-reward combination neurons at the intermediate level of the sensory-intermediate neuron-reward module. It is an important part of the architecture that at the intermediate level of the sensory-intermediate neuron-reward module one set of neurons fire if an object being presented is currently associated with reward, and a different set if the object being presented is currently associated with punishment. This representation means that these neurons can be used for different functions, such as the elicitation of emotional or autonomic responses, which occur for example to stimuli associated with particular reinforcers (Rolls, 1999).

In the rule module, there are two different types of excitatory pools: context-tuned (rule pools), and the non-selective pools. The rule pools encode the context. In our case, one pool represents: object 1 is rewarding (in that glucose taste reward is obtained if a lick is made) and object 2 is punishing (associated with aversive saline taste so that licking should be avoided); and the other pool represents the reverse associations held currently.

In both modules, the remaining excitatory neurons do not have specific sensory, response or biasing inputs, and are in a non-selective pool. They have some spontaneous firing, and help to introduce some noise into the simulation, which aids in generating the almost Poisson spike firing patterns of neurons in the simulation that are a property of many neurons recorded in the brain (Brunel and Wang, 2001). All the inhibitory neurons are clustered into a common inhibitory pool for each module, so that there is global competition throughout each module.



**Figure 1.** Cortical architecture in the model. There is a rule module (top) and a sensory-intermediate neuron-reward module (below). Neurons within each module are fully connected, and form attractor states. The sensory-intermediate neuron-reward module consists of three hierarchically organised levels of attractor network, with stronger synaptic connections in the forward than the backprojection direction. The intermediate level of the sensory-intermediate neuron-reward module contains neurons that respond to combinations of an object and its association with reward or punishment, e.g. object 1-reward (O1R, in the direct association set of pools), and object 1-punishment (O1P in the reversed association set of pools). The rule module acts as a biasing input to bias the competition between the object-reward combination neurons at the intermediate level of the sensory-intermediate neuron-reward module. 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, respectively mediated by AMPA and NMDA receptors. These two glutamatergic excitatory synapses are on the pyramidal cells and interneurons. The external background is mediated by AMPA synapses on pyramidal cells and interneurons. Each neuron receives  $N^{\text{ext}}$  excitatory AMPA synaptic connections from outside the network. The visual input is also introduced by AMPA synapses on specific pyramidal cells. Inhibitory GABAergic synapses on pyramidal cells and interneurons yield corresponding IPSPs. (See text and appendix for details). OFC, orbitofrontal cortex.

We assume that the synaptic coupling strengths between any two neurons in the network act as if they were established by Hebbian learning, i.e. the coupling will be strong if the pair of neurons have correlated activity, and weak if they are activated in an uncorrelated way. As a consequence of this, neurons within a specific excitatory pool are mutually coupled with a strong weight  $w_s = 2.1$ . Neurons in the inhibitory pool are mutually connected with an intermediate weight  $w = 1$  (forming the inhibitory to inhibitory connections that are useful in achieving non-oscillatory firing). They are 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 is weak and given by

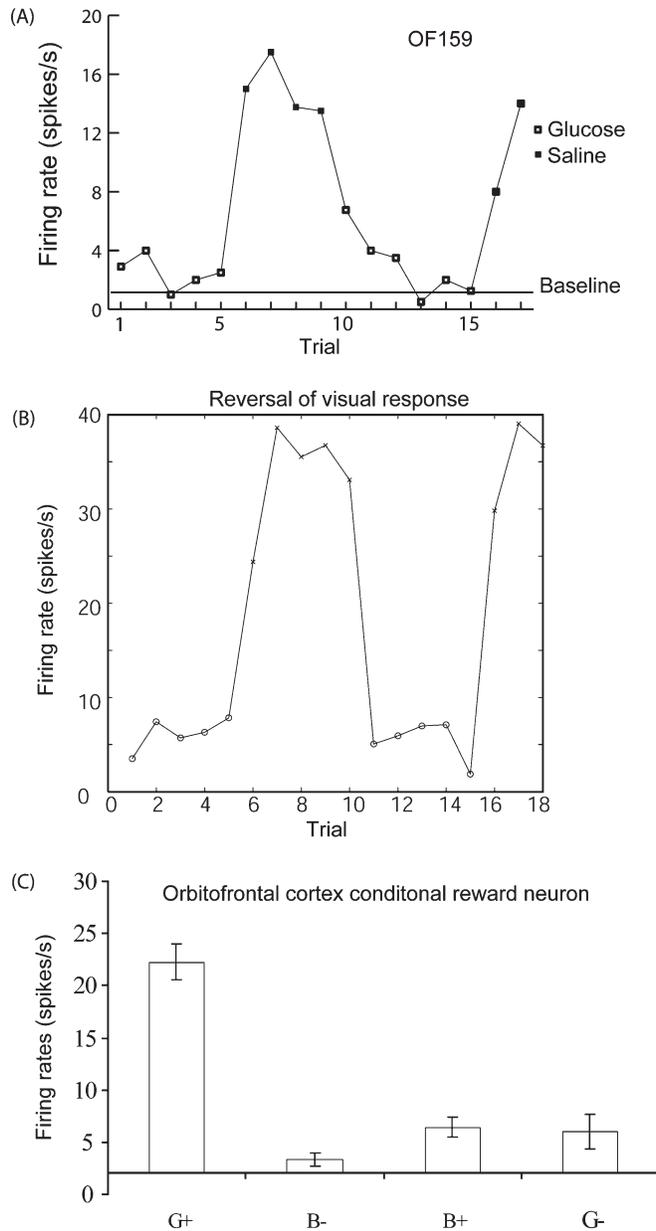
$$w_w = 1 - 2f(w_s - 1)/(1 - 2f)$$

unless otherwise specified (see next paragraph) (where  $f$  is the fraction of the excitatory neurons that are in each pool as described above, e.g.  $f = 0.05$  for the sensory-intermediate neuron-reward module). Neurons

in a specific excitatory pool are connected to neurons in the non-selective pool with a feedforward synaptic weight  $w = 1$  and a feedback synaptic connection of weight  $w_w$ .

The connections between the different pools are set up so that each specific intermediate or associative pool is connected with the corresponding specific sensory-tuned pool, reward rule context-tuned pool from the rule module, and reward/punishment-tuned pool, as if they were based on Hebbian learning of the activity of individual pools while the different tasks are being performed. The strengths of the feedforward and feedback connections between different pools are indicated in Table 1 for the associative module and in Table 2 for the rule module. The connection between the rule pools of the rule module and the associative pools of the associative module were  $w_{\text{intermodule}} = 1.1$  and used only AMPA synapses.

Each neuron (pyramidal cells and interneurons) receives  $N_{\text{ext}} = 800$  excitatory AMPA synaptic connections from outside the network. These connections provide two different types of external interactions: (i) a background noise due to the spontaneous firing activity of neurons



**Figure 2.** Reversal of visual responses in an orbitofrontal cortex single neuron. (A) Experimental results of Thorpe *et al.* (1983). On trials 1–5, no response of the neuron occurred to the sight of a syringe from which the monkey had been given orally glucose reward solution to drink on the previous trial. On trials 6–9, the neuron responded to the sight of the same syringe from which he had been given aversive hypertonic saline drink on the previous trial. Two more reversals (trials 10–15 and 16–17) were performed. (B) Simulated results, with the firing rate being shown of the neurons in the Punishment pool of the orbitofrontal cortex final (Reward/Punishment) neuronal pools in the sensory–intermediate neuron–reward module. The neuron signalled punishment only when the object, O1 (the syringe used to feed the monkey), predicted punishment, and the neuronal response reversed in one trial. (C) A differential reward neuron recorded in the orbitofrontal cortex by Thorpe *et al.* (1983) which responded only to the Green stimulus when it was associated with reward (G+), and not to the Blue stimulus when it was associated with Reward (B+), or to either stimuli when they were associated with Punishment (G– and B–).

outside the network; and (ii) a sensory-related input (object-specific). The external inputs are given by a Poisson train of spikes. In order to model the background spontaneous activity of neurons in the network (Brunel and Wang, 2001), we assume that Poisson spikes arrive at each external synapse with a rate of 3Hz, consistent with the spontaneous activity observed in the cerebral cortex (Wilson *et al.*, 1994; Rolls and

**Table 1**

Neuronal connectivity between different neuronal pools in the stimulus–intermediate neuron–reward/punishment module

Pools	O1	O2	O1-R	O2-P	O1-P	O2-R	R	P	Unsp.	Inh.
O1	$w_s$	$w_w$	$w_{ff}$	$w_w$	$w_{ff}$	$w_w$	$w_w$	$w_w$	1	1
O2	$w_w$	$w_s$	$w_w$	$w_{ff}$	$w_w$	$w_{ff}$	$w_w$	$w_w$	1	1
O1-R	$w_{fb}$	$w_w$	$w_s$	$w_w$	$w_w$	$w_w$	$w_s$	$w_w$	1	1
O2-P	$w_w$	$w_{fb}$	$w_w$	$w_s$	$w_w$	$w_w$	$w_w$	$w_s$	1	1
O1-P	$w_{fb}$	$w_w$	$w_w$	$w_w$	$w_s$	$w_w$	$w_w$	$w_s$	1	1
O2-R	$w_w$	$w_{fb}$	$w_w$	$w_w$	$w_w$	$w_s$	$w_s$	$w_w$	1	1
R	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	1	1
P	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	$w_s$	1	1
Unsp.	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	$w_w$	1	1
Inh.	1	1	1	1	1	1	1	1	1	1

O1: object 1 sensory pool (syringe, or triangle); O2: object 2 sensory pool (other object, or square); O1-R: intermediate pool for object 1 associated with reward; O2-P: intermediate pool for object 2 associated with punishment; O1-P: intermediate pool for object 1 associated with punishment; O2-R: intermediate pool for object 2 associated with reward; R: Reward (Go) pool; P: Punishment (NoGo) pool; Unsp: 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).

**Table 2**

Neuronal connectivity between different neuronal pools in the rule module

Pools	Rule direct (O1-R)	Rule reversal (O1-P)	Unsp.	Inh.
Rule direct (O1-R)	$w_s$	$w_w$	1	1
Rule reversal (O1-P)	$w_w$	$w_s$	1	1
Unsp.	$w_w$	$w_w$	1	1
Inh.	1	1	1	1

Rule direct (O1-R): object 1 is currently associated with reward; Rule reversal (O1-P): object 1 is currently associated with punishment; Unsp: non-specific neuronal pool; Inh: inhibitory neuron pool.  $w_w$ : weak synaptic strength (= 0.878).  $w_s$ : strong synaptic strength (= 2.1).

Treves, 1998). In other words, the effective external spontaneous background input rate of spikes to each cell is  $v_{ext} = N_{ext} \times 3 \text{ Hz} = 2.4 \text{ kHz}$ . The sensory input is encoded by increasing the external input Poisson rate  $v_{ext}$  to  $v_{ext} + \lambda_{input}$  to the neurons in the appropriate specific sensory pools (Brunel and Wang, 2001).  $\lambda_{input}$  is 200 Hz, which corresponds across  $n$  synapses to an average increase of  $200/n$  Hz at each synapse, or on average a change of rate at the 800 excitatory synapses from 3 to 3.25 Hz on each synapse.

The cortical architecture introduced above for the sensory–intermediate neuron–reward module presents the characteristic that its different global attractors corresponding to the different sensory cue–reward context–response situations, are each composed of a set of single pool attractors, where the single pools that are active represent a particular combination of sensory, associative and reward/punishment pools. The cue stimulus and the biasing top–down rule or the context information from the rule module, drive the system into the corresponding attractor. In fact, the system is dynamically driven according to the biased competition hypothesis (Moran and Desimone, 1985; Spitzer *et al.*, 1988; Chelazzi *et al.*, 1993; Miller *et al.*, 1993; Motter, 1993; Chelazzi, 1998; Reynolds and 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 within the sensory–intermediate neuron–reward module. The top–down interactions bias this competition in favour of specific pools, resulting in the build up of the global attractor that corresponds to the context–specific stimulus–reward mapping required.

All neuronal and synaptic equations were integrated using the second–order Runge–Kutta method, with an integration step of  $dt = 0.1 \text{ ms}$ . 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.

### Reward Reversal: The Operation of the Rule Module Neurons

The cue-response mapping required under a specific context is achieved via the biasing effect of the spiking information coming from the rule pools. For a specific context a specific rule pool will be activated, and the other rule pools are inactive. When a reward reversal occurs, the rule pools switch their activity, i.e. the previously activated context-specific rule pool is inactivated, and a new rule pool (that was inactive) is now activated, to encode the new context. Switching the rule pools switches the bias being applied to the intermediate pools, which effectively represent when a stimulus is shown whether it is (in the context of the current rule) associated with reward or with aversive saline. From the intermediate pool, the mapping is then straightforward to the reward/punishment pools (which have implied connections to produce a Go response (of licking) if a currently reward-associated visual stimulus is being shown, and a NoGo response of not licking if a stimulus currently associated with aversive saline is shown. To achieve the reversal in the rule module, we assume that the attractor state in the rule module is reset via a non-specific global inhibitory signal, which is received after each punishment or absence of an expected reward. Neurons which respond in just this way, i.e. when an expected reward is not obtained, and a stimulus-reinforcement reversal must occur, were found in the orbitofrontal cortex by Thorpe *et al.* (1983). These neurons can be described as error neurons. In our implementation, we implemented the effects of this error signal by increasing for 50 ms the external AMPA-input to the inhibitory pool of the rule module (see Fig. 1). (The increase was from  $v_{\text{ext}}$  to  $v_{\text{ext}} + \lambda_{\text{punish}}$  with  $\lambda_{\text{punish}} = 900$  Hz, which corresponds to an increase of 1.125 Hz to each of the 800 external synapses impinging on the neurons of the inhibitory pool. This compares to the mean value of the spontaneous external input of 3 Hz per synapse.) This increased the global inhibition of the rule module, and suppressed the activity of all the excitatory neuronal pools in the rule module. Effectively, the firing of the error neurons activated the inhibitory neurons in the rule module. (The effect could be implemented in the brain by the active error neurons activating inhibitory interneurons which influence, among other neurons, the rule module excitatory neurons. The system would work just as well if this inhibitory feedback was applied to both the modules shown in Figure 1, not just the rule module, and this might be more parsimonious in terms of the connectivity required in the brain.) We incorporate into the excitatory synaptic connections between the neurons in the rule module the property that they show some spike-frequency-adaptation process (with details provided below). This provides a mechanism that implements a temporal memory of the previously activated pool. When the attractor state of the rule module is shut down by the inhibitory input, then the attractor state that subsequently emerges when firing starts again will be different from the state that has just been present, because of the synaptic adaptation in the synapses that supported the previous attractor state. In order to assure that one of the rule pools is active, and to promote high competition between the possible reward contexts, we excite externally all rule pools with the same non-specific input by increasing the external input Poisson firing rate impinging on the excitatory pools of the rule module (from  $v_{\text{ext}}$  to  $v_{\text{ext}} + B$  with  $B = 200$  Hz). In a non-essential part of the model, after each response (rewarding or not) we assumed that the dopamine level increases, so that the whole dynamics is reinforced by increasing the NMDA and GABA conductivity [as in the models of Law-Tho *et al.* (1994), Zheng *et al.* (1999) and Brunel and Wang (2001)] by a factor  $Dp = 2$ , which decays according an exponential function with a time constant of 5 s. This non-essential process helps to stabilize the attractors in the system. It was not used in the simulation shown in Figure 5.

We describe now the specific implementation of the spike-frequency-adaptation mechanism that we used in the rule module. One implementation (used for the simulations shown in Figs 3, 4 and 6) was a sodium inactivation based spike-frequency-adaptation mechanism. A full statistical analysis of a model of sodium inactivation in the framework of integrate-and-fire models was introduced by Giugliano *et al.* (2002) as a realistic candidate for long lasting non-monotonic effects in current-to-rate response functions observed *in vitro* (Rauch *et al.*, 2003) and associated with spike-frequency-mechanisms.

The model was called an Integrate-and-may-fire (IMF) model, and takes into account the inactivation of sodium channels after spike generation. The integrate-and-fire model is modified, by changing the condition that when the membrane potential reaches the threshold  $\theta$ , the emission of a spike at that time is an event occurring with an activity-dependent probability  $q$ . After the spike emission, the membrane potential is clamped to the value  $V_{\text{reset}} = -55$  mV, for an absolute refractory time, after which the current integration starts again. However, each time the excitability threshold  $\theta$  is crossed and no spike has been generated (i.e. an event with probability  $1 - q$ ), the membrane potential is reset to  $H_2$  ( $0 < V_{\text{reset}} < H_2 < \theta$ ) and no refractoriness occurs. Additionally,  $q$  is a decreasing function of a slow voltage-dependent variable  $w$  ( $0 < w < 1$ ), reminiscent of the sigmoidal voltage dependence of the fast inactivation state variables that characterize conductance-based model neurons:

$$q = [1 + e^{(\omega - \omega_0)/\sigma_\omega}]^{-1} \quad (1)$$

where  $w$  evolves by,

$$\tau_\omega \frac{d\omega(t)}{dt} = \frac{V(t)}{\theta} - \omega \quad (2)$$

which corresponds to a first approximation to the average transmembrane electric field experienced by individual ion channels and affecting their population-level activation and inactivation. In our simulations we utilized  $\tau_\omega = 10\,000$  ms,  $\sigma_\omega = 0.01$ ,  $\omega_0 = 0.87$  and  $H_2 = -52$  mV.

We also tried other spike-frequency-adapting mechanisms, including  $\text{Ca}^{2+}$ -activated  $\text{K}^+$  hyper-polarizing currents (Liu and Wang, 2001) and short-term synaptic depression (Abbott and Nelson, 2000). Both of them were successfully used for producing the desired reward reversal based context-switching. The synaptic depression mechanism was used for Figure 5, and the details were after Dayan and Abbott (2002), page 185. In particular, the probability of release  $P_{\text{rel}}$  was decreased after each presynaptic spike by a factor  $P_{\text{rel}} = P_{\text{rel}} \cdot f_D$  with  $f_D = 0.994$ . Between presynaptic action potentials the release probability  $P_{\text{rel}}$  is updated by

$$\tau_P \frac{dP_{\text{rel}}}{dt} = P_0 - P_{\text{rel}} \quad (3)$$

with  $P_0 = 1$ .

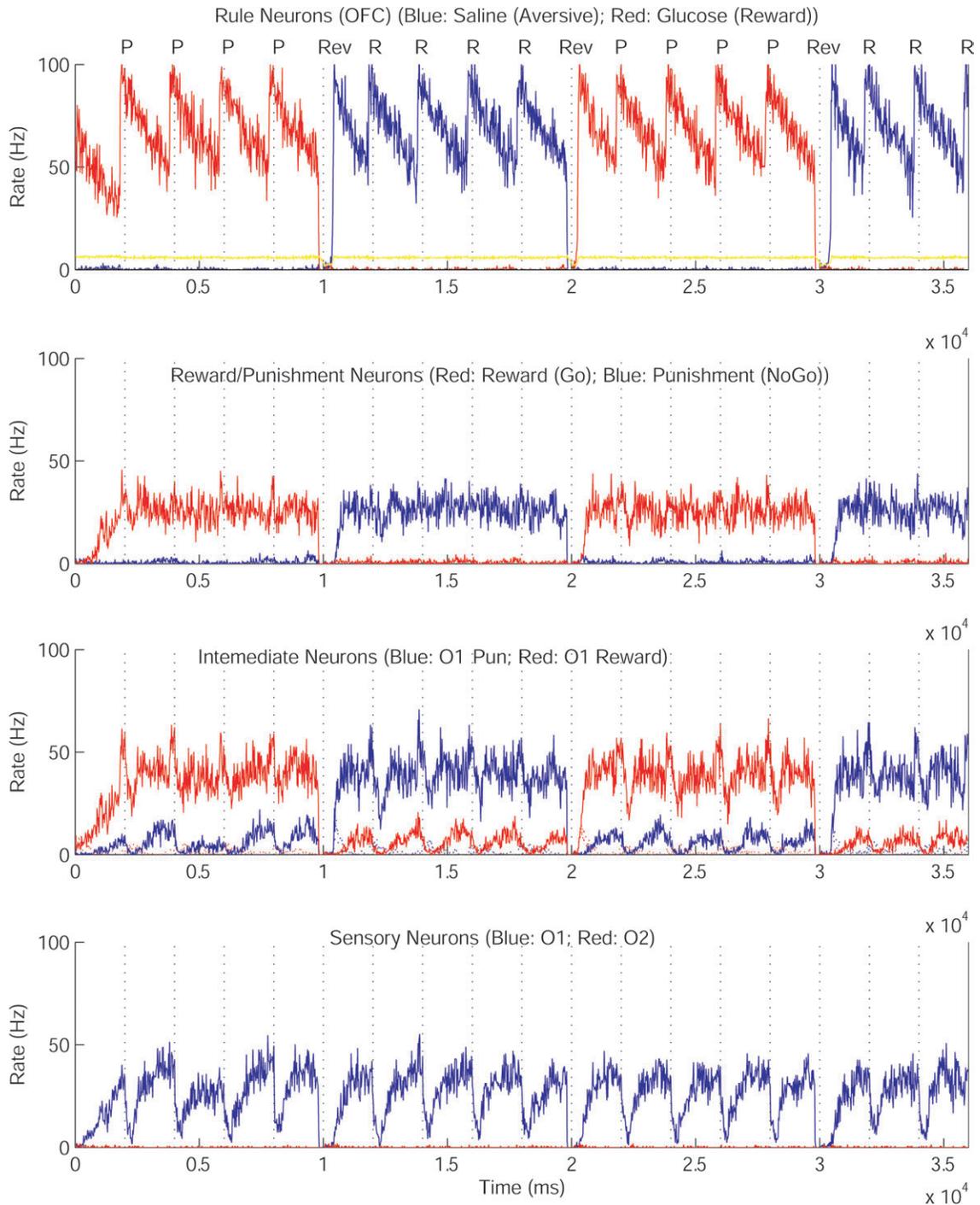
### The Neuron

We use leaky integrate-and-fire neurons for modelling the excitatory pyramidal cells and the inhibitory interneurons. The synaptic inputs to an integrate-and-fire neuron are basically described by a capacitor  $C_m$  connected in parallel with a resistor  $R_m$  through which currents are injected into the neuron. These current injections produce excitatory or inhibitory post-synaptic potentials, EPSPs or IPSPs, respectively.

These potentials are integrated by the cell, and if a threshold  $\theta$  is reached, a  $\delta$ -pulse (spike) is fired and transmitted to other neurons, and the potential of the neuron is reset. The incoming presynaptic  $\delta$ -pulse current from another neuron is low-pass filtered by the synaptic and membrane time constants, and an EPSP or IPSP in the one-compartment neuronal model. We use biologically realistic parameters (McCormick *et al.*, 1985). We take for both excitatory and inhibitory neurons a resting potential  $V_l = -70$  mV, a firing threshold  $\theta = -50$  mV, and a reset potential  $V_{\text{reset}} = -55$  mV. The membrane capacitance  $C_m$  is 0.5 nF for the pyramidal neurons and 0.2 nF for the inhibitory interneurons. The membrane leak conductance  $g_m$  is 25 nS for pyramidal cells and 20 nS for interneurons. The refractory period  $\tau_{\text{ref}}$  is 2 ms for pyramidal cells, and 1 ms for interneurons. Hence, the membrane time constant  $\tau_m = C_m/g_m$  is 20 ms for pyramidal cells and 10 ms for interneurons.

### The Synapses

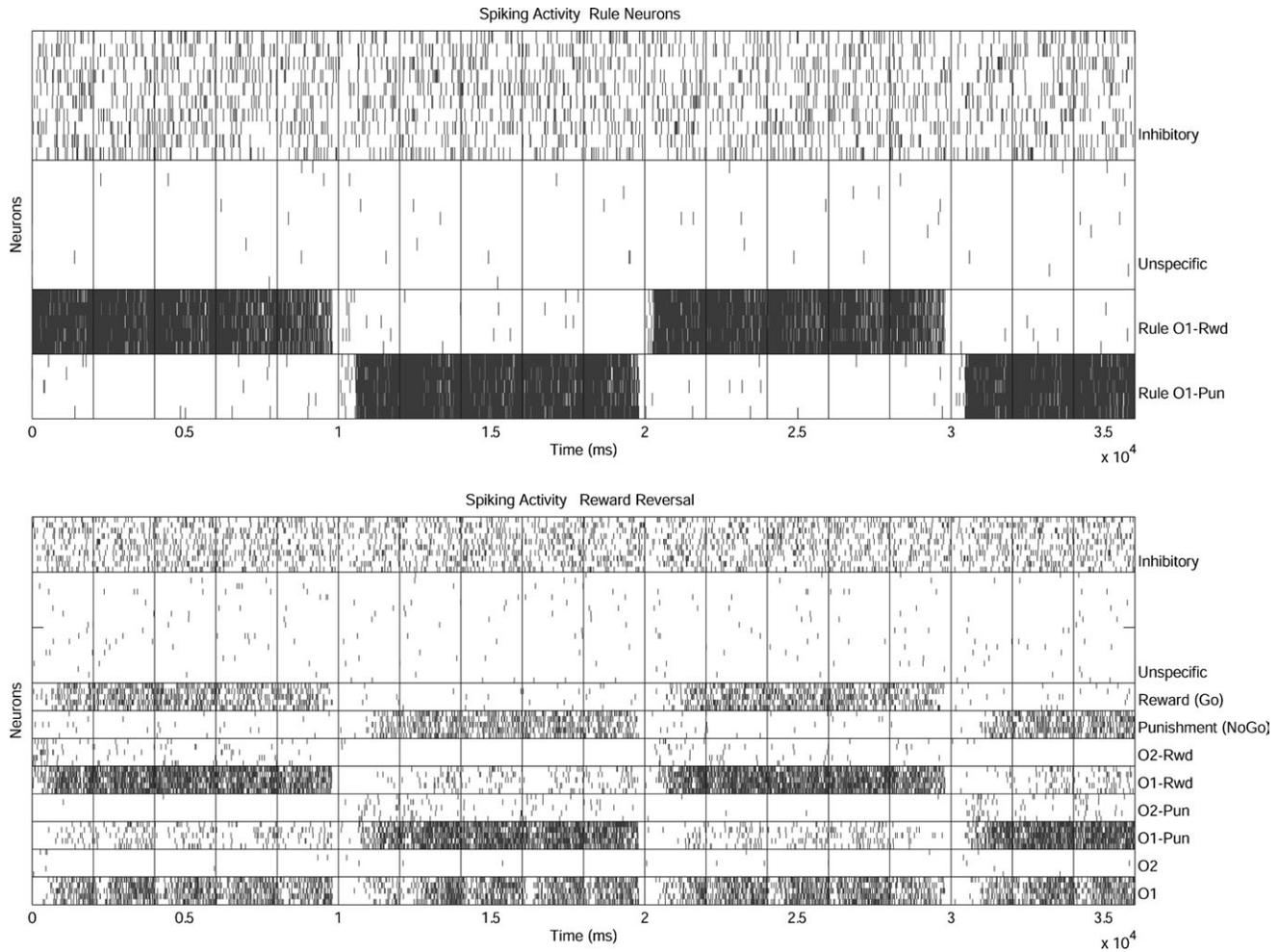
The synaptic current flows into the cells are mediated by three different families of receptors. The recurrent excitatory postsynaptic EPSPs are mediated by AMPA and NMDA (*N*-methyl-D-aspartate) receptors. These two glutamatergic excitatory synapses are on the pyramidal cells and on the interneurons. 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



**Figure 3.** Temporal evolution of the averaged population activity for all neural pools (sensory, intermediate (stimulus–reward/punishment), and Reward/Punishment) in the sensory–intermediate neuron–reward module and the rule module during the execution of the reversal of the Go/NoGo visual discrimination task shown in Figure 2B. Each vertical line indicates a new trial. The yellow curve in the top graph shows the average activity of the inhibitory interneurons in the rule module. O1 was the syringe used to present the stimuli in Figure 2A. R: Reward trial; P: Punishment Trial; Rev: Reversal trial, i.e. the first trial after the reward contingency was reversed when Reward was expected but Punishment was obtained.

synapses on pyramidal cells and interneurons yield the corresponding IPSPs. The mathematical descriptions of each synaptic channel are provided in the Appendix, and the corresponding parameters are also specified there. We consider that the NMDA currents have a voltage dependence that is controlled by the extracellular magnesium concentration (Jahr and Stevens, 1990),  $C_{Mg^{2+}} = 1$  mM. We neglect the rise time of both AMPA and GABA synaptic currents because they are typically very short ( $<1$  ms). The rise time for NMDA synapses is  $\tau_{NMDA,rise} = 2$  ms

(Hestrin *et al.*, 1990; Spruston *et al.*, 1995). All synapses have a latency (time delay) of 0.5 ms. The time constant 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 and Prince, 1996; Xiang *et al.*, 1998). The synaptic conductivities for each receptor type were taken from Brunel and Wang (2001), and were adjusted using a mean field analysis to be approximately 1 nS in magnitude, and were consistent with



**Figure 4.** Rastergrams of randomly selected neurons for each pool in the OFC network from the simulation shown in Figures 2B and 3 during the execution and reversal of the Go/NoGo visual discrimination task after Thorpe *et al.* (1983). Each vertical line indicates a new trial. O1, object 1 (the syringe) was used throughout this experiment. O1-Pun is the intermediate pool responding to O1 when it is associated with reward. The final pools are denoted Reward (which implies a Go response) and Punishment (which signifies a NoGo response).

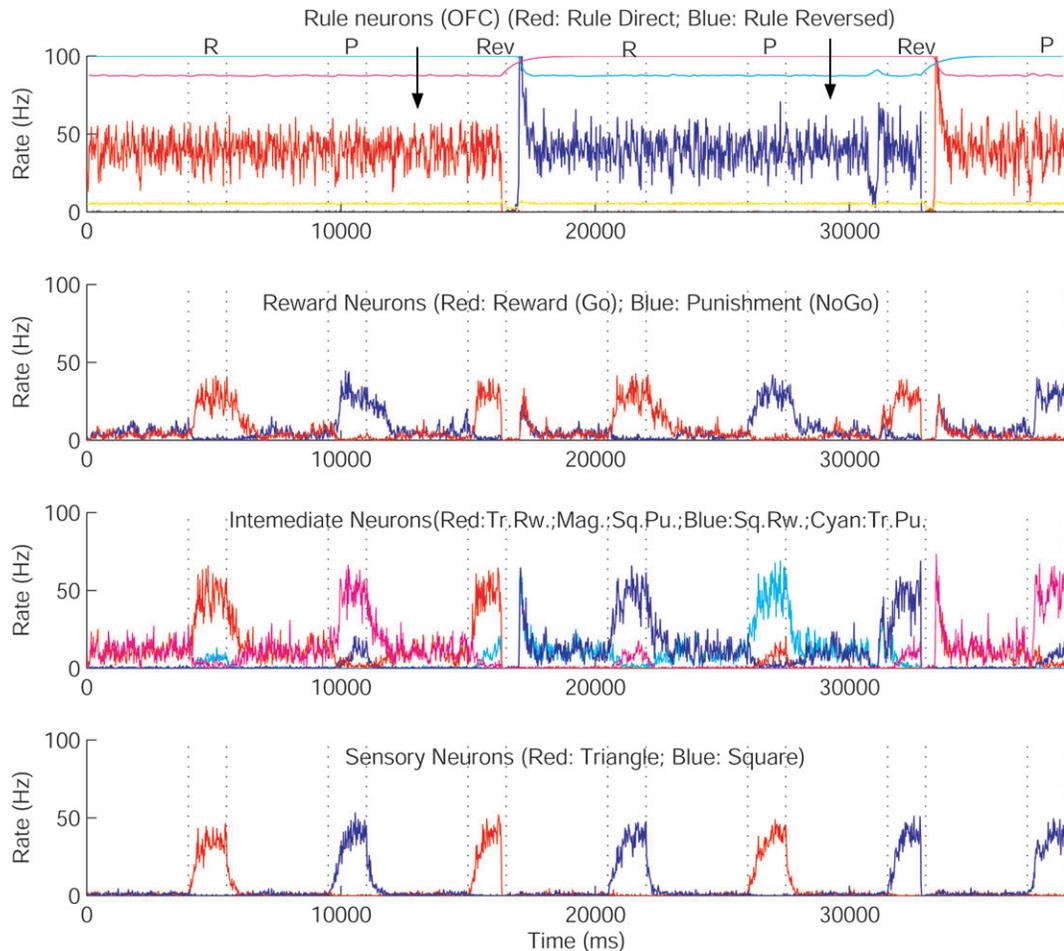
experimentally observed values (Destexhe *et al.*, 1998) (see Appendix). As was noted by Lisman *et al.* (1998), Wang (1999) and Brunel and Wang (2001), the recurrent excitation was assumed to be largely mediated by the NMDA receptors, in order to provide more robust persistent activity during the short-term memory related delay period; and the amplitude of recurrent excitation was smaller than that of feedback inhibition, and therefore the net recurrent input (i.e. the sum of these two terms) to a neuron was hyperpolarizing during spontaneous activity (i.e. without external inputs) (Amit and Brunel, 1997; Brunel and Wang, 2001). Figure 1 shows schematically the synaptic structure assumed in the orbitofrontal cortical network.

## Results

### *Single Neuron Visual Discrimination Reversal in the Orbitofrontal Cortex*

We simulated, with the architecture described in the preceding section and shown in Figure 1, the experimental set-up utilized by Thorpe *et al.* (1983) and Rolls *et al.* (1996) in order to analyse theoretically the neuronal activity in the primate orbitofrontal cortex underlying the execution of a visual discrimination reward reversal task. The model was able to reverse the behavioural responses quickly, so that if it obtained a non-

reward signal (saline for licking) to a stimulus that had previously been associated with reward, it subsequently only produced a lick to the previously punished stimulus which now indicated that reward was available. Figure 2B shows the reversal of visual responses in the orbitofrontal pool corresponding to the 'object 1-aversive' association. This figure can be compared with the neurophysiological results shown in Figure 2A. Because the significance of the visual stimuli had been altered, the neurons in the 'object 1-aversive' associative pool responded only in the cases where visual object 1 was shown, and the model did not perform a lick response because a saline solution was being expected. In fact, in this case, the 'object 1'sensory pool and the 'Punishment (NoGo)' pool of the Reward/Punishment pool were also coactivated as parts of the same global attractor. On the other hand, during the trials where object 1 was associated with a glucose reward, then the 'object 1-reward' associative pool responded with high activation, together with the sensory 'object 1' pool and the 'Reward (Go)' Reward/Punishment pool. These three pools formed the global attractor under this rewarding condition, and therefore the former attractors corresponding to the reversal condition were suppressed through inhibition within the global attractor.



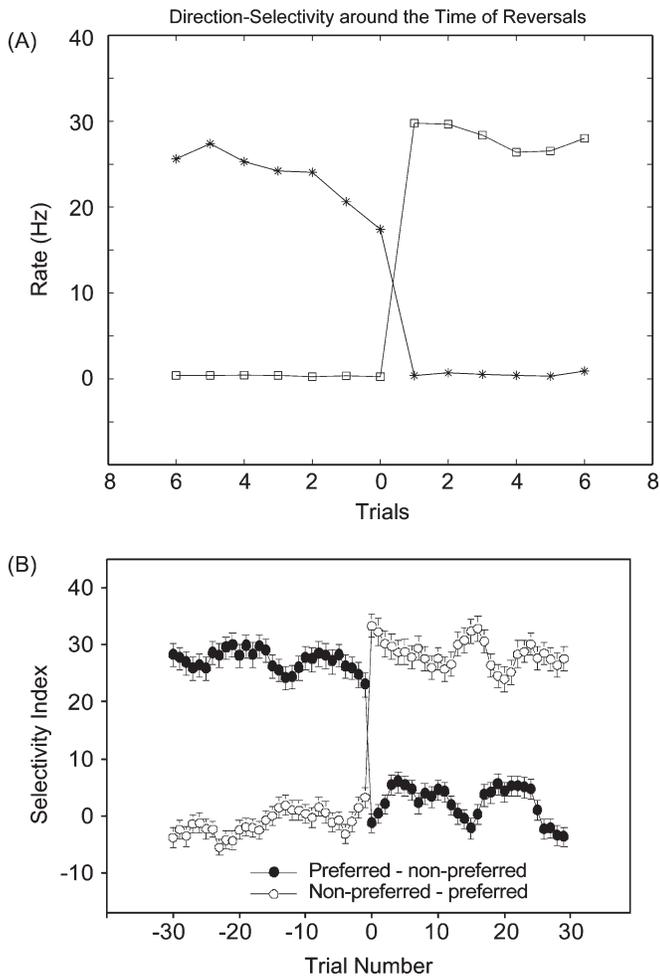
**Figure 5.** Temporal evolution of the averaged population activity for all neural pools [sensory, intermediate (stimulus–reward) and Reward/Punishment] in the stimulus–intermediate neuron–reward module and the rule module during the execution and the reversal of the Go/NoGo visual discrimination task with a pseudorandom trial sequence after Thorpe *et al.* (1983) and Rolls *et al.* (1996). Bottom row: the sensory neuronal populations, one of which responds to object 1, a triangle (red), and the other to object 2, a square (blue). For example, the intermediate neurons respond to object 1 (Tr) when it is associated with reward (Rw) (e.g. on trial 1). The top row shows the firing rate activity in the rule module, with the thin line at the top showing the mean probability of release  $P_{rel}$  of transmitter from the synapses of each population of neurons. The arrows show when the reversal contingencies reversed. R: Reward trial; P: Punishment Trial; Rev: Reversal trial, i.e. the first trial after the reward contingency was reversed when Reward was expected but Punishment was obtained. The inter-trial interval was 4 s. The yellow line shows the average activity of the inhibitory neurons. (See text for further details.)

The rule input thus biased the competition in the correct direction.

Figures 3 and 4 show the whole spatio-temporal picture by plotting the firing rates and the corresponding rastergrams. The rastergrams (Fig. 4) show randomly selected neurons for each pool in the stimulus–intermediate neuron–reward module (5 for each sensory, intermediate and Reward (Go)/Punishment (NoGo) pool, 20 for the non-selective excitatory pool and 10 for the inhibitory pool), and in the rule module (5 for each rule pool, 10 for the non-selective excitatory pool and 10 for the inhibitory pool). The spatio-temporal spiking activity shows both attractors described above, i.e. those that were present when the task was run non-reversed and reversed. The cue stimulus and the biasing rule top-down synaptic connections applied to the associative neurons, drive the system into the corresponding global attractor utilizing biased competition mechanisms. The activity of the rule pools was switched correctly by the external non-reward signal. It is of interest that although the inputs produced by the error signal were applied only to the inhibitory neurons in the rule module (via

their AMPA receptors), most of these inhibitory neurons in fact decreased their firing rates on the coarse time scale. This is because there are inhibitory to inhibitory neuron connections, and because the inhibition produced in the excitatory neurons itself caused less drive to the inhibitory neurons. Although the total firing in the modules thus was decreased, the synaptic activity produced at least by the inhibitory external input was strong. This is of interest, for total synaptic activity, rather than average firing rate, may be reflected in fMRI signals (Deco *et al.*, 2004). In any case, the switching of the attractor in the rule module acts as a reversal of the biasing context input to the stimulus–reward–response module, and therefore a switching of the attractors in the associative module. This is a biased competition operation. As shown in Figures 3 and 4, this switching is very fast, within one trial of when a non-reward signal was obtained.

Figure 5 shows the results of a simulation of the more usual Go/NoGo task design with a pseudorandom sequence of trials. On each trial, either Object 1 (a triangle) or Object 2 (a square) was shown. In Figure 5, on trial 1 the rule network was



**Figure 6.** Reversal of a conditional object-response task with a delay. Average firing activity of the two motor direction-selective pools around the time of reversal which occurred at trial zero. (A) Simulation results. The stars show the responses to the object that, having indicated a saccade in the neuron's preferred direction, begins at trial zero to indicate a saccade in the neuron's non-preferred direction. The squares show the opposite, namely the activity produced by the object that before reversal cued a saccade in the non-preferred direction, and after reversal required an eye movement in the neuron's preferred direction. (B) Neurophysiological results obtained in the experiment of Asaad *et al.* (1998) during reversal.

operating in the direct mapping state, the sensory pool responded to the triangle, the intermediate pool that was selected based on this sensory input and the direct rule bias was the triangle-reward pool, this pool led to activation of the Reward (or Go) pool, and a reward (R) was obtained. On trial 2 the sensory pool for the square responded, and this with the direct rule bias led to the intermediate square-Non-reward pool to be selected, and this in turn led to Punishment neurons being active, leading to a NoGo response (i.e. no action). On trial 3 the sensory triangle pool was activated, leading because of the direct rule to activation of the intermediate triangle-reward pool, and Reward was decoded (leading to a Go response being made). However, because this was a reversal trial, punishment was obtained, leading to activation of the error input, which increased the inhibition in the rule module, and quenching of the rule module attractor. When the rule module attractor started up again, it started with the reverse rule neurons active, as they won the competition with the direct rule neurons, whose excitatory synapses had adapted during the previous few

trials. On trial 4 the sensory-square input neurons were activated, and the intermediate neurons representing square-reward were activated (due to the biasing influence of the reversed rule input to these intermediate neurons, the Reward neurons in the third layer were activated (leading to a Go response), and reward was obtained. On trial 5 the sensory-triangle neurons activated the triangle-Non-reward intermediate neurons under the biasing influence of the reversed rule input, and Punishment was decoded by the third layer (resulting in a NoGo response). On trial 6, the sensory-square neurons were activated leading to activation of the intermediate square-reward neurons, and Reward (and a Go response) was produced. However, this was another reversal trial, non-reward or punishment activated the error inputs, and the rule neurons in the rule module were quenched, and started up again with the direct rule neurons active in the rule module, due to the synaptic depression of the synapses between the reversed rule neurons.

### Single Neuron Recordings in the Dorsolateral Prefrontal Cortex during the Reversal of a Conditional Object-Response Task

In this subsection, we present a theoretical analysis of neuronal activity in the primate dorsolateral prefrontal cortex (PFC) underlying the execution of the reversal of a conditional 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 that 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 visuomotor task* with a delay. The task required the monkeys to one response (e.g. a left oculomotor saccade) after a delay following the presentation of one object (e.g. A) at the fovea, and a different response (e.g. a right oculomotor saccade) after a delay following the presentation of another object (e.g. B) at the fovea. 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.

We adapted our network structure to simulate this task. The architecture is similar to that shown in Figure 1, but what is coded in the intermediate layer of the lower module, which is now a stimulus—stimulus-response combination—response module is different. The sensory pools now encode information about objects. The object or feature-based sensory pools are feature-specific, encoding for example the identity of an object (e.g. form, colour). The premotor pools encode the motor response (in our case the leftward or rightward oculomotor saccade), and replace the Reward/Punishment pools in the third layer in Figure 1. The intermediate pools (in that they are between the sensory and premotor pools) are task-specific and perform the mapping

between the sensory stimuli and the required motor response. The intermediate pools respond to combinations of the sensory stimuli and the response required, e.g. to object 1 requiring a left oculomotor saccade. The intermediate (or associative) pools receive a top-down biasing input that comes from the rule module, where now the two rule pools correspond to the direct and reversed conditional response association.

Figure 6A shows the simulation results corresponding to Figure 5a in the paper of Asaad *et al.* (1998) (of which we show an example in Fig. 6B). Figure 6A plots the average firing activity of the two motor direction-selective response pools around the time of reversal at trial zero. The stars show the responses to the object that, having indicated a saccade in the neuron's preferred direction before the reversal, starts after the reversal at trial zero to indicate a saccade in the neuron's non-preferred direction. The squares show the opposite, namely the activity produced by the object that before reversal at trial 0 cued a saccade in the neuron's non-preferred direction, and after reversal required an eye movement in the preferred direction. The simulation results show the same rapid reversal context switching observed in the experiments, which corresponds dynamically with a change in the whole attractor structure due to the non-rewarding inhibitory signal provided at the time of reversal. This error signal resets the whole rule module to zero firing, and, because of the intrinsic temporal memory associated with the spike-frequency adaptation mechanism or short term depression, when the rule module network starts up again, the opposite rule pool is active. This reverses the bias on the associative module, and the stimulus-response associations there are reversed.

## Discussion

This model thus shows how rapid stimulus-reinforcement association reversal learning could occur. It is an important part of the architecture that at the intermediate level of the sensory-intermediate neuron-reward module one set of neurons fire if an object being presented is currently associated with reward, and a different set fire if the same object being presented is currently associated with punishment. This is in line with what is found in the primate orbitofrontal cortex (Thorpe *et al.*, 1983), as illustrated in Figure 2C. This representation means that these intermediate neurons can be used for different functions, such as the elicitation of emotional or autonomic responses, which occur for example to stimuli associated with particular reinforcers (Rolls, 1999). (For example, this allows different emotional responses to occur to different cognitive stimuli, even if the same primary reinforcer is associated with both stimuli.) This makes the architecture quite different to that of the dorsolateral prefrontal network architecture modelled by Deco and Rolls (2003), which performs a stimulus to motor response mapping using neurons at the intermediate level that respond to combinations of stimuli and motor responses, which is the type of neuron recorded in that region (Asaad *et al.*, 1998, 2000). The dorsolateral prefrontal network thus operates by switching between two habits (stimulus-response associations). In contrast, the orbitofrontal cortex architecture described here in Figures 1-5 maps stimuli to expected rewards. The expected rewards then provide the goal for any appropriate motor response. This introduces flexibility into the response selection, and this flexible choice provides a fundamental evolutionary advantage of emotion in brain design (Rolls, 1999, 2004b).

After the conditional reward layer of intermediate neurons, the next layer represents the reward (or punishment) value of the stimulus independently of whether the current rule is direct or reversed. This third layer is thus the layer at which emotional states and responses can be elicited, as it is this layer that represents the current reward (or punishment) value of the stimulus independently of reversal. These third layer neurons will then in the real brain tend to elicit approach, which is an unlearned behaviour to a reward, and escape or avoidance, which is the natural behaviour to a punisher. However, the utility of a reward representation, which can be rapidly reversed in one trial as described here, is that any instrumental action can then be performed to obtain the reward, or avoid or escape the punisher. This is how the flexibility of behaviour arises that was referred to above as a fundamental evolutionary advantage of designing the brain with reward and punishment systems (Rolls, 1999, 2004b).

It is worth noting that the neurons described by Asaad *et al.* (1998, 2000) in the dorsolateral prefrontal cortex that respond to combinations of particular visual stimuli and responses are analogous to these orbitofrontal cortex neurons that respond to combinations of particular visual stimuli and rewards (Thorpe *et al.*, 1983).

The rapid switch of the rule in the model is produced by a single error without any synaptic modification. To set up the networks that hold the different rules, some learning is needed, and the learning of the appropriate rule pools with the correct connections to the intermediate pools is what we suggest is occurring while the reversal learning set is being acquired. Acquisition of the learning set can take a number of reversals, during which the number of trials for the reversal to occur gradually decreases to one trial. The actual implementation described here, of using an attractor network to hold the current rule, which then biases intermediate conditional neurons to achieve the correct mapping to reward (or, in the case of the dorsolateral prefrontal cortex, to responses) does appear to be quite fundamental, because an active short term (rule) memory implemented by persistent firing in an attractor does provide the necessary source of bias input for the biased competition stimulus-intermediate neuron-reward mapping network. A synaptic modification process occurring in a pattern association stimulus-reward network when an error indicated that the rule had been reversed would require large and one-trial synaptic depression to prevent the former reward-associated stimulus from still producing reward. Instead, in the approach taken here, we have shown that by reversing the state of a dynamical system when the error comes, and using this reversed state (of the rule module) to provide an active bias input to a mapping network, then very rapid, one-trial, reversal can be obtained. We note that the rule has to be available for many trials (until the next reversal). The intermediate neurons in the mapping module that respond to combinations of stimuli and reward are only active when the stimulus is applied. Thus they could not hold the current rule in any short term memory. This is why a separate rule module is required. [For comparison, Deco and Rolls (2003) model task switching in the prefrontal cortex by biased competition acting on intermediate-level stimulus-response combination-responding neurons, but do not model the reversal of the rule neurons required to reverse the biased competition signal. The reversal of the rule module is an issue treated here.]

The model introduced here makes predictions which can be tested neurophysiologically. One prediction is that there will be

rule neurons in the orbitofrontal cortex, which have high firing when one rule applies and low firing when a different rule applies. These neurons should reverse their state when the monkey reverses in the visual discrimination reversal task, and should maintain their state of firing for as long as the rule applies. Secondly, the rule neurons should, immediately after reversal to the high-firing state, have somewhat higher firing than later, reflecting some adaptation in their state. It will be of interest to measure the time course of this alteration of firing rate, which is predicted to take 10–100 s to develop, and should last for at least 60 s. Thirdly, it should be possible if a long delay occurs during testing, or if the monkey is allowed to sleep for a few minutes, which rule the monkey will use to interpret the stimuli (in terms of whether reversal applies or not) from the state of firing of the rule neurons. Fourthly, there should be different rule neurons for different stimulus pairs if several stimulus pairs are used simultaneously, and the pairs are reversed independently. Further, the rule neurons may be not only visual stimulus-specific, but also task- or context-specific, in that reversal of the association in one task need not imply the reversal of the interpretation of the stimuli in all tasks.

Next, we compare this model with other approaches to rapid reversal of behaviour, and highlight the original contributions of the theory and model described here. First, models of cognitive task switching as implemented by the prefrontal cortex have involved for example a trial and error search implemented by dopamine inputs reflecting a temporal difference learning control signal to the prefrontal cortex to transiently increase weights from posterior cortex after obtaining an unexpected reward, and modulating synaptic weights in the prefrontal cortex if an expected reward is not obtained (O'Reilly *et al.*, 2002). In contrast, the mechanism described here involves no modulation of synaptic weights by the non-reward signal known to be present in the orbitofrontal cortex (Thorpe *et al.*, 1983) (where it is probably computed from the visual neurons that respond to expected reward and the taste neurons which signal the reward or punishment actually obtained (Thorpe *et al.*, 1983; Rolls *et al.*, 1996; Rolls, 1997, 2004a). Instead, the mechanism for reversal of the rule neurons described here involves just non-specific, probably feedback, inhibition to quench the recurrently connected neurons that implement the current rule attractor, and partial synaptic adaptation which has been taking place since the previous rule change to ensure that the attractor restarts with the neurons that represent the alternative rule being active. Because the rule neurons act as a biased competition gating input to the stimulus–reward combination neurons that map the sensory input to the reward neurons, the very next time after an error trial that the other stimulus is presented, it is correctly mapped as being reward-related (Thorpe *et al.*, 1983), even though it has not been recently associated with delivery of a reward. Secondly, the theory presented in this paper is implemented at the level of biologically realistic synaptic dynamics and spiking neuronal activity, and thus provides a realistic model of the actual dynamical processes occurring in the brain which goes beyond the more artificial dynamics and learning protocols of connectionist schemes (Miller and Cohen, 2001; O'Reilly *et al.*, 2002). Thirdly, the model described here provides an explanation for the visual stimulus–reward combination neurons in the orbitofrontal cortex described by Thorpe *et al.* (1983) (and olfactory stimulus–reward combination neurons described in primates by Rolls *et al.* (1996) and reported as also being present in what

may or may not be an anatomically and functionally homologous area in rats by Schoenbaum *et al.* (1999). These stimulus–reward combination neurons respond to one stimulus when it is associated with reward, and not to a different stimulus when it is associated with reward. These neurons are important to how the model described here functions, for these combination neurons receive the biasing input from the rule module, and enable the mapping from stimulus to reward to be changed immediately when the reward rule module changes its state, because this is a biased competition mechanism which operates dynamically when the biasing input changes with no need for any further synaptic modification. This biased competition switching of the stimulus–reward mapping by the rule module is at the heart of the process described here, and it is a strength of the theory presented here that it gives an account of the presence and need for the stimulus–reward combination neurons in a part of the brain important in implementing rapid, one-trial, stimulus–reward switching, the orbitofrontal cortex (Thorpe *et al.*, 1983; Rolls *et al.*, 1994, 1996; Hornak *et al.*, 2004; Rolls, 2004a).

In this paragraph, we contrast in more detail the mechanism of switching used in our model with that used by O'Reilly *et al.* (2002) and Rougier and O'Reilly (2002). That model described the role of the prefrontal cortex in task switching in terms of the greater flexibility conferred by activity-based working memory representations in the prefrontal cortex, as compared with more slowly adapting weight-based memory mechanisms. In particular, in their model the prefrontal cortex representations could be rapidly updated when a task switches via a dynamic gating mechanism based on a temporal-difference reward–prediction learning mechanism. This dynamic gating mechanism was essential in their model for controlling the updating and maintenance of working memory representations. When the gate is open, working memory can be updated, and when it is closed, any currently active working memories are protected from interference. They claim that this gate is needed because one setting of connection strengths into the working memory system cannot satisfy both the need for rapid updating and robust maintenance (O'Reilly and Munakata, 2000). They implemented this gating mechanism by means of a multiplicative term implemented either through phasic dopamine neuromodulation of the frontal cortex by the ventral tegmental area (O'Reilly *et al.*, 2002), or through the interactions between the basal ganglia and frontal cortex (Frank *et al.*, 2001). Thus, in that model, the error signal causes synaptic weights to be modulated multiplicatively to switch in another rule. In contrast, in the new model described here, the error signal just quenches the ongoing neural activity in the short term attractor based memory (using for example inhibition implemented through GABA inputs). After this general inhibition, the other attractor then emerges from the ongoing spontaneous neuronal activity, because there has been ongoing partial neuronal or synaptic adaptation since the last rule change. Moreover, in the model described here, there is no reliance on other brain systems such as the basal ganglia or ventral tegmental dopamine neurons to compute and/or represent the non-reward signal (cf. Rolls, 1999), for this error signal is already explicitly represented in the primate orbitofrontal cortex by neurons which we have shown respond whenever an expected reward is not obtained (Thorpe *et al.*, 1983), as are the signals required to compute this of whether a reward is expected on any given trial given the visual stimulus shown (Thorpe *et al.*, 1983), and whether the

reward is obtained as shown by the firing of taste neurons (Thorpe *et al.*, 1983; Rolls *et al.*, 1990). Indeed, the neurons that change their mappings in the striatum during reward reversal (Rolls *et al.*, 1983; Schoenbaum and Setlow, 2003; Setlow *et al.*, 2003; cf. Divac *et al.*, 1967; Cools *et al.*, 2002) appear to reflect the change of mapping received from direct inputs from the orbitofrontal to the striatum rather than the computation that a change of mapping is required, as argued in the original paper of Rolls *et al.* (1983) in which the neuronal recordings in the caudate nucleus were made in the same task and in some of the same macaques as those used in the original orbitofrontal cortex recordings of Thorpe *et al.* (1983). Whereas all the signals required for the computation were present in the orbitofrontal cortex, they were not in the striatum (Rolls *et al.*, 1983; Thorpe *et al.*, 1983). Taking these findings and the connective anatomy together, it is much more likely that neuronal activity in the basal ganglia, including the dopamine neurons in the tegmentum (which receive inputs from the striatum), reflects activity in the orbitofrontal cortex, rather than being computed in the basal ganglia, and this has important implications for understanding the functions of the orbitofrontal cortex versus the basal ganglia (Rolls, 1999).

In this paragraph, we contrast in more detail the actual implementation of our model, which goes right to the level of synaptic and spiking dynamics, and can therefore describe the dynamics accurately and can allow the neuronal spiking activity in the model to be directly compared to single neuron recordings, with previous models. There are several essential differences in our model. Our implementation is based on explicit and realistic biophysical process, namely on explicit description of the synaptic dynamics (AMPA, NMDA and GABA) and spiking mechanisms, constrained by the experimentally measured biophysical parameters (e.g. latencies, synaptic and membrane conductances, reversal potential etc.). The implementations of Frank *et al.* (2001), O'Reilly *et al.* (2002) and Rougier and O'Reilly (2002) are based on the 'Leabra' framework (O'Reilly and Munakata, 2000), which is biologically motivated but does not consider the neuronal spiking mechanisms explicitly. In fact, they use a rate-based approach, which is not consistent with the spiking level, as has been thoroughly analysed recently (Brunel and Wang, 2001). Even more, rate-based approaches are only valid under stationary conditions, and are not able to describe the non-stationary temporal dynamics, which is the main goal here. The use of rate-based approaches has been recently extensively studied (Brunel and Wang, 2001; Del Giudice *et al.*, 2003), and the importance of deriving rate-response functions which are consistent with the spiking and dynamics background has been stressed, a procedure which we used here to find the parameters of the model. In order to do this, a mean-field approach has to be used. The essence of the mean-field approximation is to simplify the integrate-and-fire equations by replacing, after the diffusion approximation (Tuckwell, 1988), the sums of the synaptic components by the average DC component and a fluctuation term. The stationary dynamics of each population can be described by the *population transfer function*, which provides the average population rate as a function of the average input current. The set of stationary, self-reproducing rates  $v_i$  for the different populations  $i$  in the network can be found by solving a set of coupled self-consistency equations. This enables *a posteriori* selection of the parameter region which shows in the bifurcation diagram the emergent behaviour that is to be investigated (e.g. biased

competition). After that, with this set of parameters, we perform the full non-stationary simulations using the *true dynamics* only described by the full integrate-and-fire scheme. The mean field study assures us that this dynamics will converge to a stationary attractor that is consistent with what is to be investigated (Brunel and Wang, 2001; Del Giudice *et al.*, 2003). Therefore, in the work described here, we used a mean-field approximation to explore how the different operational regimes of the network depend on the values of certain parameters. The mean-field analysis performed in this work uses the formulation derived in (Brunel and Wang, 2001), which is consistent with the network of neurons used. Their formulation departs from the equations describing the dynamics of one neuron to provide a stochastic analysis of the mean-first passage time of the membrane potentials, which results in a description of the population spiking rates as functions of the model parameters. In conclusion, the only way to perform a non-stationary analysis of the temporal dynamics, for direct comparison with the neuronal recording results, is via the 'explicit' use of the synaptic and spiking mechanisms as incorporated in the model described here, after a prior analysis of a consistent mean-field derived approach for the analysis of the stationary attractors.

An interesting computational model was introduced by Dehaene and Changeaux (1991) for modelling task switching in the context of the Wisconsin Card Sorting Test. This model also predicts the existence of rule-neurons, and can perform more complex problems than single-trial reversals. The model keeps a memory of the rules, so that a rule is not tested twice if it was rejected in the immediate past, and can even discard rules by 'reasoning'. The model is biologically motivated and its architecture is biologically plausible, but its underlying mechanisms are not based on explicit synaptic and spiking mechanisms, but again on simple (inconsistent with the spiking level) forms of rate-based response functions. The switching mechanism incorporated was also essentially different, because the reward response, in the case of Dehaene and Changeaux (1991), elicited little synaptic adaptation that was a direct cause of the switching. In our case, the switching is due to a bi-stable dynamics in the rule module. In our model, the unspecific reward error signal just resets the system, and the rate frequency adaptation, or short term synaptic depression mechanism, just maintains the information about the last rule used, and is always continuously active and not influenced by reward responses. So, even in the case when we use short-term synaptic adaptation rather than neuronal adaptation, the synaptic mechanisms are not responsible for the switching, which is implemented by a combination of the error signal quenching the rule attractor, and then the other rule attractor emerging out of the noise.

The main goal of this paper was to study a putative complex dynamics underlying simple switching (two rules) with realistic synaptic and spiking mechanisms. The extension to more complex tasks with more than one rule is a challenging goal, because the mechanisms studied by Dehaene and Changeaux (1991) (like reasoning and keeping the memory of more than one rule previously tried) would be extremely relevant. Even more, we believe that an optimal extension of our model would be in integrating the basic ideas of Dehaene *et al.* (1991) in our dynamical model. This does appear to be feasible in principle, in that we have already performed investigations which show that the mechanisms utilized here for keeping memory of the previously activated rules can be extended straightforwardly to the case of at least 3-4 rules.

The research described in this paper also adds considerably to our own previous model of the operation of the prefrontal cortex (Deco and Rolls, 2003) as follows. First, that model was of the primate dorsolateral prefrontal cortex, and the model described here is (apart from Fig. 6) of the orbitofrontal cortex. Secondly, Deco and Rolls (2003) modelled how a contextual biasing input can select different fixed mappings between stimuli and responses in a way which is consistent with neuronal recordings in the macaque dorsolateral prefrontal during these tasks. In contrast, the model described here is of stimulus–reward association learning, which is stimulus–stimulus learning, given that rewards are stimuli such as taste (Rolls, 1999). Stimulus–reward association learning is prototypical of that involved in emotional learning, in which we update our representations of whether a stimulus in the world is still currently associated with reward, and use this to control many functions, including autonomic and endocrine functions, and whether the stimulus should be a goal for flexible actions which can be produced by new instrumental learning (Rolls, 2000c, 1999, 2004b). Third, the model described here shows how the rule module can be switched from one state to another when expected reward is not obtained, which was not addressed at all in our earlier work (Deco and Rolls, 2003). A mechanism for switching the rule module when expected reward is not obtained is crucial as the biasing input to the biased competition process, whether that process is involved in stimulus–reward switching, as described in this paper, or in stimulus–response switching as described by Deco and Rolls (2003) and in Figure 6 of this paper.

## Notes

This research was supported by the Medical Research Council (grants to E.T.R.) and by the German Ministry for Research and the European Union (grants to G.D.).

Address correspondence to Professor Edmund T. Rolls, University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, UK. Email: Edmund.Rolls@psy.ox.ac.uk.

## Appendix

In this appendix we give the mathematical equations that describe the spiking activity and synapse dynamics in the network, following in general the formulation described by Brunel and 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 equation:

$$C_m \frac{dV(t)}{dt} = -g_m(V(t) - V_L) - I_{\text{syn}}(t) \quad (4)$$

where  $I_{\text{syn}}(t)$  is the total synaptic current flow into the cell,  $V_L$  is the resting potential,  $C_m$  is the membrane capacitance and  $g_m$  is the membrane conductance. When the membrane potential  $V(t)$  reaches the threshold  $\theta$  a spike is generated, and the membrane potential is reset to  $V_{\text{reset}}$ . The neuron is unable to spike during the first  $\tau_{\text{ref}}$  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 ( $I_{\text{AMPA,ext}}$ ), while the excitatory recurrent synapses operate through AMPA and NMDA receptors ( $I_{\text{AMPA,rec}}$  and  $I_{\text{NMDA,rec}}$ ). The total synaptic current is therefore given by:

$$I_{\text{syn}}(t) = I_{\text{AMPA,ext}}(t) + I_{\text{AMPA,rec}}(t) + I_{\text{NMDA,rec}}(t) + I_{\text{GABA}}(t) \quad (5)$$

where

$$I_{\text{AMPA,ext}}(t) = g_{\text{AMPA,ext}}(V(t) - V_E) \sum_{j=1}^{N_{\text{ext}}} s_j^{\text{AMPA,ext}}(t) \quad (6)$$

$$I_{\text{AMPA,rec}}(t) = g_{\text{AMPA,rec}}(V(t) - V_E) \sum_{j=1}^{N_E} w_j s_j^{\text{AMPA,rec}}(t) \quad (7)$$

$$I_{\text{NMDA,rec}}(t) = \frac{g_{\text{NMDA}}(V(t) - V_E)}{(1 + C_{Mg^{2+}} \exp(-0.062V(t)/3.57))} \times \sum_{j=1}^{N_E} w_j s_j^{\text{NMDA}}(t) \quad (8)$$

In the preceding equations,  $V_E = 0$  mV and  $V_L = -70$  mV. The synaptic strengths  $w_j$  are specified in the Methods section and in Table 1. The fractions of open channels  $s$  are given by:

$$\frac{ds_j^{\text{AMPA,ext}}(t)}{dt} = -\frac{s_j^{\text{AMPA,ext}}(t)}{\tau_{\text{AMPA}}} + \sum_k \delta(t - t_j^k) \quad (9)$$

$$\frac{ds_j^{\text{AMPA,rec}}(t)}{dt} = -\frac{s_j^{\text{AMPA,rec}}(t)}{\tau_{\text{AMPA}}} + \sum_k \delta(t - t_j^k) \quad (10)$$

$$\frac{ds_j^{\text{NMDA}}(t)}{dt} = -\frac{s_j^{\text{NMDA}}(t)}{\tau_{\text{NMDA,decay}}} + \alpha x_j(t)(1 - s_j^{\text{NMDA}}(t)) \quad (11)$$

$$\frac{dx_j(t)}{dt} = -\frac{x_j(t)}{\tau_{\text{NMDA,rise}}} + \sum_k \delta(t - t_j^k) \quad (12)$$

$$\frac{ds_j^{\text{GABA}}(t)}{dt} = -\frac{s_j^{\text{GABA}}(t)}{\tau_{\text{GABA}}} + \sum_k \delta(t - t_j^k) \quad (13)$$

where the sums over  $k$  represent a sum over spikes emitted by presynaptic neuron  $j$  at time  $t_j^k$ . The value of  $\alpha = 0.5$  ms<sup>-1</sup>.

The values of the conductances for pyramidal neurons in the associative module were:  $g_{\text{AMPA,ext}} = 2.08$ ,  $g_{\text{AMPA,rec}} = 0.052$ ,  $g_{\text{NMDA}} = 0.164$  and  $g_{\text{GABA}} = 0.72$ ; and for interneurons:  $g_{\text{AMPA,ext}} = 1.62$ ,  $g_{\text{AMPA,rec}} = 0.0405$ ,  $g_{\text{NMDA}} = 0.129$  and  $g_{\text{GABA}} = 0.487$  nS.

The values of the conductances for pyramidal neurons in the rule module were:  $g_{\text{AMPA,ext}} = 2.08$ ,  $g_{\text{AMPA,rec}} = 0.104$ ,  $g_{\text{NMDA}} = 0.328$  and  $g_{\text{GABA}} = 1.44$ ; and for interneurons:  $g_{\text{AMPA,ext}} = 1.62$ ,  $g_{\text{AMPA,rec}} = 0.081$ ,  $g_{\text{NMDA}} = 0.258$  and  $g_{\text{GABA}} = 0.973$  nS.

## References

- Abbott LF, Nelson SB (2000) Synaptic plasticity: taming the beast. *Nat Neurosci* 3:1178–1183.
- Abeles A (1991) *Corticonics*. New York: Cambridge University Press.
- Amit D, Brunel N (1997) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb Cortex* 7:237–252.
- Asaad WF, Rainer G, Miller EK (1998) Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21:1399–1407.
- Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451–459.
- Battaglia F, Treves A (1998) Stable and rapid recurrent processing in realistic autoassociative memories. *Neural Computat* 10:431–450.
- Brunel N, Wang X (2001) Effects of neuromodulation in a cortical networks model of object working memory dominated by recurrent inhibition. *J Computat Neurosci* 11:63–85.
- 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* 363:345–347.
- Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *J Neurosci* 22:4563–4567.

- Corchs S, Deco G (2002) Large-scale neural model for visual attention: integration of experimental single cell and fMRI data. *Cereb Cortex* 12:339-348.
- Critchley HD, Rolls ET (1996) Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J Neurophysiol* 75:1673-1686.
- Damasio AR (1994) *Descartes' error*. New York: Putnam.
- Dayan P, Abbott LF (2002) *Theoretical neuroscience*. Cambridge, MA: MIT Press.
- Deco G, Lee T (2002) A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing* 44-46:775-781.
- Deco G, Rolls ET (2003) Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. *Eur J Neurosci* 18:2374-2390.
- Deco G, Zihl J (2001) Top-down selective visual attention: a neuro-dynamical approach. *Vis Cogn* 8:119-140.
- Deco G, Rolls ET, Horwitz B (2004) 'What' and 'where' in visual working memory: a computational neurodynamical perspective for integrating fMRI and single-neuron data. *J Cogn Neurosci* 16:683-701.
- Dehaene S, Changeaux JP (1991) The Wisconsin card sorting test: theoretical analysis and modelling in a neural network. *Cereb Cortex* 1:62-79.
- Del Giudice P, Fusi S, Mattia M (2003) Modelling the formation of working memory with networks of integrate-and-fire neurons connected by plastic synapses. *J Physiol Paris* 97:659-681.
- Destexhe A, Mainen Z, Sejnowski T (1998) Kinetic models of synaptic transmission. In: *Methods in neural modelling: from ions to networks*, 2nd edn (Koch C, Segev I, eds), pp. 1-25. Cambridge, MA: MIT Press.
- Divac I, Rosvold HE, Szwarcbart MK (1967) Behavioural effects of selective ablation of the caudate nucleus. *J Comp Physiol Psychol* 63:184-190.
- Frank M, Loughry B, O'Reilly R (2001) Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cogn Affect Behav Neurosci* 1:137-160.
- Giugliano M, La Camera G, Rauch A, Luescher H-R, Fusi S (2002) Non-monotonic current-to-rate response function in a novel integrate-and-fire model neuron. In: *Proceedings of ICANN 2002, LNCS 2415* (Dorransoro J, ed.), pp. 141-146. New York: Springer.
- Hansel D, Mato G, Meunier C, Neltner L (1998) On numerical simulations of integrate-and-fire neural networks. *Neural Computat* 10:467-483.
- 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 JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Nat Acad Sci USA* 79:2554-2558.
- Hornak J, Bramham J, Rolls ET, Morris R G, O'Doherty J, Bullock PR, Polkey CE (2003) Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain* 126:1691-1712.
- Hornak J, O'Doherty J, Bramham J, Rolls ET, Morris RG, Bullock PR, Polkey CE (2004) Reward-related reversal learning after surgical excisions in orbitofrontal and dorsolateral prefrontal cortex in humans. *J Cogn Neurosci* 16:463-478.
- Iversen S D, Mishkin M (1970) Perseverative interference in monkey following selective lesions of the inferior prefrontal convexity. *Exp Brain Res* 11:376-386.
- Jahr C, Stevens C (1990) Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *J Neurosci* 10:3178-3182.
- Kringelbach ML, Rolls ET (2003) Neural correlates of rapid reversal learning in a simple model of human social interaction. *Neuroimage* 20:1371-1383.
- 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.
- Lisman JE, Fellous JM, Wang XJ (1998) A role for NMDA-receptor channels in working memory. *Nat Neurosci* 1:273-275.
- Liu Y, Wang X-J (2001) Spike-frequency adaptation of a generalized leaky integrate-and-fire model neuron. *J Computat Neurosci* 10:25-45.
- 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 EK, Cohen JD (2001) An integrative theory of prefrontal functions. *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 (1964) Some effects of frontal lobectomy in man. In: *The frontal granular cortex and behavior* (Warren JM, Akert K, eds), pp. 313-314. New York: McGraw-Hill.
- Mishkin M, Manning FJ (1978) Non-spatial memory after selective prefrontal lesions in monkeys. *Brain Res* 143:313-324.
- 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.
- Nauta WJ (1971) The problem of the frontal lobe: a reinterpretation. *J Psychiatry Res* 8:167-187.
- O'Reilly J, Munakata Y (2000) *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.
- O'Reilly RC, Noelle DC, Braver TS, Cohen JD (2002) Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. *Cereb Cortex* 12:246-257.
- Rauch A, La Camera G, Luescher H-R, Senn W, Fusi S (2003) Neocortical pyramidal cells respond as integrate-and-fire neurons in *in vivo*-like input currents. *J Neurophysiol* 90:1598-1612.
- Reynolds J, Desimone R (1999) The role of neural mechanisms of attention in solving the binding problem. *Neuron* 24:19-29.
- Rolls ET (1974) The neural basis of brain-stimulation reward. *Prog Neurobiol* 3:71-160.
- Rolls ET (1975) *The brain and reward*. Oxford: Pergamon Press.
- Rolls ET (1990) A theory of emotion and its application to understanding the neural basis of emotion. *Cogn Emotion* 4:161-190.
- Rolls ET (1997) Taste and olfactory processing in the brain and its relation to the control of eating. *Crit Rev Neurobiol* 11:263-287.
- Rolls ET (1999) *The brain and emotion*. Oxford: Oxford University Press.
- Rolls ET (2000a) Neurophysiology and functions of the primate amygdala and the neural basis of emotion. In: *The amygdala: a functional analysis*, 2nd edn (Aggleton, J ed.), pp. 447-478. Oxford: Oxford University Press.
- Rolls ET (2000b) The orbitofrontal cortex and reward. *Cereb Cortex* 10:284-294.
- Rolls ET (2000c) *Précis of The Brain and Emotion*. *Behav Brain Sci* 23:177-233.
- Rolls ET (2004a) The functions of the orbitofrontal cortex. *Brain Cogn* 55:11-29.
- Rolls ET (2004b) What are emotions, why do we have them and what are their brain mechanisms. In: *Who needs emotion: the brain meets the machine* (Fellous M, Arbib M, eds). New York: Oxford University Press (in press).
- Rolls ET, Deco G (2002) *Computational neuroscience of vision*. Oxford: Oxford University Press.
- Rolls ET, Treves A (1998) *Neural networks and brain function*. Oxford: Oxford University Press.
- Rolls ET, Thorpe SJ, Maddison SP (1983) Responses of striatal neurons in the behaving monkey I head of the caudate nucleus. *Behav Brain Res* 7:179-210.
- Rolls ET, Yaxley S, Sienkiewicz ZJ (1990) Gustatory responses of single neurons in the orbitofrontal cortex of the macaque monkey. *J Neurophysiol* 64:1055-1066.
- Rolls ET, Hornak J, Wade D, McGrath J (1994) Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. *J Neurol Neurosurg Psychiatry* 57:1518-1524.
- Rolls ET, Critchley HD, Mason R, Wakeman EA (1996) Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J Neurophysiol* 75:1970-1981.

- Rougier N, O'Reilly R (2002) Learning representations in a gated prefrontal cortex model of dynamic task switching. *Cogn Sci* 26:503-520.
- Salin P, Prince D (1996) Spontaneous GABA-A receptor mediated inhibitory currents in adult rat somatosensory cortex. *J Neurophysiol* 75:1573-1588.
- Schoenbaum G, Setlow B (2003) Lesions of nucleus accumbens disrupt learning about aversive outcomes. *J Neurosci* 23:9833-9841.
- Schoenbaum G, Chiba AA, Gallagher M (1999) Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J Neurosci* 19:1876-1884.
- Setlow B, Schoenbaum G, Gallagher M (2003) Neural encoding in ventral striatum during olfactory discrimination learning. *Neuron* 38:625-636.
- 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* 482:325-352.
- Sutton RS, Barto AG (1998) Reinforcement learning. Cambridge, MA: MIT Press.
- Thorpe SJ, Rolls ET, Maddison S (1983) Neuronal activity in the orbitofrontal cortex of the behaving monkey. *Exp Brain Res* 49:93-115.
- Tuckwell H (1988) Introduction to theoretical neurobiology. Cambridge: Cambridge University Press.
- Wang X (1999) Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J Neurosci* 19:9587-9603.
- Wilson FAW, O'Scalaidhe SP, 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* 506:715-730.
- Zheng P, Zhang X-X, Bunney BS, 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.