Progress in Neurobiology 76 (2005) 236–256

progress in Neurobiology

www.elsevier.com/locate/pneurobio

Attention, short-term memory, and action selection: A unifying theory

Gustavo Deco[a], Edmund T. Rolls[b,*]

[a] Institució Catalana de Recerca i Estudis Avançats (ICREA), Universitat Pompeu Fabra, Department of Technology, Computational Neuroscience, Passeig de Circumval.lació, 8, 08003 Barcelona, Spain

[b] University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, UK

Received 25 February 2005; received in revised form 22 June 2005; accepted 24 August 2005

Abstract

Cognitive behaviour requires complex context-dependent processing of information that emerges from the links between attentional perceptual processes, working memory and reward-based evaluation of the performed actions. We describe a computational neuroscience theoretical framework which shows how an attentional state held in a short term memory in the prefrontal cortex can by top-down processing influence ventral and dorsal stream cortical areas using biased competition to account for many aspects of visual attention. We also show how within the prefrontal cortex an attentional bias can influence the mapping of sensory inputs to motor outputs, and thus play an important role in decision making. We also show how the absence of expected rewards can switch an attentional bias signal, and thus rapidly and flexibly alter cognitive performance. This theoretical framework incorporates spiking and synaptic dynamics which enable single neuron responses, fMRI activations, psychophysical results, the effects of pharmacological agents, and the effects of damage to parts of the system to be explicitly simulated and predicted. This computational neuroscience framework provides an approach for integrating different levels of investigation of brain function, and for understanding the relations between them. The models also directly address how bottom-up and top-down processes interact in visual cognition, and show how some apparently serial processes reflect the operation of interacting parallel distributed systems.

Keywords: Attention; Short-term memory; Executive function; Task switching; Biased competition; Decision-making

Contents

* Corresponding author. Tel.: +44 1865 271348; fax: +44 1865 310447.
E-mail address: edmund.rolls@psy.ox.ac.uk (E.T. Rolls).

0301-0082/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.pneurobio.2005.08.004

Understanding the fundamental principles underlying higher brain functions requires the integration of different levels of experimental investigation in cognitive neuroscience (from the operation of single neurons and neuroanatomy, neurophysiology, neuroimaging and neuropsychology to behaviour) via a unifying theoretical framework that captures the neural dynamics inherent in the computation of cognitive processes. A theoretical framework that fulfils these requirements can be obtained by developing explicit mathematical neurodynamical models of brain function based at the level of neuronal spiking and synaptic activity (Rolls and Deco, 2002; Dayan and Abbott, 2001). In this article we show how this approach is being used to produce a unified theory of visual attention and working memory, and how these processes are influenced by rewards to influence decision making.

In particular, this review focusses on a computational neuroscience perspective that adds to the general framework of the *Biased Competition Hypothesis* (Moran and Desimone, 1985; Spitzer et al., 1988; Motter, 1993; Miller et al., 1993; Chelazzi et al., 1993; Reynolds and Desimone, 1999; Chelazzi, 1998; Rolls and Deco, 2002). In this approach (Rolls and Deco, 2002) multiple activated populations of neurons which may be hierarchically organized engage in competitive interactions, and external top-down effects bias this competition in favour of specific neurons. The aim of the framework is to show how the integration of evidence from low levels of investigation (including detailed evidence of what is represented by the computing elements of the brain, single neurons) is taken to be very relevant to the computational theory. This is somewhat in contrast to the approach advocated by Marr (1982), who proposed that the first stage should be specification of the computational theory, the second specification of the algorithm, and the third specification of the implementation. In this review, we show that the experimental evidence and what is plausible in biological systems provide important constraints on the computational theory. Moreover, again based on the experimental evidence, instead of adopting a primarily bottom-up or feed-forward approach (Marr, 1982), we incorporate top-down and interactive effects in the theory. Computational neuroscience approaches to other aspects of visual cognition including bottom-up saliency map approaches to attention (Itti and Koch, 2001; Wolfe, 1994) and the computation of invariant representations (Wallis and Rolls, 1997; Riesenhuber and Poggio, 2000; Salinas and Abbott, 1997; Wiskott and Sejnowski, 2002; Rolls and Milward, 2000; Rolls and Deco, 2002) are described elsewhere (Buracas et al., 1996; O'Reilly and Munakata, 2000; Rao et al., 1997).

1. The cognitive neuroscience of attention

One type of attentional process operates when salient features in a visual scene attract attention (Itti and Koch, 2001). This visual processing is described as feedforward and bottom-up, in that it operates forward in the visual pathways from the visual input (see Fig. 1). A second type of selective attentional process, with which we are concerned here, involves actively maintaining in short-term memory a location or object as the target of attention, and using this by top-down processes to influence earlier cortical processing.

One approach to top-down selective attention utilizes the metaphor of a spotlight that 'illuminates' a portion of the field of view where stimuli are processed in higher detail (Helmholtz, 1967; Treisman, 1982). This has been developed into a *Feature Integration Theory* of visual selective attention with two processing stages (Treisman, 1988; Treisman and Gelade, 1980). The first *preattentive* process runs in *parallel* across the complete visual field extracting single primitive features without integrating them. The second *attentive* stage corresponds to the *serial* specialized integration of information from a limited part of the field at any one time. Evidence for these two stages of attentional visual processing comes from psychophysical experiments using visual search tasks where subjects examine a display containing randomly positioned items in order to detect an a priori defined target.

A second approach, the *biased competition* model (Duncan and Humphreys, 1989; Duncan, 1996), states that the multiple stimuli in the visual field activate populations of visual cortical neurons that engage in competitive interactions, and that attentional signals from outside the visual cortex bias the competition such that the cells representing the attended stimulus *win* (Duncan and Humphreys, 1989; Duncan, 1996). This competition works in parallel across the visual field as shown by psychophysical experiments (Duncan et al., 1997; Mozer and Sitton, 1998; Phaf et al., 1990; Tsotsos, 1990). Neurophysiological experiments from extrastriate areas are consistent with the biased competition hypothesis in showing that attention serves to modulate the suppressive interaction between two or more stimuli within the receptive field (Moran and Desimone, 1985; Spitzer et al., 1988; Sato, 1989; Motter, 1993; Miller et al., 1993; Chelazzi et al., 1993; Motter, 1994; Reynolds and Desimone, 1999; Chelazzi, 1998). For example, Moran and Desimone (1985) showed that the firing activity of visually tuned neurons in the extrastriate cortex was modulated if monkeys were instructed to attend to the location of the target stimulus, and this result is found in V4 (Luck et al., 1997; Reynolds et al., 1999), V2 (Reynolds et al., 1999), and even weakly in V1 (McAdams and Maunsell, 1999) (see Fig. 1). Direct physiological evidence for the existence of an attentional bias was provided by Luck et al. (1997) and by Spitzer et al. (1988). Luck et al. (1997) showed that attending to a location within the receptive field of a V2 or V4 neuron increased its spontaneous firing activity by a small amount. When a single stimulus is presented in the receptive field of a V4 neuron, Spitzer et al. (1988) observed an increase of the neuronal response to that stimulus when the monkey directed attention inside the receptive field, compared to when attention was directed outside the field. Evidence for the biased competition hypothesis as a mechanism for the selection of non-spatial, object-related attributes is that inferior temporal cortex (IT) neurons in monkeys respond more to a target than to a distractor (Chelazzi et al., 1993). Attention also influences the responses of neurons in the dorsal visual stream (Duhamel et al., 1992; Colby et al., 1993; Gnadt and Andersen, 1988), and Maunsell

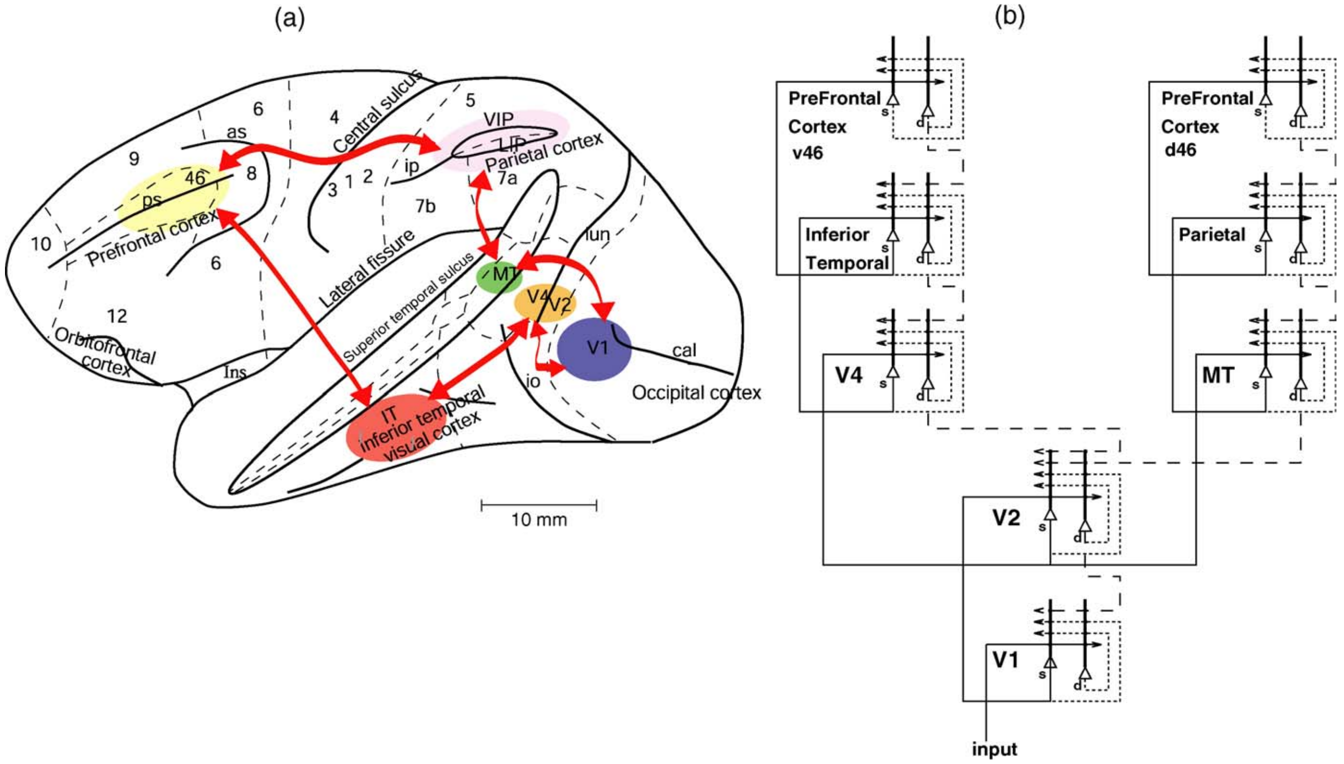


Fig. 1. (a) Lateral view of the macaque brain showing the connections in the ventral and dorsal visual streams (V1: primary visual area, V2, V4: extrastriate visual areas, IT: inferior temporal, PP: posterior parietal) and the anterior (PFCv: ventrolateral prefrontal cortex, PFCd: dorsolateral prefrontal cortex) brain areas. (b) The systems-level architecture of a model of the cortical mechanisms of visual attention and memory. The system is essentially composed of six modules which model the two known main visual pathways of the primate visual cortex. Information from the retina via the lateral geniculate nucleus enters the visual cortex through area V1 in the occipital cortex and proceeds into two forward or bottom-up processing streams. The occipital-temporal stream leads ventrally through V2–V4 and IT, and is mainly concerned with object recognition. The occipito-parietal stream leads dorsally into PP and is responsible for maintaining a spatial map of an object's location. Both posterior visual pathways send connections to and receive connections from the lateral prefrontal cortex, where short-term memory functions take place. Forward connections are indicated by solid lines; backprojections, which could implement top-down processing, by dashed lines; and recurrent connections within an area by dotted lines. s—superficial pyramidal cells; d—deep pyramidal cells.

(1995) demonstrated that the biased competitive interaction due to spatial and object attention exists not only for objects within the same receptive field of MT neurons, but also for objects in spatially distant receptive fields. This suggests that mechanisms exist to provide biased competition of a more global nature in the dorsal visual stream.

Further evidence comes from functional magnetic resonance imaging (fMRI) in humans (Kastner et al., 1998, 1999) which indicates that when multiple stimuli are present simultaneously in the visual field, their cortical representations within the object recognition pathway interact in a competitive, suppressive fashion, which is not the case when the stimuli are presented sequentially. It was also observed that directing attention to one of the stimuli counteracts the suppressive influence of nearby stimuli.

There is evidence that visual object recognition can operate in a largely feedforward mode (Marr, 1982; Rolls, 2000a). For example, Rolls et al. (1994) (see Rolls (2003)) showed in a backward masking paradigm that when humans could just identify faces, inferior temporal cortex neurons fire for only 30 ms, leaving insufficient time for information to return from the inferior temporal cortex to V1 to contribute to object identification. Also, Thorpe et al. (1996) demonstrated that people and monkeys could perform categorization tasks very rapidly, and that the event-related potentials (ERP) relevant to decision making emerge in the prefrontal areas within 150 ms, leaving little time for feedback to work its way back down the visual hierarchy. Further, many object recognition algorithms (Riesenhuber and Poggio, 2000; Wallis and Rolls, 1997; Rolls and Milward, 2000; Rolls and Deco, 2002) are based only on feedforward algorithms. In contrast, with respect to attention, our proposal here stresses that feedback is an inherent and integrated part of covert attentional processes including object localization and recognition when there are multiple objects in a visual scene.[1] Consistent with this proposal, there is evidence that top-down effects can influence early cortical visual processing quite rapidly (Haenny and Schiller, 1988; Motter, 1993; Lamme, 1995; Zipser et al., 1996; Lee et al., 1998; Ito and Gilbert, 1999; Roelfsema et al., 1998; Hupe et al., 1998; Lee and Nguyen, 2001; Lee et al., 2002).

[1] Covert attention refers to attentional processes that occur in the absence of eye movements. Active visual search using eye movements (overt search) is often necessary in complex natural scenes, because the inferior temporal visual cortex represents primarily what is close to the fovea (Rolls et al., 2003).

2. The computational neuroscience of visual attention

What are the theoretical tools for achieving the proper level of description of the neurodynamical mechanisms that underlie brain functions? On one hand, the level of description should be accurate enough to allow the relevant mechanisms at the level of neurons and synapses to be properly taken into account. On the other hand, the description should be simple enough, so that we can really infer by abstraction the *relevant principles* substantiating perception and cognition. A mere reproduction of phenomena of a complex system, like the brain, just by simulating the same kind of artificial complex system, is most of the time not useful, because there are usually no explicit underlying *First Principles*, and it is also unrealistic.

We assume, with others (Tuckwell, 1988; Brunel and Wang, 2001; Amit and Brunel, 1997; Del Giudice et al., 2003; Brunel and Wang, 2001), that a proper level of description at the microscopic level is captured by the spiking and synaptic dynamics of one-compartment, point-like models of neurons, such as *Integrate-and-Fire Models*. The realistic dynamics allow the use of realistic biophysical constants (like conductances, delays, etc.) in a thorough study of the realistic time scales and firing rates involved in the evolution of the neural activity underlying cognitive processes, for comparison with experimental data. We believe that it is essential of a *biologically plausible* model that the different time scales involved are properly described, because the system that we are describing is a dynamical system that is sensitive to the underlying different spiking and synaptic time courses, and the non-linearities involved in these processes. For this reason, it is convenient to include a thorough description of the different time courses of the synaptic activity, by including fast and slow excitatory receptors (AMPA and NMDA) and GABA-inhibitory receptors. A second reason why this temporally realistic and detailed level of description of synaptic activity is required is the goal to perform realistic fMRI-simulations. These involve the realistic calculation of BOLD-signals that are intrinsically linked with the synaptic dynamics, as recently found by Logothetis et al. (2001). A third reason is that one can consider the influence of neurotransmitters and pharmacological manipulations, e.g. the influence of dopamine on the NMDA and GABA receptor dynamics (Zheng et al., 1999; Law-Tho et al., 1994), to study the effect on the global dynamics and on the related cortical functions (e.g. working memory Deco et al. (2004), and Deco and Rolls (2003)). A fourth reason for analysis at the level of spiking neurons is that the computational units of the brain are the neurons, in the sense that they transform a large set of inputs received from different neurons into an output spike train, that this is the single output signal of the neuron which is connected to other neurons, and that this is therefore the level at which the information is being transferred between the neurons, and thus at which the brain's representations and computations can be understood (Rolls and Treves, 1998; Rolls and Deco, 2002).

The ways in which we implement the models at the Integrate-and-Fire level are described in Appendix A.

2.1. A functional architecture for visual attention

We consider here top-down processes in attention, and how they interact with bottom-up processing, in a model of visual attentional processing which has multiple hierarchically organized modules in the architecture shown schematically in Fig. 1b. The model shows how the dorsal (sometimes called *where*) visual stream (reaching the posterior parietal cortex, PP) and the ventral (*what*) visual stream (via V4 to the inferior temporal cortex, IT) could interact through early visual cortical areas (such as V1 and V2) to account for many aspects of visual attention (Deco, 2001; Deco and Zihl, 2001; Rolls and Deco, 2002; Deco and Lee, 2002; Corchs and Deco, 2002; Heinke et al., 2002; Deco and Lee, 2004). The system modelled is essentially composed of six modules (V1 (the primary visual cortex), V2–V4, IT, PP, ventral prefrontal cortex v46, and dorsal prefrontal cortex d46). These six modules are reciprocally connected in a parallel (dorsal and ventral) hierarchy in accord with anatomical data (Felleman and Van Essen, 1991). Some details of the background for and architecture of the model (Deco and Lee, 2004; Deco and Rolls, 2004; Rolls and Deco, 2002) are as follows.

2.1.1. Details of the architecture

Information from the retino-geniculo-striate pathway enters the visual cortex through area V1 in the occipital lobe and proceeds into two processing streams. The occipital-temporal stream leads ventrally through modules V2, V4 to IT (the inferior temporal cortex), and is mainly concerned with object recognition, independently of position and scaling. The occipito-parietal stream leads dorsally into PP (the posterior parietal complex) and is responsible for maintaining a spatial map of an object's location and/or the spatial relationship of an object's parts as well as for moving the spatial location of attention. In the model, the ventral stream consists of the four modules V1, V2, V4 and IT. This part of the architecture is similar to VisNet in architecture and training (Wallis and Rolls, 1997; Rolls and Milward, 2000; Elliffe et al., 2002; Rolls and Deco, 2002), except that backprojections are incorporated, and the numbers of neurons are reduced for simplicity. These different modules allow combinations of features or inputs that occur in a given spatial arrangement to be learned by neurons, ensuring that higher-order spatial properties of the input stimuli are represented in the network (Elliffe et al., 2002). This is implemented via convergent connections to each part of a layer from a small region of the preceding layer, thus allowing the receptive field size of cells to increase through the ventral visual processing areas, as is observed in the primate ventral visual stream (see Fig. 1b). An external top-down bias, coming it is postulated from a short-term memory for shape features or objects in the more ventral part of the prefrontal cortex area v46, generates an object-based attentional component that is fed back down through the recurrent connections from IT through V4 and V2 to V1 (see Fig. 1b). The V1 module contains hypercolumns, each covering a pixel in a topologically organized model of the scene. Each hypercolumn contains orientation columns of orientation-tuned (complex) cells with

Gabor filter tuning at octave intervals to different spatial frequencies. V1 sends visual inputs to both the ventral and dorsal streams, and in turn receives backprojections from each stream, providing a high-resolution representation for the two streams to interact. This interaction between the two streams made possible by the backprojections to V1 is important in the model for implementing attentional effects. In the brain, there may be contributions to this interaction from further cross-links between the processing streams, occurring for example in V2, but the principle of the interaction is captured in the model by the common V1 module. The V2, V4 and IT modules each receive inputs from a small region of the preceding module, allowing the receptive field sizes of the neurons to increase gradually through the pyramidal structure of the network. Each of these modules acts like a competitive network (see Rolls and Treves, 1998, Rolls and Deco, 2002 and Wallis and Rolls, 1997) which enables neurons to learn to respond to spatially organized combinations of features detected at the preceding stage, thus helping to solve the binding problem (Elliffe et al., 2002), and also implementing a certain degree of localized competitive interaction between different targets. All the feedforward connections are trained by an associative (Hebb-like) learning rule with a short-term memory (the trace learning rule) in a learning phase in order to produce invariant neuronal responses (Rolls, 1992; Wallis and Rolls, 1997; Rolls and Milward, 2000). The backprojections between modules, a feature of cortical connectivity (Rolls and Treves, 1998; Rolls and Deco, 2002) are symmetric and reciprocal in their connectivity with the forward connections. The average strength of the backprojections is set to be a specified fraction of the strength of the forward connections (by a single parameter in the model) so that the backprojections can influence but not dominate activity in the input layers of the hierarchy (Renart et al., 1999a, b). Intramodular local competition is implemented in all modules by lateral local inhibitory connections between a neuron and its neighboring neurons via a Gaussian-like weighting factor as a function of distance (see Deco and Rolls, 2004).

The inputs to module V1 of the network are provided by neurons with simple cell-like receptive fields. This input filtering enables real images to be presented to the network. Following Daugman (1988) the receptive fields of these input neurons are modelled by 2D-Gabor functions. The Gabor receptive fields have five degrees of freedom given essentially by the product of an elliptical Gaussian and a complex plane wave. The first two degrees of freedom are the 2D-locations of the receptive field's centre; the third is the size of the receptive field; the fourth is the orientation of the boundaries separating excitatory and inhibitory regions; and the fifth is the symmetry. This fifth degree of freedom is given in the standard Gabor transform by the real and imaginary part, i.e. by the phase of the complex function representing it, whereas in a biological context this can be done by combining pairs of neurons with even and odd receptive fields. This design is supported by the experimental work of Pollen and Ronner (1981), who found simple cells in quadrature-phase pairs. Even more, Daugman (1988) proposed that an ensemble of simple cells is best modelled as a family of 2D-Gabor wavelets sampling the frequency domain in a log-polar manner as a function of eccentricity. Experimental neurophysiological evidence constrains the relation between the free parameters that define a 2D-Gabor receptive field (De Valois and De Valois, 1988). There are three constraints fixing the relation between the width, height, orientation, and spatial frequency (Lee, 1996). The first constraint posits that the aspect ratio of the elliptical Gaussian envelope is 2:1. The second constraint postulates that the plane wave tends to have its propagating direction along the short axis of the elliptical Gaussian. The third constraint assumes that the half-amplitude bandwidth of the frequency response is about 1–1.5 octaves along the optimal orientation. Further, we assume that the mean is zero in order to have an admissible wavelet basis (Lee, 1996). The neuronal pools in our V1 module cells are modelled by the power modulus of a 2D-Gabor function sensitive to a particular location, orientation, symmetry, and spatial frequency according to the constraints described above. The V1 module contains $N_{\mathrm{V1}} \times N_{\mathrm{V1}}$ hypercolumns, covering a $N \times N$ pixel scene. Each hypercolumn contains L orientation columns of complex cells with K octave levels corresponding to different spatial frequencies. The cortical magnification factor is explicitly modelled by introducing more high spatial resolution neurons in a hypercolumn the nearer this hypercolumn is to the fovea. The density of the fine spatial resolution neurons across the visual field decreases in the model according to a Gaussian function centred on the fovea. In other words, in the periphery far from the fovea only coarse spatial resolution V1 pools are in the respective hypercolumn, whereas in regions near to the fovea, the V1 hypercolumns include also high spatial resolution input neurons.

The modules V2, V4 and IT consist also of C-dimensional columns of neuronal pools (i.e., each column contains C pools) distributed in a topographical lattice with $N_{\mathrm{V2}} \times N_{\mathrm{V2}}$, $N_{\mathrm{V4}} \times N_{\mathrm{V4}}$, and $N_{\mathrm{IT}} \times N_{\mathrm{IT}}$ neurons, respectively. The connectivity between modules V1–V2, V2–V4 and V4–IT is intended to mimic the convergent forward connectivity of the cerebral cortex. This connectivity helps to implement the gradually increasing receptive field size as one proceeds up the cortical hierarchy, and the formation of neurons that respond to combinations of inputs with features in a defined spatial configuration (Rolls, 1992; Wallis and Rolls, 1997; Elliffe et al., 2002). The connections to neuronal pools in a column in an upper module are limited to neuronal pools in a column in the immediately connected lower module that are within a certain radius around the focal point of connection (see Fig. 1). This connectivity is reciprocated by the backprojections.

In this model system (Deco and Lee, 2004; Deco and Rolls, 2004; Rolls and Deco, 2002), one particular aspect of ventral stream function was modelled: translation invariant object recognition. One particular aspect of parietal cortex function was modelled: the encoding of visual space in retinotopic coordinates. (This is obviously a great simplification of the complex hierarchical architecture and functions of the primate visual system, but the model is sufficient to test and demonstrate our fundamental proposals for how the object and spatial streams interact to implement visual attentional processes. The way in which connectivity in the ventral visual stream could self-organize appropriately to implement translation invariance by using a modified associative learning rule with a short-term memory trace of preceding neuronal activity which statistically is likely to be a transform of the same object has been studied by Rolls and colleagues (Wallis and Rolls, 1997; Rolls and Milward, 2000; Elliffe et al., 2000; Rolls and Stringer, 2001; Elliffe et al., 2002; Stringer and Rolls, 2002; Rolls and Deco, 2002), and these ideas were incorporated into the dynamical model of ventral and dorsal visual stream operation in attention by Deco and Rolls (2004). Further details of the model, including the equations for the dynamics and the connectivity, are provided by Deco and Lee (2004) and Rolls and Deco (2002).) In the model, top-down connections from prefrontal cortex area 46 (modules d46 and v46) provide the external top-down bias that specifies the processing conditions of earlier modules. In particular, the feedback connections from area v46 with the IT module specify the target object in a visual search task; and the feedback connections from area d46 with the PP module generate the bias to a targeted spatial location in an object recognition task given a spatial attentional cue. These top-down inputs produce effects in the whole system by a biased competition process, which has been modelled only in an individual module such as IT previously (Usher and Niebur, 1996).[2] Each node or computational unit in the modules represents a population or pool of neurons. The activity of each computational unit is described by a dynamical equation derived from the mean-field approximation (see Appendix A and Rolls and Deco, 2002 for details). The mean-field approximation consists of replacing the temporally averaged discharge rate of a neuron with the instantaneous ensemble average of the activity of the neuronal population or pool (corresponding to the assumption of ergodicity). The dynamical evolution of activity at the level of a cortical area can be simulated in the framework of the present model by integrating the pool activity in a given area over space and time.

2.1.2. Operation of the model

The system operates in two different modes: the learning mode and the recognition mode. During the learning mode the synaptic connections between V4 and IT are trained by means of Hebbian (associative) learning during several presentations of a specific object at changing random positions in the visual field. This is the simple way in which translation invariant representations are produced in IT in this model. During the recognition mode there are two possibilities for running the system, illustrated in Fig. 2.

First, in *visual spatial search mode* (Fig. 2b), an object can be found in a scene by biasing the system with an external top-down (backprojection) component (from e.g. prefrontal area v46) to the IT module. This drives the competition in IT in favour of the pool associated with the specific object to be searched for. Then, the intermodular backprojection attentional modulation IT–V4–V1 will enhance the activity of the pools in V4 and V1 associated with the component features of the specific object to be searched for. This modulation will add to the visual input being received by V1, resulting in greater local activity where the features in the topologically organized visual input features match the backprojected features being facilitated. Finally, the enhanced firing in a particular part of V1 will lead to increased activity in the topologically mapped forward pathway from V1 to V2–V4 to PP, resulting in increased firing in the PP module in the location that corresponds to where the object being searched for is located. In this way, the architecture automatically finds the location of the object being searched for, and the location found is made explicit by which neurons in the spatially organized PP module are firing.

Second, in *visual object identification mode* (Fig. 2a), the PP module receives a top-down (backprojection) input (from e.g. prefrontal area d46) which specifies the location in which to identify an object. The spatially biased PP module then drives by its backprojections the competition in the V2–V4 module in favour of the pool associated with the specified location. This biasing effect in V1 and V2–V4 will bias these modules to have a greater response for the specified location in space. The shape feature representations which happen to be present due to the visual input from the retina at that location in the V1 and V2–V4 modules will therefore be enhanced, and the enhanced firing of these shape features will by the feedforward pathway V1–V4–IT favour the IT object pool that contains the facilitated features, leading to recognition in IT of the object at the attentional location being specified in the PP module. The operation of these two attentional modes is shown schematically in Fig. 2.

The operation of this system is illustrated in Fig. 3. (The whole system was described by Deco and Rolls (2004), and the simulation illustrated in Fig. 3 is a simplified simulation including only the IT–V1–PP modules performed by Deco and Lee (2004) just to show the main aspects of the dynamics.) Fig. 3I illustrates operation in the *spatial attention mode* when the object at that location was to be identified (cf. Fig. 2a; see figure legend for details). Fig. 3II shows the operation of the system in *object attention mode*, i.e. in the visual search task when the system was looking for the location of a particular object in a visual scene (cf. Fig. 2b; see figure legend for details, and Rolls and Deco (2002)). (The parameters for this simulation were obtained from the mean-field analysis, and the source of the fluctuations was noise introduced at the mean-field level, which can be interpreted as the effect of finite-size noise (Mattia and Del Giudice, 2002, 2004).)

2.2. Linking computational and single-neuron data

Is this theoretical framework (described in Section 2.1 and in detail by Rolls and Deco (2002)) relevant to understanding the visual system? By design, our model attempts to explain a potential functional role of the backprojection connections in the visual cortex. But do the computational units in this model

[2] The hypothesis for this mechanism can be traced back to the ‘adaptive resonance’ model (Grossberg, 1987) in the neural network literature and the ‘interactive activation’ model (McClelland and Rumelhart, 1981) in the connectionist literature.

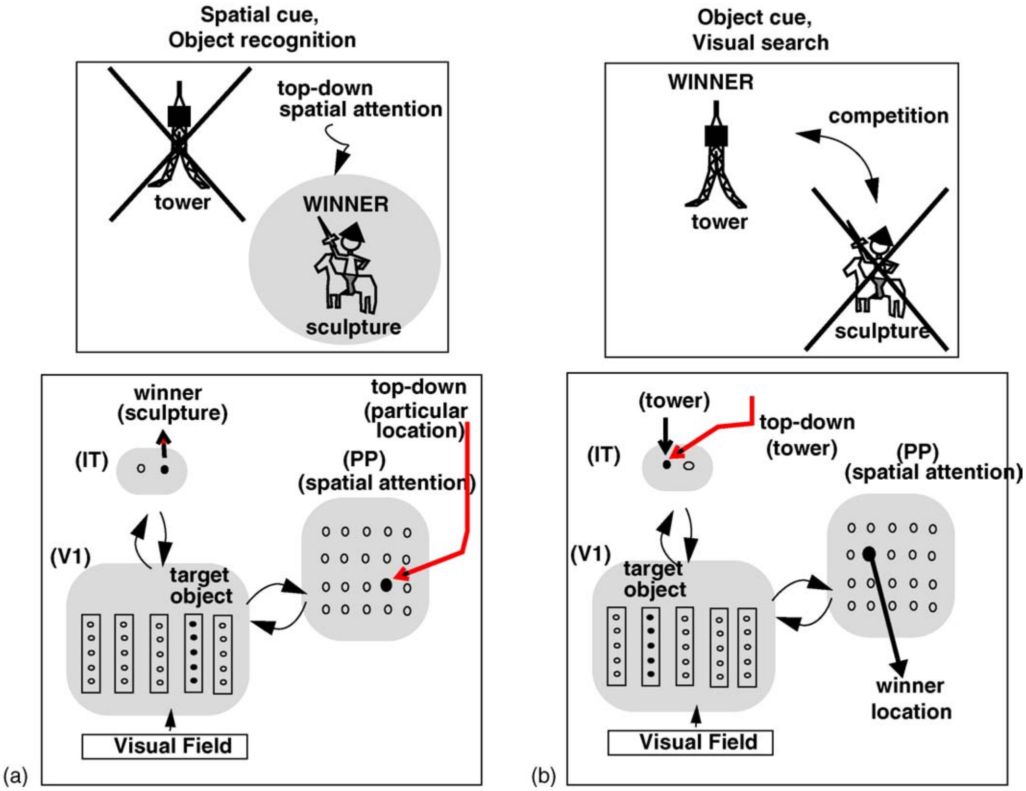


Fig. 2. Attentional modulation for finding an object (in visual object identification mode) at a specific spatial location using spatial bias to the PP module from for example a prefrontal module d46. See text for details. (b) Attentional modulation for finding a visual spatial location (in visual spatial search mode) when an object is specified by bias to the IT module from for example a prefrontal module v46.

system behave in a similar way to the neurons observed in neurophysiological experiments? We address this in this Section X. We then show that the same theoretical framework and model can also be directly related to functional neuroimaging data (Section 2.3), to psychophysical data on serial versus parallel processing (Section 2.4), and to neuropsychological data (Section 2.5). The framework and model operate in the same way to account for findings at all these levels of investigation, and in this sense provide a unifying framework.

In neurophysiological experiments, it is found that neurons in V4 and IT show competitive interactions between stimuli presented within the receptive field, and that the suppression of a neuronal response to a feature within the receptive field can be overcome if attention is paid to the location of the feature (Moran and Desimone, 1985; Chelazzi et al., 1993; Reynolds et al., 1999). Similar effects are found for single neuron responses in the model (as shown in Fig. 6.4 of Rolls and Deco, 2002). In another comparison, Deco and Lee (2004) and Corchs and Deco (2002) showed a long-latency enhancement effect on V1 units in the model under top-down attentional modulation (compare Fig. 3Ib with Fig. 3Ia) which is similar to the long-latency contextual modulation effects observed in early visual cortex (Lamme, 1995; Zipser et al., 1996; Lee et al., 1998; Roelfsema et al., 1998; Lee et al., 2002). Interestingly, in our simulation, we found that the observed spatial or object attentional enhancement is stronger for weaker stimuli. This predicted result has been confirmed neurophysiologically by Reynolds et al. (2000). The mechanism for this may be that the top-down attentional influence can dominate the firing of the neurons relatively more as there are fewer feedforward forcing and shunting effects on the neurons. An extension of this model (Deco and Rolls, 2004) can account for the reduced receptive fields of inferior temporal cortex neurons in natural scenes (Rolls et al., 2003), and makes predictions about how the receptive fields are affected by interactions of features within them, and by object-based attention.

The model has also been extended to the level of spiking neurons which allows biophysical properties of the ion channels affected by synapses, and of the membrane dynamics, to be incorporated, and shows how the non-linear interactions between bottom-up effects (produced for example by altering stimulus contrast) and top-down attentional effects can account for new neurophysiological results in areas MT and V4 (Deco and Rolls, 2005a). The model and simulations show that attention has its major modulatory effect at intermediate levels of bottom-up input, and that the effect of attention disappears at low and high levels of contrast of the competing stimulus. The model assumed no kind of multiplicative attentional effects on the gain of neuronal responses. Instead, in the model, both top-down attention and bottom-up input information (contrast) are implemented in the same way, via additive synaptic effects in the postsynaptic neurons. There is of course a non-linearity in the effective activation function of the integrate-and-fire neurons, and this is what we identify as the source of the apparently multiplicative effects (Martinez-Trujillo and Treue,

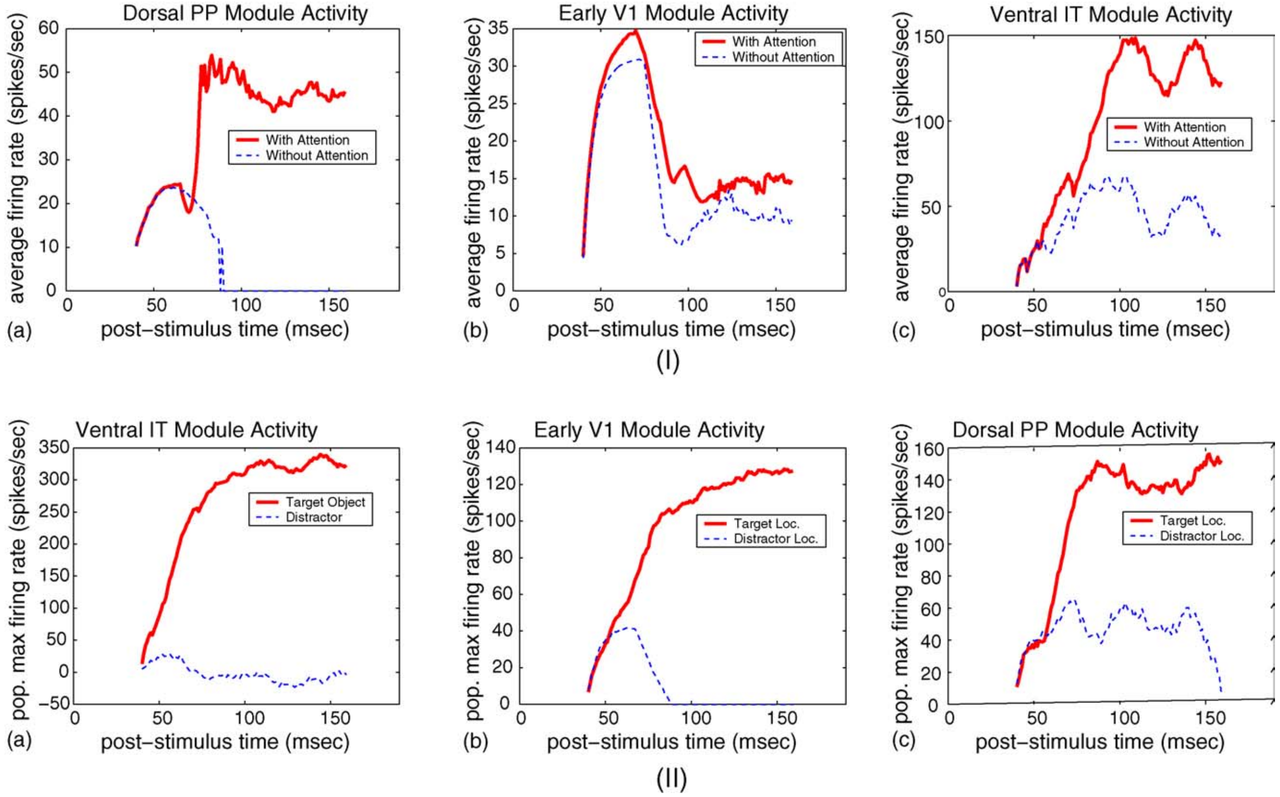


Fig. 3. I. Effect of spatial attention on neural activities shown in three modules when performing object recognition. A top-down spatial attentional bias was introduced to the dorsal PP module pool that encodes the location of the sculpture in the scene (curves marked ‘With Attention’), or there was no attentional input (curves marked ‘Without Attention’). The responses of neuronal pools in the corresponding locations in the dorsal stream PP module and the V1 module, and the response of the neuronal pool encoding the sculpture in the ventral stream IT module were compared with and without spatial attention. (a) The evolution of the neuronal pool response at the dorsal stream PP module with and without the top-down spatial attentional bias. When the top-down bias was absent, the competition in the dorsal stream PP module was entirely determined by the bottom-up signals from V1. In this case, the sculpture shape did not provide the strongest input and was rapidly suppressed by the competitive interactions from the neuronal activities in the other locations in the scene. When there was a top-down spatial bias to the PP neuronal pool, the neuronal pool’s activity rose rapidly and was maintained at a sustained level. (b) The increase in the activity of the neuronal pool in the dorsal stream PP module enhanced the activity at the corresponding retinotopic location in the early V1 module, particularly in the latter phase of the response, producing a highlighting effect. (c) The response of the ventral stream IT module neuron pool coding for the sculpture was substantially stronger when the spatial attention bias to the PP module was allocated to the location of the sculpture than when there was no spatial attention. The spatial highlighting gated the sculpture image to be analyzed and recognized by the ventral stream IT module neurons. Without spatial attention, the sculpture did not have bottom-up saliency to enable the domination of the sculpture neuron in the ventral stream IT module. II. Neuronal activities shown in three modules during visual search in object attention mode. The maximum population activity of the neuronal pools corresponding to the identity or location of the sculpture in the scene in the three modules was compared against the maximum activity in pools coding any other locations or objects when sculpture was the target object of attention. The effect of attention, i.e. differentiation of neuronal response between the target and distractor conditions, was observed to start at the ventral stream IT module, and then to propagate to the early V1 module and the dorsal stream PP module. (a) The neuronal activity of the top-down biased ‘sculpture’ neuronal pool in the ventral-stream module was compared against the maximum activity of all other object pools. The increase in the response of the sculpture pool relative to the response in all other pools was observed to rise rapidly as a result of the top-down object attentional bias applied to the IT module. (b) The maximum population activity at the sculpture location in the early V1 visual module was compared against the maximum activity of pools at all other locations in the scene. (c) The maximum activity of the pools coding in the dorsal stream PP module for the sculpture location was compared against the maximum activity of the pools encoding the other locations.

2002) of top-down attentional biases on bottom-up inputs. The relevant part of the effective activation function of the neurons (the relation between the firing and the injected excitatory currents) is the threshold non-linearity, and the first steeply rising part of the activation function, where just above threshold the firing increases markedly with small increases in synaptic inputs (cf. Amit and Brunel (1997) and Brunel and Wang (2001)). Attention could therefore alternatively be interpreted as a phenomenon that results from purely additive synaptic effects, non-linear effects in the neurons, and cooperation-competition dynamics in the network, which together yield a variety of modulatory effects, including effects that appear (Martinez-Trujillo and Treue, 2002) to be multiplicative. In addition, we were able to show that the non-linearity of the NMDA receptors may facilitate non-linear attentional effects, but is not necessary for them. This was shown by disabling the voltage-dependent non-linearity of the NMDA receptors in the simulations (Deco and Rolls, 2005a).

2.3. Linking computational and functional neuroimaging data

This type of computational neuroscience approach is also able to simulate and predict functional neuroimaging data obtained with for example functional magnetic resonance imaging (fMRI), as shown in Appendix B. For example, Corchs and Deco (2002) simulated fMRI signals from V4 in the attentional paradigm studied by Kastner et al. (1999), and reproduced the experimental result that when multiple stimuli are present simultaneously in the visual field, their cortical representations within the object recognition pathway interact in a competitive, suppressive fashion. In addition, directing attention to one of the stimuli counteracted the suppressive influence of nearby stimuli. In these simulations, the activity of neuronal populations or pools in a given area in the model described above was integrated over space and time, to make it relevant to interpreting fMRI signals. Simulations of this type can help to interpret fMRI data, and indeed Deco et al. (2004) showed that differences in the fMRI BOLD-signal from the dorsal as compared to the ventral prefrontal cortex in working memory tasks may reflect a higher level of inhibition in the dorsolateral prefrontal cortex, rather than necessarily spatial working memory functions dorsally, and object working memory functions ventrally.

2.4. Linking computational and psychophysical data: ‘serial’ versus ‘parallel’ processing

In the visual search tasks we consider, subjects examine a display containing randomly positioned items in order to detect an a priori defined target (i.e. a target that the subject is paying attention to and must search for), and other items in the display which are different from the target serve the role of distractors. In a feature search task the target differs from the distractors in one single feature, e.g. only colour. In a conjunction search task the target is defined by a conjunction of features, and each distractor shares at least one of those features with the target. Conjunction search experiments show that search time increases linearly with the number of items in the display, which has been taken to imply a serial process, such as an attentional spotlight moving from item to item in the display (Treisman, 1988). An example of a display with this ‘serial’ search is shown in Fig. 4b, in which an E is the target and Fs are the distractors. On the other hand, search times in a feature search can be independent of the number of items in the display, and this is described as a preattentive parallel search (Treisman, 1988) (see further (Deco and Zihl, 2001; Rolls and Deco, 2002; Deco and Lee, 2004)). An example of a display with this ‘parallel’ search is shown in Fig. 4a, in which an E is the target and Xs are the distractors.

In a simulation of these feature and conjunction search tasks just described with the architecture of the type described in Section 2.1, the results shown in Fig. 4c and d were obtained (Deco and Zihl, 2001) (see also Rolls and Deco, 2002). The results of the simulations are consistent with existing experimental results (Quinlan and Humphreys, 1987). Although the

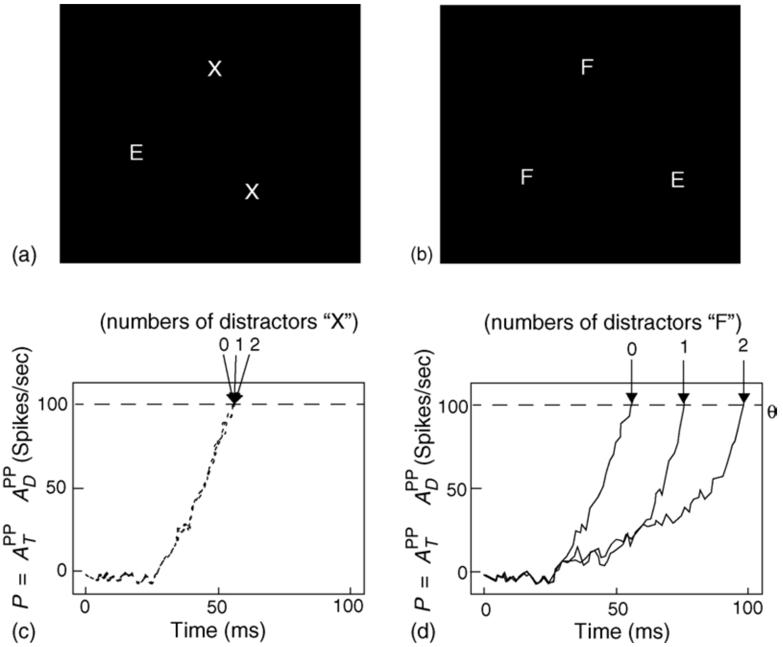


Fig. 4. (a) Parallel search example: an image that contains a target E in a field of X distractors. Since the elementary features in E and X are distinct, i.e. their component lines have different orientations, E pops out from X, and its location can be rapidly localized independently of the number of distracting X shapes in the image. This is called parallel search. (b) Serial search example: an image that contains a target E in a field of F distractors. Since both letters are composed of vertical and horizontal lines, there is no difference in the elementary features to produce a preattentive pop-out, so the E and F can be distinguished from each other only after their features are bound or glued by attention. The time required to locate the target in such an image increases linearly with the number of distractors in the image. This is called serial search and has been thought to involve the scanning of the scene with a covert attentional spotlight. (c)–(d) Simulation result of the network performing visual search on images (a) and (b) respectively. The difference (polarization) between the maximum activity in the neuronal pool corresponding to the target locations and the maximum activity of all other neuronal pools in the dorsal PP module is plotted as a function of time. (c) shows that the difference signal rose to a threshold, corresponding to localization of the object, at about the same time independently of the numbers of distractor items. (d) shows that the time for the polarization signal to rise to a threshold increased linearly with the number of distracting items, with an additional 25 ms required per item.

whole simulation is parallel, and involves no serial moving spotlight process, the conjunction search takes longer with more distractors because the constraints are then more difficult to satisfy, and the dynamics of the coupled set of networks takes longer to settle. (The constraints are more difficult to satisfy in the conjunction search task in that the features, which include oriented line elements, are more similar to each other in the E versus F case, and may use feature conjunctions to solve the fact that the F uses a subset of the feature combinations present in the E, see (Elliffe et al., 2002).) This is an important result, for it provides direct computational evidence that some apparently serial cognitive tasks may in fact be performed by fully parallel processing neuronal networks with realistic dynamics. Deco et al. (2002) extended this approach to deal with serial search when there are conjunctions of different numbers of features. The model predicted different slopes for different kinds of conjunction search, and these predictions were confirmed by psychophysical investigations (Deco et al., 2002).

The results of this neurodynamical model thus extend the ideas of Duncan and Humphreys (1989) who used the biased competition mechanism to conceptually explain serial visual search as involving a top-down process influencing the competition between neurons that code for different object attributes in the early visual areas, thus leading to the selection of one object in the visual field. The basic hypothesis for this mechanism can be traced back to earlier neural network models such as the adaptive resonance model (Grossberg, 1987) or the interactive activation model (McClelland and Rumelhart, 1981). The idea that some relatively early visual areas, such as MT, can potentially mediate the interaction of dorsal and ventral streams in visual search can be traced back to a workshop paper by Buracas et al. (1996).

2.5. *Linking computational and neuropsychological data on attention*

By artificially damaging the connections in the model, it is possible to account for a number of attentional deficits occurring in humans after brain damage. For example, with graded damage increasing in severity towards the right of the topologically organized parietal module, the model reproduces some of the symptoms of the left hemineglect of space that occurs after right parietal cortex damage in humans (Rolls and Deco, 2002; Heinke et al., 2002). If this lesion is combined with local lateral inhibition (rather than the global inhibition) within a processing module, then it is even possible to account for object-based neglect, in which the left half only of object in a series of objects arranged across the visual field is not seen (Deco and Rolls, 2002). Moreover, the model makes predictions about how patients with object-based neglect will perceive objects when they are brought towards each other, or are joined by cross-links (Deco and Rolls, 2002), and these predictions are useful, in part because they allow the model to be tested, as described by Rolls and Deco (2002) on pages 394–398, and by Heinke et al. (2003).

3. A unified model of attention and working memory

3.1. *A model of the effects of attention on working memory and decision making*

There is much evidence that the prefrontal cortex is involved in at least some types of working memory and related processes such as planning (Fuster, 2000). Working memory refers to an active system for maintaining and manipulating information in mind, held during a short period, usually of seconds. Recently, Asaad et al. (2000) investigated the functions of the prefrontal cortex in working memory by analyzing neuronal activity when the monkey performs two different working memory tasks using the same stimuli and responses. In a *conditional object–response (associative) task* with a delay, the monkey was shown one of two stimulus objects (O1 or O2), and after a delay had to make either a rightward or leftward oculomotor saccade response depending on which stimulus was shown. In a *delayed spatial response task* the same stimuli were used, but the rule required was different, namely to respond after the delay towards the left or right location where the stimulus object had been shown (Asaad et al., 2000). The main motivation for such studies was the fact that for real-world behaviour, the mapping between a stimulus and a response is typically more complicated than a one-to-one mapping. The same stimulus can lead to different behaviours depending on the situation, or the same behaviour may be elicited by different cueing stimuli. In the performance of these tasks populations of neurons were found that respond in the delay period to the stimulus object or its position (‘sensory pools’), to combinations of the response and the stimulus object or position (‘intermediate pools’), and to the response required (left or right) (‘premotor pools’). Moreover, the particular intermediate pool neurons that were active depended on the task, with neurons that responded to combinations of the stimulus object and response active when the mapping was from object to behavioural response, and neurons that responded to combinations of stimulus position and response when the mapping rule was from stimulus position to response (Asaad et al., 2000). In that different sets of intermediate population neurons are responsive depending on the task to be performed, PFC neurons provide a neural substrate for responding appropriately on the basis of an abstract rule or context.

Neurodynamics helps to understand the underlying mechanisms that implement this rule-dependent mapping from the stimulus object or the stimulus position to a delayed behavioural response. Deco and Rolls (2003) formulated a neurodynamical model that builds on the integrate-and-fire attractor network by introducing a hierarchically organized set of different attractor network pools in the lateral prefrontal cortex. The hierarchical structure is organized within the general framework of the biased competition model of attention (Chelazzi, 1998; Rolls and Deco, 2002). There are different populations or pools of neurons in the prefrontal cortical network, as shown in Fig. 5. (In order to be able to analyze and constrain the problem, we adopt a minimalistic approach and assume a minimal number of neuronal pools for coding two

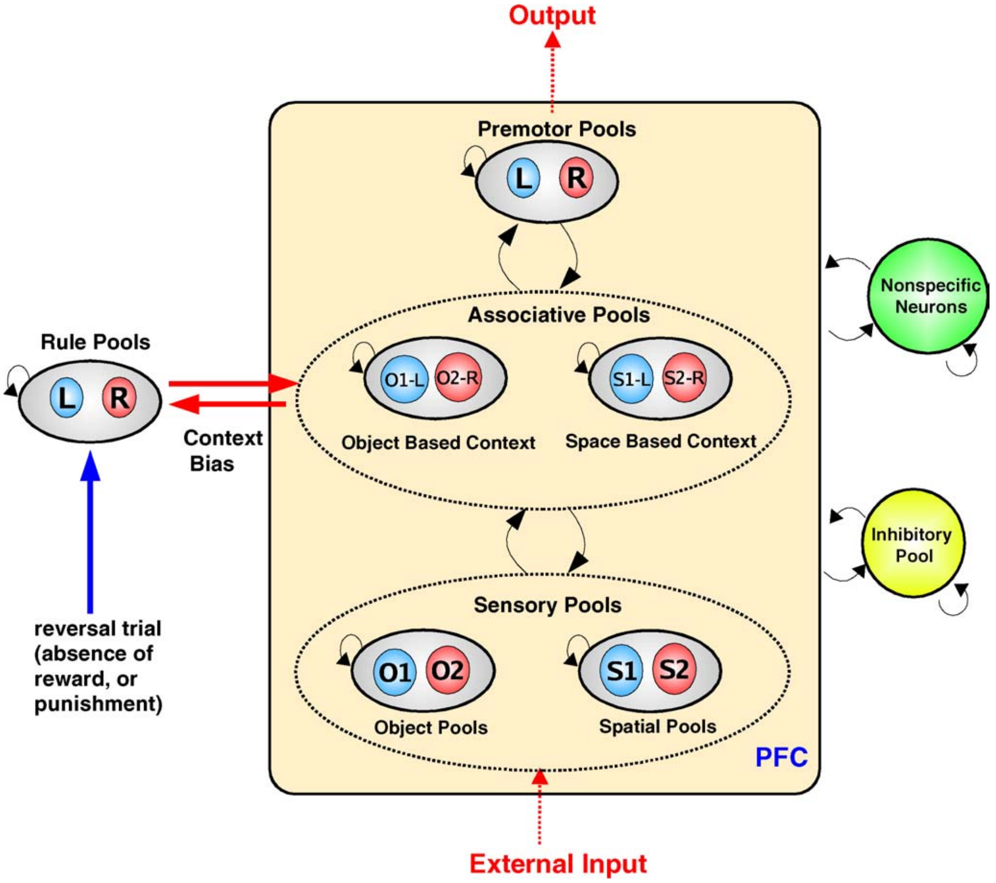
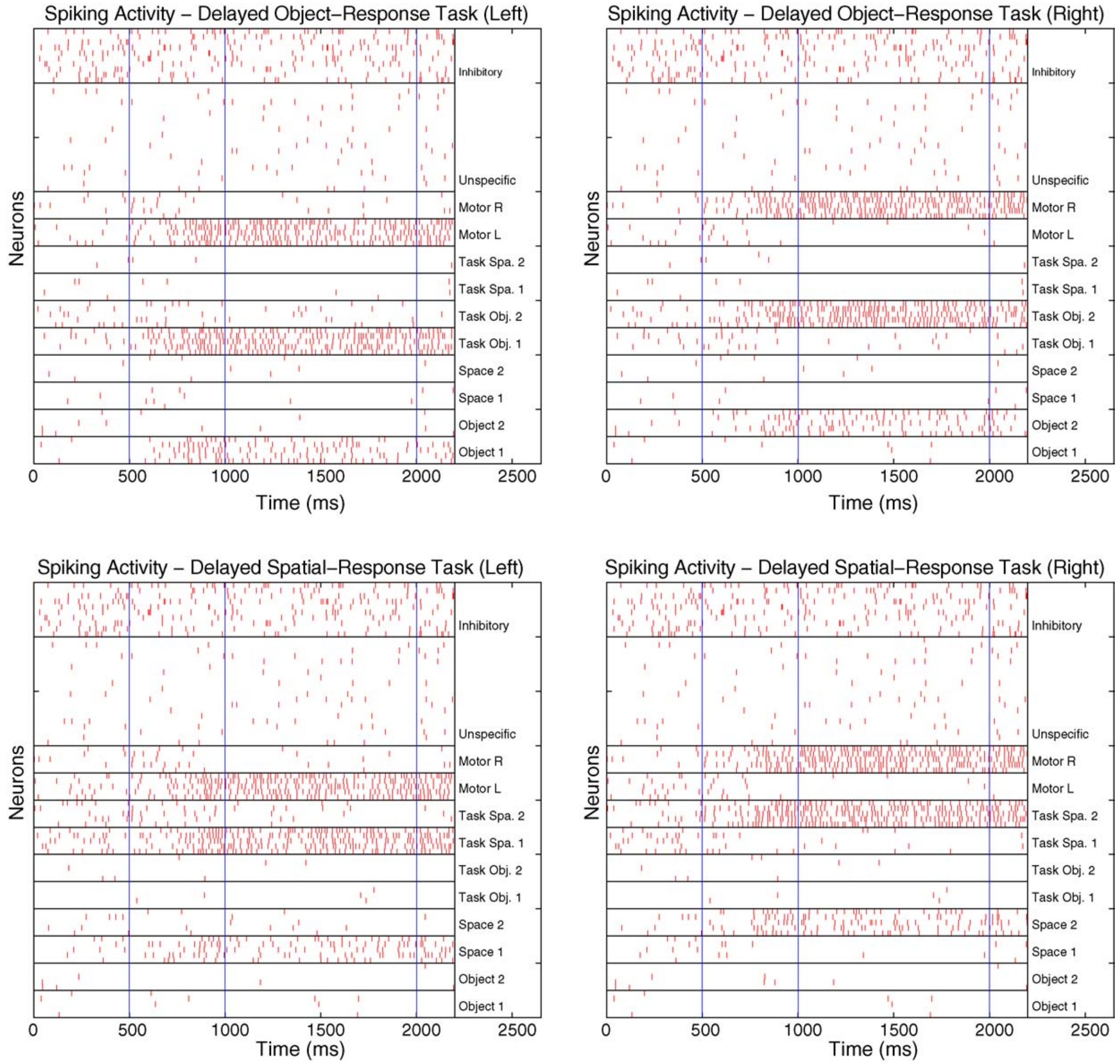


Fig. 5. Network architecture of the prefrontal cortex unified model of attention, working memory, and decision making. There are sensory neuronal populations or pools for object type (O1 or O2) and spatial position (S1 or S2). These connect hierarchically (with stronger forward than backward connections) to the intermediate or 'associative' pools in which neurons may respond to combinations of the inputs received from the sensory pools for some types of mapping such as reversal, as described by Deco and Rolls (2003). For the simulation of the data of Asaad et al. (2000) these intermediate pools respond to O1-L, O2-R, S1-L, or S2-R. These intermediate pools receive an attentional bias, which in the case of this particular simulation biases either the O pools or the S pools. The intermediate pools are connected hierarchically to the premotor pools, which in this case code for a Left or Right response. Each of the pools is an attractor network in which there are stronger associatively modified synaptic weights between the neurons that represent the same state (e.g. object type for a sensory pool, or response for a premotor pool) than been neurons in the other pools or populations. However, all the neurons in the network are associatively connected by at least weak synaptic weights. The attractor properties, the competition implemented by the inhibitory interneurons, and the biasing inputs, result in the same network implementing both short term memory and biased competition, and the stronger feed forward than feedback connections between the sensory, intermediate, and premotor pools results in the hierarchical property by which sensory inputs can be mapped to motor outputs in a way that depends on the biassing contextual or rule input.

different objects and two positions. This simplification allows a detailed study of the dynamics, and is enough to capture the main aspects of the dynamics associated with the memory functions that we aim to study here.) There are four types of excitatory pool, namely: sensory, intermediate (or associative, in that they respond to combinations), premotor, and non-selective. The sensory pools encode information about objects (O1 and O2), or spatial location (S1 and S2). The premotor pools encode the motor response (in our case the leftward (L) or rightward (R) oculomotor saccade). The intermediate (associative) pools are task-specific or rule-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, with one pool for each of the four possible stimulus–response combinations (O1-L, O2-R, S1-L and S2-R). The intermediate pools can be considered as being in two groups, one for the delayed object–response associative task, and the other for the delayed spatial response task. The intermediate pools receive an external biasing context or rule input (see Fig. 5) that reflects the current rule. The remaining excitatory neurons do not have specific sensory, response or biasing inputs, and are in a non-selective pool. All the inhibitory neurons are clustered into a common inhibitory pool, so that there is global competition throughout the network.

The parameters of the model (Deco and Rolls, 2003) were set using a mean-field analysis as follows, as this analysis enables the dependencies of specific network behaviours on the network parameters to be assessed. In particular, we showed that the inter-pool connection strengths along the processing pathway cannot be equal, and an asymmetry is needed so that a response is computed by the network. The mean-field analysis shows the states that the system can reach, and enables further constraints to be identified to produce behaviour of the system



Spiking Activity – Delayed Object–Response Task (Left)

Spiking Activity – Delayed Object–Response Task (Right)

Spiking Activity – Delayed Spatial–Response Task (Left)

Spiking Activity – Delayed Spatial–Response Task (Right)

Fig. 6. Delayed spatial–response and delayed object–response simulations. Rastergrams of randomly selected neurons for each pool in the PFC network (5 for each sensory, intermediate and premotor pool, 20 for the non-selective excitatory pool, and 10 for the inhibitory pool) and for all task conditions after the experimental paradigms of Asaad et al. (2000). The spatio-temporal spiking activity shows that during the short-term memory delay period only the sensory cue, associated future oculomotor response, and intermediate neurons maintain persistent activity and build up the stable global attractor of the network. The underlying biased competition mechanisms are explicit. Pool names: Task Spa 2-intermediate pool responding in the delayed spatial response task to a combination of spatial location 2 and a Right motor response; Task Obj 1-intermediate pool responding in the delayed object–response task to a combination of object 1 and a left motor response; Space 1-sensory pool responding to the stimulus when it is in spatial position 1; Object 1-sensory pool responding to the stimulus when it is object 1 (after Deco and Rolls, 2003).

that correlates best to the corresponding data, and also better theoretical descriptions for the dependencies of specific behaviours on the network parameters (see Loh et al., 2004).

Fig. 6 plots the rastergrams of randomly selected neurons for each pool in the network. The spatio-temporal spiking activity shows that during the short-term memory delay period only the relevant sensory cue, associated future oculomotor response, and intermediate neurons maintain persistent activity, and build up a stable global attractor in the network. The underlying biased competition mechanisms that operate as a result of the rule/context input (see Fig. 5) biasing the correct intermediate pool neurons are very explicit in this experiment. Note that neurons in pools for the irrelevant input sensory dimension (location for the object–response associative task, and object for the delayed spatial response task), are inhibited during the sensory cue period and are not sustained during the delay short-term memory period. Only the relevant single pool attractors, given the rule context, that are suitable for the cue–response mapping survive the competition and are persistently maintained with high firing activity during the short-term memory delay period.

This model thus shows how a rule or context input can influence decision making by biasing competition in a hierarchical network that thus implements a flexible mapping from input stimuli to motor outputs (Rolls and Deco, 2002; Deco and Rolls, 2003). The model also shows how the same network can implement a short-term memory, and indeed how the competition required for the biased competition selection process can be implemented in an attractor network which itself requires inhibition implemented through the inhibitory neurons. The short-term memory of the system, evident by the continuing firing of the relevant neuronal sensory, intermediate combination, and motor pools in the delay period from 1000 to 2000 ms in Fig. 6, is implemented both by the recurrent associatively modified connections within each pool of neurons shown in Fig. 5, but also by the associatively modified feedforward and feedback connections between the pools at the different levels of the hierarchy shown in Fig. 5. It is an interesting property of the model that the system consists of sets of hierarchically but reciprocally connected attractor networks which together contribute to the short-term memory implemented in the system. We note that the types of neuron included in the model are the types recorded by Asaad et al. (2000) during the performance of the task, and that a combinatorial explosion of required neuron types may be avoided by the fact that neurons in the prefrontal cortex adapt their response properties to the task demands (Freedman et al., 2003).

Another issue is how the connectivity between the different pools is set up during the learning phase. One process, which works for many of the connections, is Hebbian associative learning. However, even in the cases when the forward and backward connections are not identical (as in our model), the pattern of the synaptic weights required can still be set up during training by simple associative learning. All that is required even in these cases is some asymmetry in the gain of the forward and backward connections. This could be implemented by the forward and backward connections terminating on different parts of the dendrite, as occurs for connections between cortical areas (see Rolls and Treves, 1998 and Rolls and Deco, 2002). The implication of this would be that there would be a trend through prefrontal cortical areas, from those closer to the sensory input, through areas between the sensory and motor-related areas, to areas with response-related neuronal activity. With the connections through the networks in this direction, there would be stronger connections in the correct (forward) direction, because the forward connections are more likely to end on the main part of the dendrites of pyramidal cells, and the backprojections are more likely to end on the apical dendrites of cortical pyramidal cells (Rolls and Treves, 1998; Rolls and Deco, 2002). Such a trend, from prefrontal cortical areas that receive from posterior perceptual areas, through regions that are intermediate, through to regions closer to motor output, could in fact be one of the principles of prefrontal cortical connectivity, which would not be inconsistent with what is known about prefrontal connectivity. For example, the orbitofrontal cortex has mainly sensory inputs (with little response-related neuronal activity) (Rolls, 1999, 2005), and so does the ventrolateral prefrontal cortex (area 47/12). The dorsolateral prefrontal cortex is more of a mixed area, with neurons that respond to combinations of sensory inputs and responses, and where effects of biassing attentional signals are evident (Asaad et al., 2000, 1998). Finally the more dorsal and posterior prefrontal cortical areas may be more closely related to the responses being made, including oculomotor responses (Kandel et al., 2000).

Synaptic plasticity is dependent on the timing of the spikes in the pre- and postsynaptic neuron (Markram et al., 1997; Bi and Poo, 2001; Senn et al., 2001), and a theoretical and computational analysis of these effects in the context of working memory formation in the prefrontal cortex has been performed by Fusi and colleagues (Fusi, 2003, 2002; Fusi et al., 2000; Fusi and Mattia, 1999). They have shown how Hebbian dynamic learning can cope with both stability of the network states and stability of the learning process. They have shown that a spike-time based learning rule can result in a rate dependent long-term synaptic modification, and that a working memory prefrontal architecture similar to ours (i.e. excitatory pools of neurons strongly connected within a pool, and weakly connected between other excitatory pools, and with a common inhibitory pool) can indeed be formed by this kind of spike-time based learning. Further, there is accumulating evidence (Sjöström et al., 2001) that a more realistic description of the protocols for inducing LTP and LTD probably requires a combination of dependence on spike-timing – to take into account the effects of the backpropagating action potential – and dependence on the sub-threshold depolarization of the postsynaptic neuron. However these spike-timing-dependent synaptic modifications may be evident primarily at low firing rates (Sjöström et al., 2001), and may not be especially reproducible in the cerebral neocortex.

3.2. The topographical structure of the prefrontal cortex

A 'what'/'where' hypothesis proposes that visual working memory is organized into two networks within the PFC, with spatial working memory supported by the dorsolateral PFC, and object working memory supported by the ventrolateral PFC of the lateral convexity (Leung et al., 2002; Goldman-Rakic, 1987). A second hypothesis proposes a hierarchical organization of the PFC by which non-mnemonic higher-order functions (e.g. manipulation) are ascribed to dorsolateral prefrontal areas, and short-term memory maintenance functions are allocated to inferior prefrontal areas (Postle and D'Esposito, 2000; Rao et al., 1997).

Deco et al. (2004) modelled the underlying mechanisms that implement the working memory-related activity observed in the primate PFC as evidenced by single-cell data and in the human PFC as evidenced with event-related fMRI during the execution of delay tasks with a 'what'-then-'where' design. Their main finding was that the model could account for the neurophysiological activity seen in both the ventrolateral and dorsolateral PFC during the delay periods of Working Memory tasks (as found in the neurophysiological results of Rao et al. (1997)), and at the same time could provide simulated fMRI signals that matched experimental findings during a 'what-then-where' short-term memory task for both PFC sectors (as shown by the fMRI findings of Postle and D'Esposito (2000)). However, the fMRI data were most easily modelled by hypothesizing that the differences between these two prefrontal regions resulted from assigning a greater amount of inhibition to the dorsolateral portion of the prefrontal cortex. Both brain areas may show short-term memory maintenance capabilities related to their capacities to maintain stable attractors during delay periods, but the increased level of inhibition assumed in the dorsolateral PFC may be associated with the capacity of this brain region to support more complex functions. Exactly what those more complex functions may be is not revealed by the studies of Deco et al. (2004), but higher inhibition in the dorsolateral prefrontal cortex might be useful for maintaining several separate representations and preventing the formation of a global attractor, which could be useful if several items must be held in memory for manipulation. The same model made the prediction that drugs acting at D1 receptors would impair both short-term memory and attention, for by reducing the maximum current that can flow through NMDA receptor activated channels, the attractors that represent different short-term memory states become less distinct.

4. Closing the cognitive loop: reward reversal and decision making

The prefrontal architecture shown in Fig. 5 uses a context or rule bias to influence the mapping from stimulus to response, and thus to influence decision making. The question we now address is how the change in the rewards being received when the task is changed could implement a switch from one rule to another. Deco and Rolls (2005c) proposed that the rule is held in a separate attractor network as shown in Fig. 5 which is the source of the biased competition. Different sets of neurons in this network are active according to which rule is current. Consistent with this Wallis et al. (2001) have described neurons in the primate PFC that reflect the explicit coding of abstract rules. We propose that the synapses in the recurrent attractor rule network show some adaptation, and that when an error signal is decoded by neurons in the orbitofrontal cortex (a part of the prefrontal cortex) that respond when an expected reward is not obtained (Thorpe et al., 1983; Rolls, 1999, 2000b, 2004, 2005), activation of the inhibitory interneurons reduces activity in the attractor networks, including that which holds the current rule. This reduction of firing is sufficient in the simulations to stop the dynamical interactions between the neurons that maintain the current attractor. When the attractor network starts up again after the inhibition, it starts with the other attractor state active, as the synapses of this other attractor have not adapted as a result of previous activity (Deco and Rolls, 2005c).

In order to model reward reversal in a Go-NoGo paradigm, we used the architecture illustrated in Fig. 7, with the different neuron pools as follows (Deco and Rolls, 2005c). In a 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: 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: triangle-rewarding, square-punishing, and two for the reversal condition: triangle-punishing, square-rewarding.) These intermediate pools respond to combinations of the sensory stimuli and the expected reward, e.g. to triangle and an expected reward (glucose obtained after licking). The sensory–intermediate–reward module consists of three hierarchically organized levels of attractor network, with stronger synaptic connections in the forward than the backprojection direction. The rule module acts as a biassing input to bias the competition between the object–reward combination neurons at the intermediate level of the sensory–intermediate–reward module. It is an important part of the architecture that at the intermediate level of the sensory–intermediate–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: the triangle is rewarding (in that glucose taste reward is obtained if a lick is made) and the square is punishing (associated with aversive saline taste so that licking should be avoided); and the other pool represents the reverse associations hold currently.

The cue–response mapping required under a specific context is achieved via the biassing 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 (which was previously 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

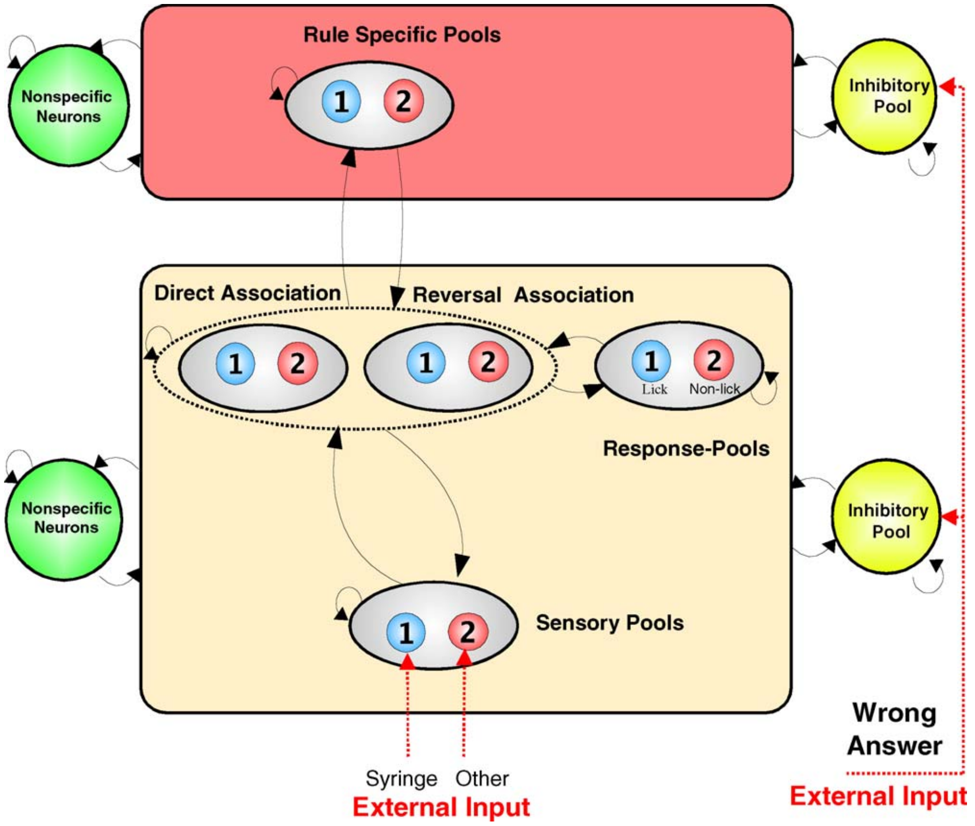


Fig. 7. Cortical architecture or the reward reversal model. There is a rule module (top) and a sensory–intermediate neuron–reward module (below). Neurons within each module are fully connected, and form attractor states. The sensory–intermediate neuron–reward module consists of three hierarchically organized 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. object1–reward (O1R, in the direct association set of pools), and object1–punishment (O1P in the reversed association set of pools). The rule-module acts as a biassing 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_{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 Deco and Rolls, 2005c for details).

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 (Rolls, 2005). 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. 7). (The increase was from ν_{ext} to $\nu_{\text{ext}} + \lambda_{\text{Punish}}$ with $\lambda_{\text{Punish}} = 500$ Hz.) (The increase was from ν_{ext} to $\nu_{\text{ext}} + \lambda_{\text{Punish}}$ 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 Fig. 7, 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 ν_{ext} to $\nu_{\text{ext}} + B$ with $B = 200$ Hz).

We describe now the specific implementation of the spike-frequency-adaptation mechanism that we used in the rule-module: a short-term synaptic depression (Abbott and Nelson, 2000), following the details provided by Dayan and Abbot (2001, p. 185). In particular, the probability of release P_{rel} was decreased after each presynaptic spike by a factor $P_{\text{rel}} = P_{\text{rel}} f_{\text{D}}$ with $f_{\text{D}} = 0.994$. Between presynaptic action potentials the release probability P_{rel} is updated by

$$\tau_P \frac{\mathrm{d}P_{\text{rel}}}{\mathrm{d}t} = P_0 - P_{\text{rel}} \qquad (1)$$

with $P_0 = 1$.

Fig. 8 shows the results of a simulation of the more usual Go/NoGo task design with a pseudorandom sequence of trials (Deco and Rolls, 2005c). On each trial, either a triangle or a square was shown. In Fig. 8, on trial 1 the rule network was 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 biassing 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

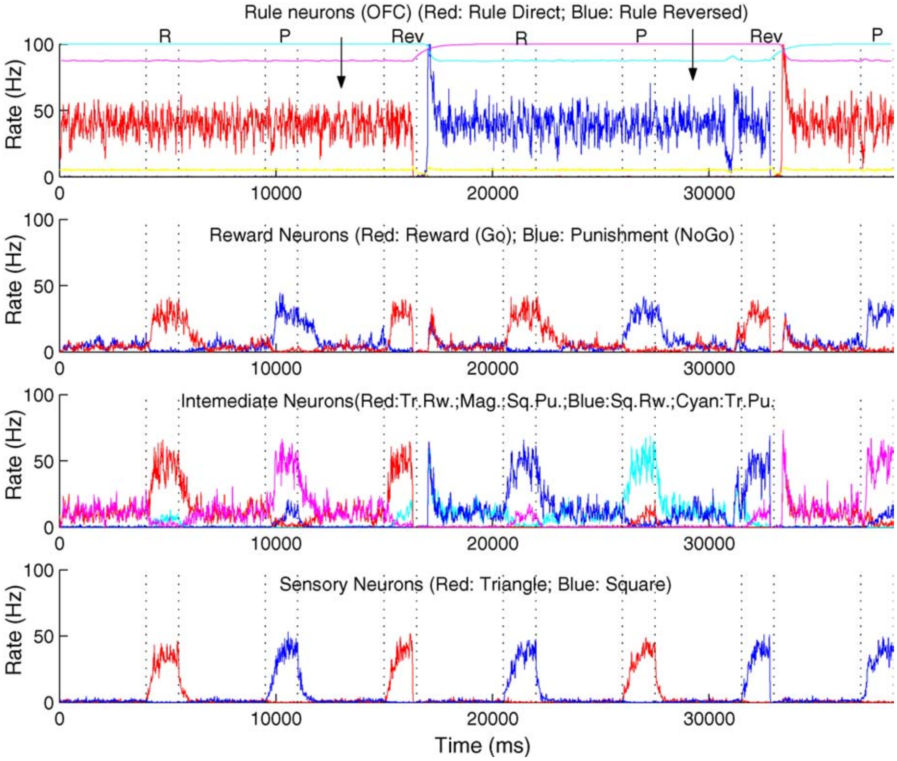


Fig. 8. Model of reward reversal. Temporal evolution of the averaged population activity for all neural pools (sensory, intermediate (stimulus-reward), and Reward/Punishment) in the stimulus– intermediate–reward module and the rule-module during the execution and the reversal of the Go/NoGo visual discrimination 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). The intermediate neurons respond to for example 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 of this graph 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 intertrial interval was 4 s. The yellow line shows the average activity of the inhibitory neurons (see text and Deco and Rolls, 2005c for further details).

the biassing 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.

The results of this dynamical simulation (Deco and Rolls, 2005c) demonstrate one way in which punishers, or the absence of expected rewards, can produce rapid, one trial reversal, and thus close the cognitive loop in such a way that cognitive functions are guided by the rewards being received. The rapid reversal is in this case the reversal of a reward, but applies equally to the rule-biased mapping from stimulus to response in the architecture illustrated in Fig. 5 (Deco and Rolls, 2003), and this stimulus-to-response reversal was also simulated by Deco and Rolls (2005c).

The model can in principle be extended to more than one rule, and indeed, we have described a network that can recall the order of a sequence of several items that are presented (Deco and Rolls, 2005b). The principle is that after a sequence of items has been presented to the attractor network with some habituation in the synapses or neurons, each time the network is quenched, the least recently presented item will be retrieved from the spontaneous firing, as its synapses or neurons are the least adapted. More that one rule has also been simulated in a model of the Wisconsin task (Stemme et al., 2005).

5. Conclusions

In summary, computational neuroscience provides a mathematical framework for studying the mechanisms involved in brain function, such as visual attention, working memory, and the control of behaviour by reward mechanisms, as reviewed here. Analysis of networks of neurons each implemented at the integrate-and-fire neuronal level, and including non-linearities, enables many aspects of brain function, from the spiking activity of single neurons and the effects of pharmacological agents on synaptic currents, through fMRI and neuropsychological findings, to be integrated (see Appendices A and B), and many aspects of cognitive function including visual cognition and attention to be modelled and understood. The theoretical approach described makes explicit predictions at all these levels which can be tested to develop and refine our understanding of the underlying processing, and its dynamics. We believe that this kind of analysis is fundamental for a deep understanding in neuroscience of how the brain performs complex tasks.

Acknowledgements

Gustavo Deco was supported by ICREA and research was supported by the Medical Research Council, the Wellcome Trust, and the Oxford McDonnell Centre for Cognitive Neuroscience.

Appendix A. The mathematics of neurodynamics

For the reasons set out at the start of Section 2, the non-stationary temporal evolution of the spiking dynamics is addressed by describing each neuron by an integrate-and-fire model. The subthreshold membrane potential $V(t)$of each neuron evolves according to the following equation:

$$C_\mathrm{m}\frac{\mathrm{d}V(t)}{\mathrm{d}t} = -g_\mathrm{m}(V(t) - V_\mathrm{L}) - I_\mathrm{syn}(t) \tag{A.1}$$

where $I_\mathrm{syn}(t)$ is the total synaptic current flow into the cell, V_L the resting potential, C_m the membrane capacitance, and g_m the membrane conductance. When the membrane potential $V(t)$ reaches the threshold θ a spike is generated, and the membrane potential is reset to V_reset. The neuron is unable to spike during the first τ_ref which is the absolute refractory period.

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

$$I_\mathrm{syn}(t) = I_\mathrm{AMPA,ext}(t) + I_\mathrm{AMPA,rec}(t) + I_\mathrm{NMDA,rec}(t) + I_\mathrm{GABA}(t) \tag{A.2}$$

where the current generated by each receptor type follows the general form:

$$I(t) = g(V(t) - V_\mathrm{E})\sum_{j=1}^{N} w_j s_j(t) \tag{A.3}$$

and $V_\mathrm{E} = 0$ mV for the excitatory (AMPA and NMDA) synapses and −70 mV for the inhibitory (GABA) synapses. The synaptic strengths w_j are specified by the architecture. The time course of the current flow through each synapse is dynamically updated to describe its decay by altering the fractions of open channels s according to equations with the general form:

$$\frac{\mathrm{d}s_j(t)}{\mathrm{d}t} = -\frac{s_j(t)}{\tau} + \sum_k \delta(t - t_j^k) \tag{A.4}$$

where the sums over k represent a sum over spikes emitted by presynaptic neuron j at time t_j^k, and τ is set to the time constant for the relevant receptor. In the case of the NMDA receptor, the rise time as well as the decay time is dynamically modelled, as it is slower. Details are provided by Deco and Rolls (2003).

The problem now is how to analyze the dynamics and how to set the parameters which are not biologically constrained by experimentally determined values. The standard trick is to simplify the dynamics via the *mean-field* approach at least for the stationary conditions, i.e. for periods after the dynamical transients, and to analyze there exhaustively the bifurcation diagrams of the dynamics. This enables a posteriori selection of the parameter region which shows in the bifurcation diagram the emergent behaviour that we are looking for (e.g. sustained delay activity, biased competition, etc.). 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 we were looking for (Del Giudice et al., 2003; Brunel and Wang, 2001; Fusi and Mattia, 1999). This philosophy is schematized in Fig. A.1.

In the standard mean-field approach, the network is partitioned into populations of neurons, which share the same statistical properties of the afferent currents, and fire spikes independently at the same rate. 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 *p*opulation transfer function $F(\cdot)$, which provides the average population rate as a function of the average input current. The set of stationary, self-reproducing rates ν_i for the different populations i in the network can be found by solving a set of coupled self-consistency equations:

$$\nu_i = F(\mu_i(\nu_1, \ldots, \nu_N), \sigma_i(\nu_1, \ldots, \nu_N)) \qquad \text{(A.5)}$$

where $\mu_i(\cdot)$ and $\sigma_i(\cdot)$ are the mean and standard deviation of the corresponding input current, respectively. To solve these equations, a set of first-order differential equations, describing a *fake dynamics* of the system, whose fixed point solutions correspond to the solutions of Eq. (A.5), is used:

$$\tau_i \frac{d\nu_i(t)}{dt} = -\nu_i(t) + F(\mu_i(\nu_1, \ldots, \nu_N), \sigma_i(\nu_1, \ldots, \nu_N)) \qquad \text{(A.6)}$$

The standard mean-field approach neglects the temporal properties of the synapses, i.e. considers only delta-like spiking input currents. Consequently, after this simplification, the transfer function $F(\cdot)$ is an Ornstein-Uhlenbeck solution for the simplified integrate-and-fire equation $\tau_x \frac{dV(t)}{dt} = -V(t) + \mu_x + \sigma_x \sqrt{\tau_x}\eta(t)$, as detailed by Brunel and Wang (2001). An extended mean-field framework, which is consistent with the integrate-and-fire and synaptic equations described above, i.e. that considers both the fast and slow glutamatergic excitatory synaptic dynamics (AMPA and NMDA), and the dynamics of

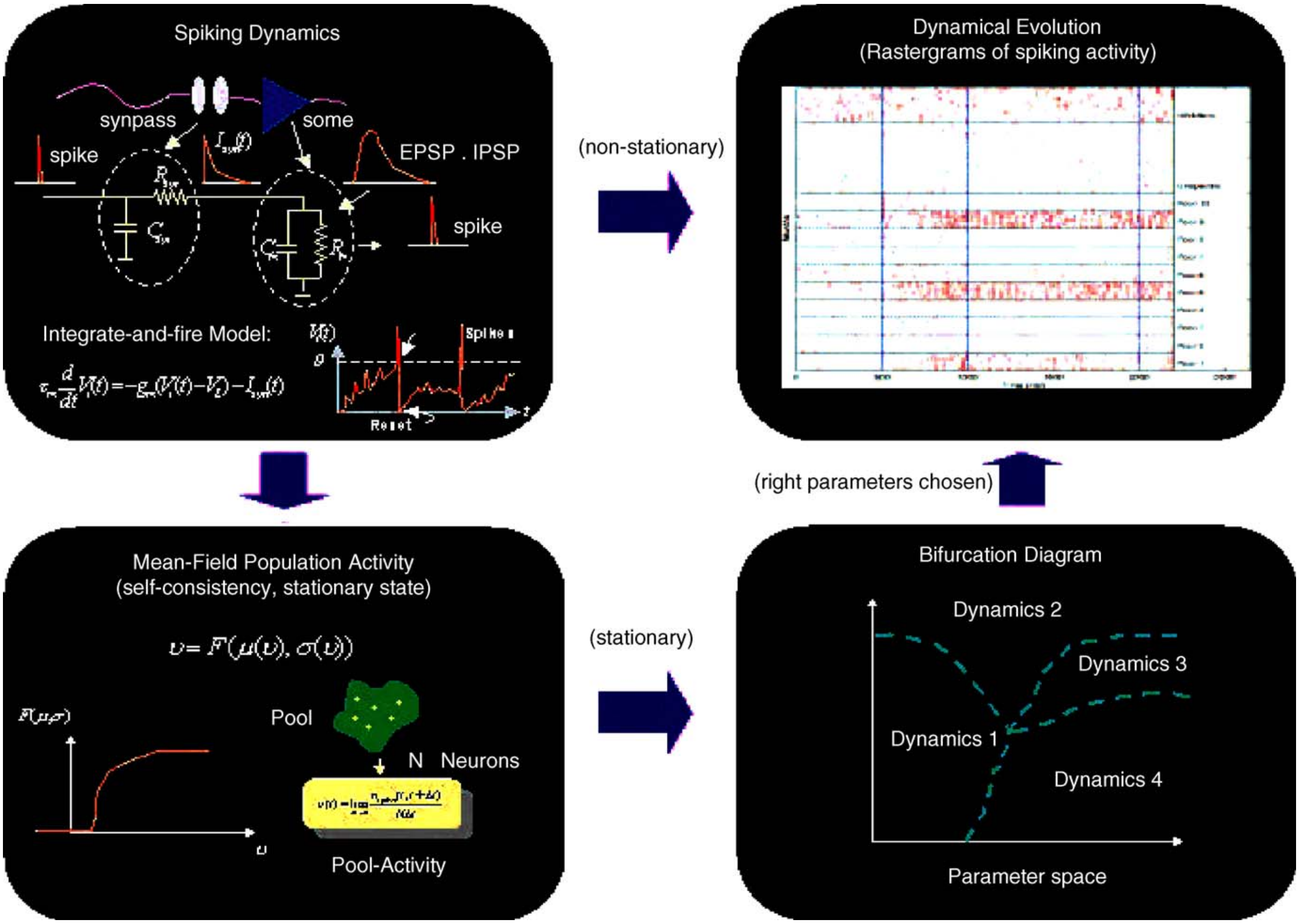


Fig. A.1. The approach to understanding the dynamical systems described in this paper utilizes a mean-field approach to establish the static parameters of the system and how the system operates for different values of the parameters; and the integrate-and-fire approach is used to identify the spiking dynamics of the system with the parameters defined in the mean-field analysis (see text).

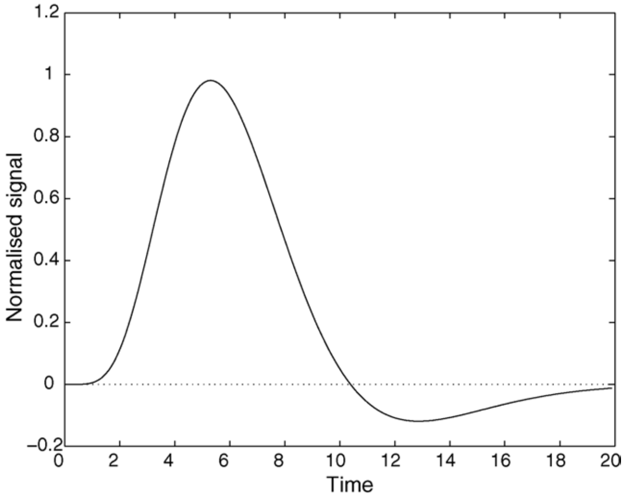


Fig. B.1. The hemodynamic standard response function $h(t)$(see text).

GABA-inhibitory synapses, was derived by Brunel and Wang (2001).

Appendix B. Simulation of fMRI signals: hemodynamic convolution of synaptic activity

The functional magnetic resonance neuroimaging (fMRI) BOLD (blood oxygen level-dependent) signal, which is likely to reflect the total synaptic activity in an area (as ions need to be pumped back across the cell membrane) rather than the spiking neuronal activity (Logothetis et al., 2001), is spatially and temporally filtered. The filtering reflects the inherent spatial resolution with which the blood flow changes, as well as the resolution of the scanner, and filtering which may be applied for statistical purposes, and the slow temporal response of the blood flow changes (Glover, 1999; Buxton and Frank, 1997; Buxton et al., 1998). Glover (1999) demonstrated that a good fitting of the hemodynamical response $h(t)$ can be achieved by the following analytic function:

$$h(t) = c_1 t^{n_1} \mathrm{e}^{-t/t_1} - a_2 c_2 t^{n_2} \mathrm{e}^{-t/t_2}, \quad c_i = \max(t^{n_i} \mathrm{e}^{-\frac{t}{t_i}})$$

where t is the time, and c_1, c_2, a_2, n_1, and n_2 are parameters that are adjusted to fit the experimentally measured hemodynamical response. Fig. B.1 plots the hemodynamic standard response $h(t)$ for a biologically realistic set of parameters (see Deco et al., 2004).

The temporal evolution of fMRI signals can be simulated from an integrate-and-fire population of neurons (see Appendix A) by convolving the total synaptic activity in the simulated population of neurons with the standard hemodynamic response formulation of Glover (1999) presented above (Deco et al., 2004; Horwitz and Tagamets, 1999)). The total synaptic current (I_{syn}) is given by the sum of the absolute values of the glutamatergic excitatory components (implemented through NMDA and AMPA receptors) and inhibitory components (GABA) (Tagamets and Horwitz, 1998; Horwitz et al., 1999; Rolls and Deco, 2002; Deco et al., 2004). As described in Appendix A, in our simulations the external excitatory contributions are produced through AMPA receptors ($I_{\mathrm{AMPA,ext}}$), while the excitatory recurrent synaptic currents are produced through AMPA and NMDA receptors ($I_{\mathrm{AMPA,rec}}$ and $I_{\mathrm{NMDA,rec}}$). The GABA-inhibitory currents are denoted by I_{GABA}. Consequently, the simulated fMRI signal activity S_{fMRI} is calculated by the following convolution equation:

$$S_{\mathrm{fMRI}}(t) = \int_0^{\infty} h(t - t') I_{\mathrm{syn}}(t')\,\mathrm{d}t'.$$

References

Abbott, L., Nelson, S., 2000. Synaptic plasticity: taming the beast. Nat. Neurosci. 3, 1178–1183.

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

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

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

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

Buracas, G.T., Albright, T.D., Sejnowski, T.J., 1996. Varieties of attention: a model of visual search. In: Proceedings of the Third Joint Symposium on Neural Computation, Institute of Neural Computation, pp. 11–25.

Buxton, R.B., Frank, L.R., 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. J. Cereb. Blood Flow Metab. 17, 64–72.

Buxton, R.B., Wong, E.C., Frank, L.R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magnet. Reson. Med. 39, 855–864.

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 (London) 363, 345–347.

Colby, C.L., Duhamel, J.R., Goldberg, M.E., 1993. Ventral intraparietal area of the macaque—anatomic location and visual response properties. J. Neurophysiol. 69, 902–914.

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.

Daugman, J., 1988. Complete discrete 2D-Gabor transforms by neural networks for image analysis and compression. IEEE Trans. Acoust. Speech Signal Process. 36, 1169–1179.

Dayan, P., Abbott, L., 2001. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. MIT Press, Cambridge, MA.

De Valois, R.L., De Valois, K.K., 1988. Spatial Vision. Oxford University Press, New York.

Deco, G., 2001. Biased competition mechanisms for visual attention. In: Wermter, S., Austin, J., Willshaw, D. (Eds.), Emergent Neural Computational Architectures Based on Neuroscience. Springer, Heidelberg, pp. 114–126.

Deco, G., Lee, T.S., 2002. A unified model of spatial and object attention based on inter-cortical biased competition. Neurocomputing 44–46, 775–781.

Deco, G., Lee, T.S., 2004. The role of early visual cortex in visual integration: a neural model of recurrent interaction. Eur. J. Neurosci. 20, 1089–1100.

Deco, G., Rolls, E.T., 2002. Object-based visual neglect: a computational hypothesis. Eur. J. Neurosc. 16, 1994–2000.

Deco, G., Rolls, E.T., 2003. Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. Eur. J. Neurosci. 18, 2374–2390.

Deco, G., Rolls, E.T., 2004. A neurodynamical cortical model of visual attention and invariant object recognition. Vision Res. 44, 621–644.

Deco, G., Rolls, E.T., 2005a. Neurodynamics of biased competition and cooperation for attention: a model with spiking neurons. J. Neurophysiol. 94, 295–313.

Deco, G., Rolls, E.T., 2005b. Sequential memory: a putative neural and synaptic dynamical mechanism. J. Cogn. Neurosci. 17, 294–307.

Deco, G., Rolls, E.T., 2005c. Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. Cereb. Cortex 15, 15–30.

Deco, G., Zihl, J., 2001. Top-down selective visual attention: a neurodynamical approach. Visual Cogn. 8, 119–140.

Deco, G., Pollatos, O., Zihl, J., 2002. The time course of selective visual attention: theory and experiments. Vision Res. 42, 2925–2945.

Deco, G., Rolls, E.T., Horwitz, B., 2004. 'What' and 'where' in visual working memory: a computational neurodynamical perspective for integrating fMRI and single-neuron data. J. Cogn. Neurosc. 16, 683–701.

Del Giudice, P., Fusi, S., Mattia, M., 2003. Modeling the formation of working memory with networks of integrate-and-fire neurons connected by plastic synapses. J. Physiol. (Paris) 97, 659–681.

Duhamel, J.R., Colby, C.L., Goldberg, M.E., 1992. The updating of the representation of visual space in parietal cortex by intended eye movements. Science 255, 90–92.

Duncan, J., 1996. Cooperating brain systems in selective perception and action. In: Inui, T., McClelland, J.L. (Eds.), Attention and Performance XVI. MIT Press, Cambridge, MA, pp. 549–578.

Duncan, J., Humphreys, G., 1989. Visual search and stimulus similarity. Psychol. Rev. 96, 433–458.

Duncan, J., Humphreys, G., Ward, R., 1997. Competitive brain activity in visual attention. Curr. Opin. Neurobiol. 7, 255–261.

Elliffe, M.C.M., Rolls, E.T., Parga, N., Renart, A., 2000. A recurrent model of transformation invariance by association. Neural Networks 13, 225–237.

Elliffe, M.C.M., Rolls, E.T., Stringer, S.M., 2002. Invariant recognition of feature combinations in the visual system. Biol. Cybernet. 86, 59–71.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb. Cortex 1, 1–47.

Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorisation. J. Neurosci. 23, 5235–5246.

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

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

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

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

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

Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.

Gnadt, J.W., Andersen, R.A., 1988. Monkey related motor planning activity in posterior parietal cortex of macaque. Exp. Brain Res. 70, 216–220.

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

Grossberg, S., 1987. Competitive learning: from interactive activation to adaptive resonance. Cogn. Sci. 11, 23–63.

Haenny, P.E., Schiller, P.H., 1988. State dependent activity in monkey visual cortex. I. Single cell activity in V1 and V4 on visual tasks. Exp. Brain Res. 69, 225–244.

Heinke, D., Deco, G., Zihl, J., Humphreys, G., 2002. A computational neuroscience account of visual neglect. Neurocomputing 44–46, 811–816.

Heinke, D., Deco, G., Humphreys, G., Zihl, J., 2003. Top-down learning effect on the saccade pattern of patient with visual neglect: a computational neuroscience based model. prediction and experimental confirmation, J. Comput. Neurosci.

Helmholtz, H.V., 1967. Handbuch der physiologischen Optik. Voss, Leipzig.

Horwitz, B., Tagamets, M.-A., 1999. Predicting human functional maps with neural net modeling. Human Brain Mapping 8, 137–142.

Horwitz, B., Tagamets, M.-A., McIntosh, A.R., 1999. Neural modeling, functional brain imaging, and cognition. Trends Cogn. Sci. 3, 85–122.

Hupe, J.M., James, A.C., Payne, B.R., Lomber, S.G., Girard, P., Bullier, J., 1998. Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394, 784–787.

Ito, M., Gilbert, C., 1999. Attention modulates contextual influences in the primary visual cortex of alert monkeys. Neuron 22, 593–604.

Itti, L., Koch, C., 2001. Computational modelling of visual attention. Nature Rev. Neurosci. 2, 194–203.

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

Kastner, S., De Weerd, P., Desimone, R., Ungerleider, L., 1998. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. Science 282, 108–111.

Kastner, S., Pinsk, M., De Weerd, P., Desimone, R., Ungerleider, L., 1999. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22, 751–761.

Lamme, V.A.F., 1995. The neurophysiology of figure-ground segregation in primary visual cortex. J. Neurosci. 15, 1605–1615.

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.

Lee, T., Nguyen, M., 2001. Dynamics of subjective contour formation in the early visual cortex. Proc. Natl. Acad. Sci. 98, 1907–1911.

Lee, T.S., 1996. Image representation using 2D Gabor wavelets. IEEE Trans. Pattern Anal. Mach. Intell. 18 (10), 959–971.

Lee, T.S., Mumford, D.R.R., Lamme, V.A.F., 1998. The role of primary visual cortex in higher level vision. Vision Res. 38, 2429–2454.

Lee, T.S., Yang, C.Y., Romero, R.D., Mumford, D., 2002. Neural activity in early visual cortex reflects behavioral experience and higher order perceptual saliency. Nat. Neurosci. 5, 589–597.

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

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150–157.

Loh, M., Szabo, M., Ameida, R., Stetter, M., Deco, G., 2004. Computational neuroscience for cognitive brain functions. In: Dubitzky, W., Azuaje, F. (Eds.), Artificial Intelligence: Methods and Tools for Systems Biology: Computational Biology Series, vol. 5. Kluwer, Reihe.

Luck, S., Chelazzi, L., Hillyard, S., Desimone, R., 1997. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J. Neurophysiol. 77, 24–42.

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

Marr, D., 1982. Vision. W. Freeman, San Francisco.

Martinez-Trujillo, J., Treue, S., 2002. Attentional modulation strength in cortical area MT depends on stimulus contrast. Neuron 35, 365–370.

Mattia, M., Del Giudice, P., 2002. Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. Phys. Rev. E 66, 51917–51919.

Mattia, M., Del Giudice, P., 2004. Finite-size dynamics of inhibitory and excitatory interacting spiking neurons. Phys. Rev. E 70, 052903.

Maunsell, J.H.R., 1995. The brain's visual world: representation of visual targets in cerebral cortex. Science 270, 764–768.

McAdams, C., Maunsell, J., 1999. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J. Neurosci. 19, 431–441.

McClelland, J.L., Rumelhart, D.E., 1981. An interactive activation model of context effects in letter perception. Part I. An account of basic findings. Psychol. Rev. 88, 375–407.

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.

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.

Motter, B., 1994. Neural correlates of attentive selection for colours or luminance in extrastriate area V4. J. Neurosci. 14, 2178–2189.

Mozer, M., Sitton, M., 1998. Computational modeling of spatial attention. In: Pashler, H. (Ed.), Attention. UCL Press, London, pp. 341–393.

O'Reilly, J., Munakata, Y., 2000. Computational Explorations in Cognitive Neuroscience. MIT Press, Cambridge, MA.

Phaf, H., Van der Heijden, A., Hudson, P., 1990. A connectionist model for attention in visual selection tasks. Cogn. Psychol. 22, 273–341.

Pollen, D., Ronner, S., 1981. Phase relationship between adjacent simple cells in the visual cortex. Science 212, 1409–1411.

Postle, B.R., D'Esposito, M., 2000. Evaluating models of the topographical organization of working memory function in frontal cortex with event-related fMRI. Psychobiology 28, 132–145.

Quinlan, P.T., Humphreys, G.W., 1987. Visual search for targets defined by combination of color, shape, and size: an examination of the task constraints on feature and conjunction searches. Percept. Psychophys. 41, 455–472.

Rao, S., Rainer, G., Miller, E., 1997. Integration of what and where in the primate prefrontal cortex. Science 276, 821–824.

Renart, A., Parga, N., Rolls, E.T., 1999a. Associative memory properties of multiple cortical modules. Network 10, 237–255.

Renart, A., Parga, N., Rolls, E.T., 1999b. Backprojections in the cerebral cortex: implications for memory storage. Neural Comput. 11, 1349–1388.

Reynolds, J., Desimone, R., 1999. The role of neural mechanisms of attention in solving the binding problem. Neuron 24, 19–29.

Reynolds, J., Chelazzi, L., Desimone, R., 1999. Competitive mechanisms subserve attention in macaque areas V2 and V4. J. Neurosci. 19, 1736–1753.

Reynolds, J.H., Pastemak, T., Desimone, R., 2000. Attention increases sensitivity of v4 neurons. Neuron 26, 703–714.

Riesenhuber, M., Poggio, T., 2000. Models of object recognition. Nat. Neurosci. Suppl. 3, 1199–1204.

Roelfsema, P.R., Lamme, V.A., Spekreijse, H., 1998. Object-based attention in the primary visual cortex of the macaque monkey. Nature 395, 376–381.

Rolls, E.T., 1992. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Philos. Trans. Roy. Soc. 335, 11–21.

Rolls, E.T., 1999. The Brain and Emotion. Oxford University Press, Oxford.

Rolls, E.T., 2000a. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron 27, 205–218.

Rolls, E.T., 2000b. The orbitofrontal cortex and reward. Cereb. Cortex 10, 284–294.

Rolls, E.T., 2003. Consciousness absent and present: a neurophysiological exploration. Prog. Brain Res. 144, 95–106.

Rolls, E.T., 2004. The functions of the orbitofrontal cortex. Brain Cogn. 55, 11–29.

Rolls, E.T., 2005. Emotion Explained. Oxford University Press, Oxford.

Rolls, E.T., Deco, G., 2002. Computational Neuroscience of Vision. Oxford University Press, Oxford.

Rolls, E.T., Milward, T., 2000. A model of invariant object recognition in the visual system: learning rules, activation functions, lateral inhibition, and information-based performance measures. Neural Comput. 12, 2547–2572.

Rolls, E.T., Stringer, S.M., 2001. Invariant object recognition in the visual system with error correction and temporal difference learning. Network: Comput. Neural Syst. 12, 111–129.

Rolls, E.T., Treves, A., 1998. Neural Networks and Brain Function. Oxford University Press, Oxford.

Rolls, E.T., Tovee, M.J., Purcell, D.G., Stewart, A.L., Azzopardi, P., 1994. The responses of neurons in the temporal cortex of primates, and face identification and detection. Exp. Brain Res. 101, 474–484.

Rolls, E.T., Aggelopoulos, N.C., Zheng, F., 2003. The receptive fields of inferior temporal cortex neurons in natural scenes. J. Neurosci. 23, 339–348.

Salinas, E., Abbott, L., 1997. Invariant visual perception from attentional gain fields. J. Neurophysiol. 77, 3267–3272.

Sato, T., 1989. Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. Exp. Brain Res. 77, 23–30.

Senn, W., Markram, H., Tsodyks, M., 2001. An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timin. Neural Comput. 13, 35–67.

Sjöström, P.J., Turrigiano, G.G., Nelson, S.B., 2001. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron 32, 1149–1164.

Spitzer, H., Desimone, R., Moran, J., 1988. Increased attention enhances both behavioral and neuronal performance. Science 240, 338–340.

Stemme, A., Deco, G., Busch, A., Schneider, W.X., 2005. Neurons and the synaptic basis of the fMRI signal associated with cognitive flexibility. Neuroimage 26, 454–470.

Stringer, S.M., Rolls, E.T., 2002. Invariant object recognition in the visual system with novel views of 3D objects. Neural Comput. 14, 2585–2596.

Tagamets, M., Horwitz, B., 1998. Integrating electrophysical and anatomical experimental data to create a large-scale model that simulates a delayed match-to-sample human brain study. Cereb. Cortex 8, 310–320.

Thorpe, S.J., Rolls, E.T., Maddison, S., 1983. Neuronal activity in the orbitofrontal cortex of the behaving monkey. Exp. Brain Res. 49, 93–115.

Thorpe, S.J., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381, 520–522.

Treisman, A., 1982. Perceptual grouping and attention in visual search for features and for objects. J. Exp. Psychol.: Human Percept. Perform. 8, 194–214.

Treisman, A., 1988. Features and objects: the fourteenth Barlett memorial lecture. Quart. J. Exp. Psychol. 40A, 201–237.

Treisman, A., Gelade, G., 1980. A feature-integration theory of attention. Cogn. Psychol. 12, 97–136.

Tsotsos, J., 1990. Analyzing vision at the complexity level. Behav. Brain Sci. 13, 423–469.

Tuckwell, H., 1988. Introduction to Theoretical Neurobiology. Cambridge University Press, Cambridge.

Usher, M., Niebur, E., 1996. Modelling the temporal dynamics of IT neurons in visual search: a mechanism for top-down selective attention. J. Cogn. Neurosci. 8, 311–327.

Wallis, G., Rolls, E.T., 1997. Invariant face and object recognition in the visual system. Prog. Neurobiol. 51, 167–194.

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

Wiskott, L., Sejnowski, T.J., 2002. Slow feature analysis: unsupervised learning of invariances. Neural Comput. 14, 715–770.

Wolfe, J.M., 1994. Guided search 2.0: A revised model of visual search. Psychon. Bull. Rev. 1, 202–238.

Zheng, P., Zhang, X.-X., Bunney, B.S., Shi, W.-X., 1999. Opposite modulation of cortical *N*-methyl-D-aspartate receptor-mediated responses by low and high concentrations of dopamine. Neuroscience 91, 527–535.

Zipser, K., Lamme, V., Schiller, P., 1996. Contextual modulation in primary visual cortex. J. Neurosci. 16, 7376–7389.