Cooperation and biased competition model can explain attentional filtering in the prefrontal cortex

Miruna Szabo,1,2 Rita Almeida,1 Gustavo Deco3 and Martin Stetter1
1Siemens AG, Corporate Technology, Information and Communications, 81730 Munich, Germany
2Department of Computer Science, Technical University of Munich, 85747 Garching, Germany
3Institució Catalana de Recerca i Estudis Avançats (ICREA), Universitat Pompeu Fabra, Department of Technology Computational Neuroscience, Passeig de Circumval.lació 8, 08003 Barcelona, Spain

Keywords: computational neuroscience, inattentional blindness, mean-field analysis, network of spiking neurons, neurocognitive model

Abstract

Recent neurophysiological experimental results suggest that the prefrontal cortex plays an important role in filtering out unattended visual inputs. Here we propose a neurodynamical computational model of a part of the prefrontal cortex to account for the neural mechanisms defining this attentional filtering effect. Similar models have been employed to explain experimental results obtained during the performance of attention and working memory tasks. In this previous work the principle of biased competition was shown to successfully account for the experimental data. To model the attentional filtering effect, the biased competition model was extended to enable cooperation between stimulus selective neurons. We show that, in a biological relevant minimal model, competition and cooperation between the neurons are sufficient conditions for reproducing the attentional effect. Furthermore, a characterization of the parameter regime where the cooperation effect is observed is presented. Finally, we also reveal parameter regimes where the network has different modes of operation: selective working memory, attentional filtering, pure competition and noncompetitive amplification.

Introduction

Selective attention may be defined as a process in which the perception of certain stimuli in the environment is enhanced relative to other concurrent stimuli of less importance. A remarkable phenomenon of selective attention, known as inattentional blindness, has been described for human vision (for a review see Simons, 2000). The inattentional blindness refers to the unawareness of a certain visual event when attention is focused on another event. Recently, Everling et al. (2002), investigated the possible underlying mechanisms of the referred effect, by measuring the activity of prefrontal cortex (PFC) neurons in awake behaving monkeys performing a focused attention task. The experimental results showed that some PFC neurons discriminate between a previously learned target and a nontarget, but that this discrimination disappears if the objects are presented in the unattended visual hemifield. Attention acts not only in a modulatory way but imposes a multiplicative effect on the neural response. This attentional filtering effect of object’s representation for the unattended hemifield is complete and might be the neuronal substrate of the referred selective attention effect studied in humans, possibly explaining the blindness to ignored inputs.

In this work we propose a neurodynamical computational model of a part of the PFC that can explain the underlying neural substrate of the visual attention filtering effect. We use the conceptual framework of the Biased Competition Hypothesis (Chelazzi et al., 1993, Chelazzi, 1998; Desimone & Duncan, 1995; Reynolds & Desimone, 1999) which assumes that multiple activated populations of neurons engage in competitive interactions, mediated by global inhibition, and that external top–down interactions can bias this competition in favour of specific neurons, representing the attended stimulus. Neurodynamical models developed within this framework have been proved to successfully account for different aspects of visual attention (Rolls & Deco, 2002; Corchs et al., 2003) and working memory context dependent tasks (Deco & Rolls, 2003; Deco et al., 2003).

In the present work we extend previously used neurocomputational models (Brunel & Wang, 2001; Deco & Rolls, 2003; Deco et al., 2003) to setup a biological relevant minimal model for the attentional filtering effect. We observed that the mechanism of biased competition alone cannot account for the experimental results and show that biased competition and cooperation between stimulus selective neurons are, in combination, required conditions for reproducing the referred effect. To model the experimental results we used a network of integrate-and-fire neurons, which allows us to reproduce the dynamics of the neurophysiological measurements. Next, we characterize the network’s modes of operation, which correspond to different parameter regimes. This analysis was performed using a mean-field formulation, which has been derived for stationary regimes, from the network of model spiking neurons (Brunel & Wang, 2001).

We will summarize the experiment modelled and the neurodynamical computational model of the prefrontal cortex and present the results obtained for the nonstationary simulations of the network of model spiking neurons and the parameter space exploration carried out using the mean-field formulation, together with a discussion of the results and the conclusions.

Correspondence: Dr Martin Stetter, as above.
E-mail: stetter@siemens.com

Received 5 September 2003, revised 3 November 2003, accepted 25 November 2003

Materials and methods

Focused attention task and inattentive blindness

We model a visual attentional experiment, performed by Everling et al. (2002), that monitors the activity level of single neurons in the prefrontal cortex (PFC) of awake behaving monkeys engaged in a focused attention task. In this experiment, a monkey, after being cued to attend one of two hemifields, had to watch a series of bilateral stimuli that consisted of different pairs of objects, and to react with a saccade if and only if a predefined target object appeared in the cued hemifield. In order to correctly perform this cognitive task, the monkey had to ignore any object in the uncued hemifield and to concentrate (focus his attention) on the cued location. In the experiment, Everling and coworkers first observed neurons which were selective for target or nontarget stimuli, and preferred the stimulus location in one hemifield over the other (cf. fig. 1 in Everling et al., 2002). Next, during the focused attention task, these neurons again discriminated between target and nontarget, but only when the stimulus changed in the attended location. The stimulus in the nonattended location had no influence on the neuronal response (cf. Fig. 3, lb, grey line). This effect we refer to as attentional filtering. In other words, attention acts in a multiplicative way upon the sensory driven neuronal response, and consequently these neurons seem to code for the behavioural relevance of a stimulus rather than for its identity.

When humans perform similar tasks, they cannot recover any information from unattended sensory stimuli, an effect known as inattentive blindness (Simons, 2000). Because the reported response properties of the described neurons strongly resemble this behaviour, they might be part of a neural correlate of inattentive blindness. Inattentive blindness is thought to be part of an important cognitive mechanism, namely that of focusing or ‘concentrating’ on a task to be performed. We use this similarity as motivation to model attentional filtering as a possible neural correlate of cognitive inattentive blindness.

Conductance-based model

We implement a model consisting of groups of neurons that show the same selectivities as found in the experimental results (Everling et al., 2002). Therefore we use four interconnected pools of neurons coding for target with preferred location left (TL), target with preferred location right (TR), nontarget (other) left (OL) and nontarget (other) right (OR). Further we consider a pool with nonselective neurons (representing neurons in the area which are involved in other tasks) and one inhibitory pool (that implements competition in the network). The weights between the pools were chosen to values which might have evolved after training, following a Hebbian learning paradigm.

We assume that a proper level of description at the microscopic level is captured by the spiking and synaptic dynamics assuming one-compartment, point-like neuron models, like Integrate-and-Fire-models. An IF neuron integrates the afferent current generated by the incoming spikes, and fires when the depolarization of the cell membrane crosses a threshold. The subthreshold membrane potential $V(t)$ of a neuron is given by

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

where $I_{syn}(t)$ is the total incoming synaptic current, $C_m$ is the membrane capacitance and $g_m$ is the membrane leak conductance. At this level of detail the model allows the use of realistic biophysical time constants, latencies, and conductances to model the synaptic current, which in turn allows to perform a thorough study of the realistic time scales and firing rates involved in the evolution of the neural activity. Consequently, the simulated dynamical processes, that putatively underly cognitive processes, can be quantitatively contrasted with experimental data. For this reason, it is convenient to include a thorough description of the different latencies of the synaptic activity.

We depart from a biologically plausible conductance-based neuron model and system architecture, which has been recently proposed by Brunel & Wang (2001) as a model for selective working memory. Cells have three types of receptors mediating the synaptic current flowing into them: AMPA, $\ N$-methyl-D-aspartate (NMDA) glutamate receptors; and $\gamma$-aminobutyric acid (GABA) receptors (Fig. 1). The recurrent excitatory postsynaptic currents (EPSCs) have two components, mediated by AMPA (fast) and NMDA (slow) receptors. External EPSCs imposed onto the network from outside are assumed to be driven only by AMPA receptors. The shunting inhibitory GABAAergic synapses inject inhibitory PSCs (IPSCs) into both pyramidal cells and interneurons. The mathematical formulation is summarized in Appendix 1.

Conductance-based model neurons are combined to a spiking neuronal network to form a minimal model of a small part of monkey PFC. The network consists of $N_T = 800$ excitatory pyramidal cells and $N_I = 200$ interneurons, which are fully connected (Fig. 1). The proportion 80% pyramidal cells and 20% interneurons is consistent with the neurophysiological experimental data (Abeles, 1991). The network is structured into distinct populations or pools of neurons, which are defined by sharing the same inputs and weights. There are three types of pools: four stimulus specific ‘selective’ pools, each with $f \times N_T$ neurons, one nonselective pool with $(1-4f) \times N_T$ neurons and one inhibitory pool. For simplicity we used the same fraction $f$ of all $N_T$ excitatory neurons for each selective pool. Throughout this work we used $f = 0.1$. Under a nonattentive control task, the selective pools encode information about the object identity (‘T’ for target, ‘O’ for other) and spatial location (‘L’ for left, ‘R’ for right).

The individual pools are driven by four different kinds of inputs, as summarized in Fig. 2. First, each neuron of the network receives spontaneous background activity from outside the module through $N_{ext} = 800$ external excitatory connections, carrying Poisson spike trains at a spontaneous rate of 3 Hz (the typical value observed in the cerebral cortex (Koch & Fuster, 1989; Wilson et al., 1994). Second,
selective pools are driven by object-specific and unilateral inputs, which are assumed to originate from lower areas which process the visual scene such as to provide these signals.

Besides specific afferent bottom-up input, selective pools are biased by two kinds of top–down inputs. The first top–down signal biases neurons that are selective for the target object. The origin of this signal is not explicitly modelled, but it might originate from a working-memory module that encodes and memorizes context in terms of rules. The second top–down signal, the attention bias, facilitates neurons that have the cued location as preferred location. Also the origin of this bias, which might be sent from a spatial working memory area, is not modelled explicitly here. The nonselective excitatory neurons do not have sensory or biasing inputs. They are thought to be involved in other cognitive tasks and to be only spontaneously and nonselectively active in the present framework. Finally, inhibitory neurons balance the overall activity and implement global competition in the network.

The network is fully connected, but weights can differ depending on the pools being connected. We model the prefrontal cortex of a monkey that has been trained already and do not explicitly model the learning process itself. Instead, the weights are chosen (not learned or optimized in any manner) such as to intuitively match Hebbian learning. The resulting weight setting is presented in Table 1 (cf. also Fig. 1). Within the same selective pool, the neurons are strongly coactivated, and are therefore connected with a stronger than average weight, \( w_\text{c}, > 1 \); between two neurons from two pools selective for different objects, activity is likely to be anticorrelated resulting in weaker than average connections \( w_\text{c}, < 1 \). Further, it is not unreasonable to hypothesize that neurons with the same object selectivity are also often coactivated and consequently linked strongly by Hebbian learning mechanisms, yet weaker than neurons which share both object selectivity and location preference. Therefore, weights between the pools that encode for the same object are set in the range \( w_\text{c}, \leq w_\text{c}, \leq w_\text{c}, \). Activities between selective and nonselective pools are likely to be close to uncorrelated. We set \( w = 1 \) for weights from selective to nonselective pools and \( w_\text{n} \) for the corresponding feedback connections. Finally, all connections from and to the inhibitory pool are weighted by \( w = 1 \). The absolute strengths of the weight parameters are explored to analyse the different operational modes of the model.

A focused attention task was modelled here by applying each of the four different stimulus combinations used in Everling et al. (2002) (Fig. 2a–d) and calculating the population-averaged spike rate of the target specific right preferred TR pool. In this situation, the attention bias set to right preferred neurons corresponds to the condition ‘preferred location attended’, a left bias corresponds to the ‘nonpreferred location attended’ condition.

Explicit simulations of the network dynamics accurately capture the temporal dynamics and any order spike statistics. However, they are computationally expensive and not easy to use for systematic parameter explorations. Therefore we used mean-field models, which represent a well-established means for efficiently analysing the approximate network behaviour (Stetter, 2002), at least for the stationary conditions (i.e. after the dynamical transients), in order to systematically explore parameter regimes of qualitatively different network response. The mean field study assures us that the network dynamics will converge to a stationary attractor that is consistent with the asymptotic behaviour of an asynchronously firing spiking network (see Del Giudice et al., 2003; Brunel & Wang, 2001; Fusi & Mattia, 1999). After identifying different regimes, it is necessary to go back and explore the full nonstationary behaviour using the model spiking network. For our mean-field analysis, we used a recent derivation by Brunel & Wang (2001), which is consistent with the type of conductance-based neuron networks we simulate. Starting from the equation for the dynamics of a single neuron, a stochastic analysis of the mean-first passage time of the membrane potentials is performed, resulting in the firing frequencies of the populations as a function of the model parameters. The mathematical framework used is summarized in Appendix 2.

**Results**

Explicit simulation of spike dynamics

Explicit simulations were carried out in the framework of the architecture presented in Fig. 2 by applying each of the four bilateral stimuli combinations for a time period of 300 ms, with a total strength of 400 Hz per stimuli-input, distributed over 800 afferent fibres. The target and attentional biases were set to 16 and 160 Hz, respectively, and were left constant throughout the entire simulation run. Using this setting, the 1000 coupled equations 1 were integrated numerically using the second order Runge–Kutta method with step size 0.1 ms. After 800 ms from stimulus onset, an excitatory flush of strength 5 Hz per afferent fibre was given to all neurons, which simulates a strong
Fig. 3. Experimental results (column 1) and model simulation (columns 2, 3 and 4) for focused attention task. Black lines: Attention focused to the preferred location (right), grey lines: attention focused to the nonpreferred location of the measured neurons and model-neurons, respectively. (a) Both target stimuli. (b) Target in preferred location only. (c) Target in nonpreferred location. (d) Both nontarget stimuli. Column 2, simulation with cooperation and competition: \( w_+ = 1.6, w' = 1.6, w_n = 0.62, w_- = 0.3 \). Column 3, simulation with competition only: \( w_+ = 1.6, w' = w_n = 0.76, w_- = 0.3 \). Column 4, simulation with cooperation only: \( w_+ = 1.6, w' = 1.6, w_- = 1, w_n = 0.65 \). For further explanations see text.

generic brain activity and causes the reset of the network activities (mediated over the inhibitory interneurons).

The left column of Fig. 3 (column 1) displays the experimental results from Everling et al. (2002) in the case of the four stimulus combinations illustrated as insets. The black lines correspond to attention directed to the preferred location and the grey lines correspond to attention directed to the nonpreferred location. In the second from left column (Fig. 3, column 2), the population-averaged responses of the model ‘target right selective’ (TR) neurons for the same stimulus conditions and attentional states as the experimental results are shown. In this network, competition and cooperation are combined by using the parameters \( w_+ = 1.6, w' = 1.6, w_n = 0.62 \) and \( w_- = 0.3 \). It can be observed, that the simulation results for this parameter set follow closely the experimental measurements in the PFC of awake behaving monkeys (Fig. 3, column 1).

In particular, our simple network shows attentional filtering of the information in the unattended hemisphere. Attentional filtering consists of four different phenomena which can be assigned to the four stimulus conditions: (i) when both hemispheres contain target stimuli, the response reflects whether the attended stimulus is in the preferred or nonpreferred location. (Fig. 3, 1a and 2a); (ii) when a target appears in the preferred location only, the response is completely shut down, as soon as attention is shifted away from the target-stimulated side (Fig. 3, 1b and 2b). We refer to this effect as attentional suppression; (iii) in contrast, when a target appears in the nonpreferred location, the neural response is increased, as soon as attention is shifted towards it (Fig. 3, 1c and 2c). We refer to this effect as attentional facilitation; (iv) finally, when both hemispheres are stimulated with nontargets, the response stays low, reflecting the target-selectivity of neurons (Fig. 3, 1d and 2d). In combination of these effects, the neurons both in the experiment and in the model encode only the contents of the attended hemisphere [compare black lines in Fig. 3 (columns 1 and 2), a and b with c and d, respectively, and compare the grey lines in Fig. 3 (columns 1 and 2), a and c with b and d, respectively], and ignore the contents of the nonattended hemisphere [compare black lines in Fig. 3 (columns 1 and 2), a with b and c with d, and compare the grey lines in Fig. 3 (columns 1 and 2), a with c and b with d]. The contents of the nonattended hemisphere is not encoded in the responses.

Next we examined what are the roles of cooperation and competition, respectively, for producing attentional filtering. For this we carried out two further simulations, when the network was equipped with competition only (Fig. 3, column 3) and when it was equipped with cooperation only (Fig. 3, column 4). We say that a network exhibits cooperation when, due to recurrent processing, the activities of different pools end up being equalized. In contrast, a network exhibits competition, when different pools try to strengthen their activities at the expense of others, in other words when differences between activities become amplified. In order to observe cooperation, the network needs a strong weight \( w' \) between pools coding for the same category. Intuitively speaking, the strong \( w' \) accounts for the...
equalization of activities of target pools (or nontarget-pools) with each other. In contrast, in order to observe competition, a weak weight \( w^* \) is required. Intuitively speaking, the small \( w^* \) between competing pools cannot counteract the global inhibition, which then succeeds in amplifying differences between activities. When the network is dominated by competition (Fig. 3, column 3), the competition causes a complete attentional suppression of unattended stimuli (Fig. 3, 3b), however, there is no attentional facilitation (see the zero activity in Figs 3, 3c). This is the case, because in the present model the facilitation effect is caused by a lateral propagation of activity from the stimulated TL pool to the nonstimulated TR pool over recurrent connections. Because these connections are too weak in the competition only setting (i.e. \( w^* \) is too small), facilitation does not occur. When the network is dominated by cooperation (Fig. 3, column 4), activities between attended and nonattended conditions are equalized, and as a consequence attentional effects are diminished (compare black with grey lines in Fig. 3, column 4). In particular, attentional suppression is no longer observed.

In summary, competition, mediated by a small weight \( w^* \), implements attentional suppression, and cooperation, mediated by a strong weight \( w^* \), implements attentional facilitation. When both mechanisms act together, our model shows a strong, all-or-none attentional filtering effect, which results from the effects of weak top-down biases.

Concrete predictions of the model, suitable for experimental testing, can be formulated by observing the results of manipulating the dopamine level. With this aim we studied the effects of an increase in dopamine concentration, leading to an increase in D2 receptor activation, which is modelled by a decrease in both NMDA and GABA conductances (Law-Tho et al., 1994; Zheng et al., 1999). A weak increase in D2 receptor activation was modelled by multiplying both NMDA and GABA conductances by a factor of 0.7. A strong increase in D2 receptor activation was modelled by multiplying the same quantities with 0.2. We found that, when the dopamine level increases, first, for the weak increase in D2 receptor activation, the attention facilitation effect previously defined becomes impaired. As the level of dopamine is further increased, the attentional suppression effect also becomes impaired. Hence, the results suggest that an increase in dopamine concentration will lead to a progressive weakening of the attentional filtering effect.

**Mean-field parameter exploration**

Motivated by the described emergent property of a biased competition network with cooperation within intracortical local connection, we next performed a systematic exploration of weight parameters of the structured network of model spiking neurons in the presence of cooperation. We used the mean-field approximation described in Appendix 2, which is fully consistent with the biophysical plausible spiking neurons network. Unless otherwise stated, mean-field simulations were carried out as follows: The firing frequencies per neuron were initialized to 3 Hz for the excitatory pools and 9 Hz for the inhibitory pool. The external input received by each neuron of the network was set to \( 800 \times 3 \text{ Hz} = 2.4 \text{ kHz} \). An attentional-bias of 100 Hz and a target-bias of 30 Hz were applied to the respective pools throughout the whole simulation. An external stimulus was presented by providing a weak extra input of 200 Hz to the stimulated selective pools, while integrating the mean-field equations over 1000 iterations.

The mentioned values for inputs and biases have been chosen after exploring the effect of different signal strengths on the results. It was found that the network behaviour does not considerably depend on their exact values, as long as they stay within certain regimes: First, the target bias should not be too strong, because otherwise it causes unattended target stimuli to win the competition implemented by

the inhibitory neurons, but it is needed to stabilize target stimuli to win generally over nontarget stimuli. A suitable target-bias can be thought of breaking the symmetry between target and nontarget but weakly affecting the network behaviour otherwise. Second, the attentional bias should be not too weak compared with the input. It causes attentional filtering to be the more complete, the stronger the attentional bias is. For the present network, the target bias should be in the range of 30 Hz or less, and the attentional bias in the range of 100 Hz or more, but it should be stressed that these numbers are only coarse estimates and because of their dependency on network size, structure and parameters cannot be taken as a quantitative predictions but only as qualitative trends.

Next we examined how different values for the weights connecting the selective and nonselective neurons affect the collective behaviour of the network. We explore the space of parameters \( w_{n} \) and \( w^{*} \) in the value range from 0 to 1 in steps of 0.1 (Fig. 4), while keeping \( w_{n} = 1.6 \) and \( w^{*} = 1.6 \) fixed to the same values used in the simulation of Fig. 3 (column 2). Seeing attentional filtering in our network in the same way as in the experiment, requires four different response properties in combination: (i) the network must show responses at all; (ii) there should be no memory trace by persistent poststimulus activity; (iii) the network should carry out competition between target and nontarget pools, to implement the attentional suppression effect; and (iv), target or nontarget pools need to cooperate, to implement the attentional facilitation effect. We tested the presence or absence of these four properties over the parameter plane, and the results are summarized in Fig. 4.

First we tested for responsiveness of the network. This was done by applying the most effective stimulus (target left and right) and testing if the activity of the target pools exceeded a threshold of 10 Hz. The dotted line in Fig. 4 marks the border, almost parallel to the \( w_{n} \) axis, above which (approximately, \( w_{n} > 0.5 \)) the network is effectively driven by a weak afferent input as used here. Above this border the total excitation in the network becomes sufficient to amplify a small external input.

Next, we tested for the regime where, besides being responsive, the stimulated pools show persistent poststimulus activity. This was done by
by initializing the activity of each neuron to 50 Hz for the two cooperating target pools, to 3 Hz for the other excitatory pools, and 9 Hz for the inhibitory pool. This scenario was chosen to simulate the immediate poststimulus phase of the network, where the stimulus just disappeared but the pools previously driven by input were still active. Persistent activity was assumed, if the activity of the pools after 1000 iterations still exceeded 3 Hz. The dashed line in Fig. 4 separates the regimes of persistent activity ($w_{in} > 0.71$) from pure stimulus coding. Above the border, the reverberations in the network become strong enough to autonomously stabilize activity.

The presence of competition was tested by applying one target and one nontarget stimulus and checking, if the cooperating nontarget pools could win against the target pools when the stimulated nontarget pool was biased by an attentional signal. Winning was tested by requiring the losing target right preferred pool to not exceed 3 Hz activity. Below the dash-dotted line in Fig. 4 the network shows competition. Below this border the common input provided by the fibres from nonselective neurons and the equalizing effect provided by $w_{in}$ is weak enough to allow for efficient competition.

Finally, we examined under which conditions cooperation between neurons selective for the same object became prominent. For this we provided both input and bias to one target specific pool, an unbiased nontarget to the other hemifield, and measured, whether the other target-specific pool could be coactivated even without receiving any afferent input or bias. Above the solid line in Fig. 4 the system shows cooperation.

From Fig. 4, a number of conclusions can be drawn. First, it demonstrates that there exists a variety of different regimes, where the network of model spiking neurons shows qualitatively different kinds of behaviour. For example it is predicted that cortical networks can show attentional filtering in combination with selective working memory, pure competition, and noncompetitive stimulus amplification. Second, in one of these regimes, in which the network exerts both competition and cooperation, but no persistent activity, attentional filtering is found. Hence, besides a weights setting implementing cooperation (here $w_{in} = w' = 1.6$), these ingredients are predicted to be necessary for attentional filtering to occur. Third, in order to obtain and stabilize this and other regimes, a pool of nonselective neurons which drives the selective pools, proves necessary as well. Fourth, the weights from this pool to the selective pools must be strong enough to allow for sufficient amplification and for stabilizing cooperation. Note that the connectivity of the nonselective pool is nonsymmetric. This weight setting was chosen in order to remain close to the original model. Changing the weights of the nonselective pool to be symmetric and equal to $w_{in}$ gives qualitatively the same results as with the nonsymmetric weights.

It turns out, that for stabilizing most of the regimes found, $w_{in}$ should be weak, but otherwise does not strongly influence the behaviour of the system. Therefore, we set $w_{in} = 0$ for the following and explored, how the parameter regimes vary when $w_{in}$ and $w'$ are changed. Figure 5 summarizes the borders of the regimes, which are plotted as a function of $w_{in}$ for a number of $w_{in}$ and $w'$ combinations. All regimes are relatively robust, when these weights are changed. Changing $w_{in} = w'$ in concert results only in a shift of the regimes along the $w_{in}$ axis. However when these weights become too small, attentional filtering destabilizes. Likewise, if $w'$ becomes too small, attentional filtering vanishes as well. In summary, cooperation, $w' > w_{in}$, in combination with competition ($w_{in}$ small, $w_{in}$ intermediate) appears to be important to generate attentional filtering.

Most of the excitatory neurons considered in the model belong to a pool of nonselective neurons. In the brain, these neurons would probably correspond to neurons contributing to the implementation of some function not related to the task we want to model. In particular some of these neurons could encode stimuli, which are irrelevant for the task we consider. This being the case, we investigated how the model behaves in the presence of task-irrelevant stimuli, by providing stronger external inputs to the nonselective neurons. For simplicity, this situation was modelled by increasing the total external input that each nonselective neuron receives from 2.4 kHz to 2.4 + $x$ kHz. In this case, the resultant increase in total external input to the nonselective neurons is equal to that achieved by increasing the total external input to just 80 neurons (the size considered for one selective pool) from 2.4 kHz to 2.4 + 6x kHz. That is, of considering that a pool of 80 selective neurons (which are, however, not strongly coupled) receives an external input coding for the presence of a given stimulus. We found out that both attentional suppression and facilitation decrease gradually when $x$ increases and hence that the presence of external irrelevant stimuli will tend to impair the attentional filtering effect. Note however, that here the effect of presenting just one stimulus is strong because the number of nonselective neurons is small. By presenting one stimulus we give extra input to a large percentage of the nonselective neurons. If the number of nonselective neurons was larger, what would be biologically more plausible, the presentation of a few task irrelevant inputs would cause a minor increase of the total inputs to the nonselective neurons and hence would not strongly impair the attentional filtering effect.

**Discussion**

In this work a network of integrate and fire neurons, characterized by biophysical realistic spiking and synaptic dynamics, was used to model the mechanisms of the visual attention filtering measured in Everling et al. (2002). To our knowledge, this is the first computational model proposed to explain the referred attentional effect. The experimental results show that the presence of a target in an unattended hemifield is not at all signaled in the target-selective neurons of the PFC. The information is totally filtered out by attention. However presenting a target in an attended hemifield will always exert a strong excitation in the target-selective neurons even if they are preferentially activated by stimuli in another hemifield. This strong effect could be explained by the presented model, through competition (implemented by global inhibition) and cooperation (implemented by stronger connections.
between all the target selective neurons) along with the attentional bias that defines the relevant information to the current behaviour.

Earlier models of attention (Rolls & Deco, 2002; Corchs et al., 2003) and plain working memory (Brunel & Wang, 2001) have been operating without cooperative weight settings. However, when working memory becomes selective, which memories are currently stored becomes dependent on the current state of the brain. Hence, context-dependent working memory requires a mechanism of association, which in general allows context to modulate working memory formation. In recent neurodynamical models it turned out, that a more complex structure of the recurrent weights than seen in Brunel and Wang is needed, in order to create selective working memory or rule-dependent working memory (Deco & Rolls, 2003; Deco et al., 2003).

In the present work we systematically explore the effects of a more complex weight setting on the network operation, and find that the most important structural feature of excitatory weight settings is cooperation between pools. Roughly speaking, cooperation allows that activity, which has been gathered by one pool, e.g. through winning a competition, can be propagated to other pools that are associated with it. In this example, the first pool serves as the context, which enables or facilitates the response of the other pools. In general, cooperation can be hypothesized as the basis of categorization (binding pools together to form categories), association between sensations or even mental manipulation. Here we suggest cooperation, besides competition, as a second fundamental principle for the neural basis of cognitive processes in the prefrontal cortex, leading to what we name the ‘Extended Biased Cooperation–Competition Hypothesis’.

A putative aspect of attentional filtering, inattentive blindness, has been studied behaviourally in humans. The neuronal mechanisms underlying this effect have been hypothesized to be the ones experimentally measured by Everling et al. (2002). However, single neuron measurements cannot be directly compared with behavioural performance. In this work, we applied a neurodynamical computational model to capture the neuronal behaviour underlying the mechanisms of visual attentional filtering, as measured in Everling et al. (2002). The network of model spiking neurons used follows biologically realistic dynamics and hence can in principle be used to make predictions concerning human psychophysical experimental results. It can also be extended to allow comparison with neuroimaging results (see Deco et al., 2003). The model can thus provide an ideal theoretical framework to link the electrophysiological measurements with the results from human studies, both measuring performance and brain activity through imaging methods.

In this work we formulate two experimentally testable predictions of our model. One prediction is that an increase in the level of dopamine will lead to a progressive impairment of the attentional filtering effect. In particular, as the dopamine concentration increases, we expect that attentional facilitation is first affected, followed by impairment of attentional suppression. According to these results, in terms of behaviour, an increase in the level of dopamine is expected to impair performance, through interference of the presence of distracting stimuli with the processing of the task-relevant information. The other prediction is that presentation of an increasing number of task irrelevant stimuli will eventually lead to the disappearance of the attentional filtering effect. The observed behaviour of the model suggests that the ability to filter out stimuli with basis on attention will degrade as the number of distracting stimuli increases.

The network model used in this work represents a small area of prefrontal cortex. The PFC is a neocortical structure connecting with all sensory and motor areas and with other cortical and subcortical systems. This connectivity structure makes PFC suitable to coordinate convergent information originating in other brain regions. In fact, neurophysiology, imaging and computational studies have suggested that PFC plays an important role in cognition and behaviour control. It is thought that the wide connectivity that characterizes PFC might determine its crucial role in cognition. However, the focus of our work is in isolating possible neural mechanisms, within the PFC, which underly complex behaviours and hence our model is restricted to a minimal network of PFC neurons. The emphasis is not put in modeling in detail the whole hierarchy of cortical regions involved in the attentional filtering effect. Therefore, in the network presented, the two biases (for attention and identity of the target) are hypothesized to come from neurons or pools of neurons not explicitly modelled. We also do not explicitly model the cortical processing of visual information before it reaches the prefrontal cortex. Instead, we consider that the inputs to the network are already signalling a specific pair of object and location. At a later stage one could model the biases explicitly by having specific pools operating in a bistability regime (i.e. able of persistent activity over a delay period) coding for the present context of the task. In principle, the processing of the visual input by early cortical areas could also be added to the present model, using for instance models based on those presented by Deco and collaborators (Rolls & Deco, 2002).

Given the complexity of PFC and its associated projections is remarkable that we could explain the experimental results using a very simple network model. Note however, that although we reproduce qualitatively the attention filtering effect some quantitative aspects of the results could not be explained by our model. In particular, the results of our simulations could not account for the values of the baseline activity and the strength of the responses. We tried to reproduce these features of the data by modifying the values considered for the conductances. The high firing rates observed in Fig. 3 (2a and b) could be decreased. However this could not be achieved without reducing also the attentional facilitation effect (Fig. 3, 2c) which tended to disappear. We think that to modify in an unproportional way the peak activations a more detailed network model, containing several processing layers, would be required. For example, the interplay between such layers might contribute to the reduction of the peak activity found when targets are presented to both hemifields. This could result from an attentional weakening of the input at an earlier stage, that is driven by top-down attentional biases. The attentional weakening would lead to a weaker input to the target encoding pool with preference for the nonattended location and hence less activity for the cooperating pair of target encoding pools.

The number of neurons used in the network (1000) is relatively small, and was chosen for computational feasibility. The network can be scaled to have larger number of neurons (see Brunel & Wang, 2001), thus reducing finite size effects, and still preserving the qualitative behaviour of the system. The system is robust to changes in the relative sizes of the neuronal pools, provided that a large part of the excitatory neurons are nonselective. This characteristic is important to assure stability of the activity of the network, as well as several important operational regimes identified in the present work. In fact, in biology, for any possible state of the system there is always a large pool of nonselective neurons, corresponding to all neurons not involved in coding the present particular state. Stability of operational modes might then be an important functional role for distributed representation and sparse coding in the brain.

We found that high weights between pools specific for the same object ('cooperative weight setting') successfully implemented cooperation. It might be hypothesized that the same effect could be obtained by divergent inputs, for example target-left neurons would also receive some external input when the target was presented at the right. This divergence of the inputs was not modelled in our network, as
preliminary results showed that it could not implement the cooperation between pools and hence account for the attentional filtering effect. The cooperation effect seems to need the reverberation between the pairs of selective pools.

Networks of model spiking neurons as the one presented here, have been analysed previously in the context of bistability to implement working memory. The mean-field explorations presented here, allowed us to characterize further the network in a number of different overlapping regimes: competition, cooperation, persistent activity and noncompetitive amplification, which in combination form quite different modes of operation. For example nonpersistent activity and competition yield an attentional module, whereas nonpersistent and noncompetition result in a noncompetitive amplification. The grouping of selective pools, effect we name cooperation, together with competition for both bistability and single stability regimes might be a more general mechanism used in the brain to implement computation. The contribution of our work is then twofold: we reveal different modes of operation of the network, that can be used to perform distinct operations, and model the neuronal mechanisms underlying visual attentional filtering.

Acknowledgements

This work was supported by the German Ministry for Research, BMBF Grant 01IBC01A, and through the European Union, grant IST-2001–38099. R.A. was financed by a Marie Curie Individual Fellowship, QLK6-CT-2002–51493. G.D. was supported by Institution Catalana de Recerca i Estudis Avançats (ICREA).

References


Appendix 1

Here we summarize the Leaky Integrate-and-fire model of neurons as described by Brunel & Wang (2001).

The pyramidal cells and interneurons are characterized by the biological realistic parameters (McCormick et al., 1985): resting potential \( V_L = -70 \) mV, firing threshold \( \theta = -50 \) mV, reset potential \( V_{\text{reset}} = -55 \) mV, membrane capacitance \( C_m = 0.5 \) nF for pyramidal cells and 0.2 nF for interneurons, membrane leak conductance \( g_m = 25 \) nS for pyramidal cells and 20 nS for interneurons, refractory period \( \tau_{\text{ref}} = 2 \) ms for pyramidal cells and 1 ms for interneurons, membrane time constant \( \tau_m = 20 \) ms for pyramidal cells and 10 ms for interneurons.

The total synaptic current is given by the sum of the glutamatergic excitatory components (NMDA, AMPA) and the GABA inhibitory component:

\[
I_{\text{syn}}(t) = I_{\text{AMPAn,ext}}(t) + I_{\text{AMPAn,rec}}(t) + I_{\text{NMDAn,rec}}(t) + I_{\text{GABA}}(t)
\]

\[
I_{\text{AMPAn,ext}}(t) = g_{\text{AMPAn,ext}}(V(t) - V_E) \sum_{j=1}^{N_E} w_j s_j^{\text{AMPAn,ext}}(t)
\]

\[
I_{\text{AMPAn,rec}}(t) = g_{\text{AMPAn,rec}}(V(t) - V_E) \sum_{j=1}^{N_E} w_j s_j^{\text{AMPAn,rec}}(t)
\]

\[
I_{\text{NMDAn,rec}}(t) = \frac{g_{\text{NMDAn}}(V(t) - V_E)}{1 + [\text{Mg}^{++}]^c \exp(-0.062 V_E/\sqrt{3})} \sum_{j=1}^{N_E} w_j s_j^{\text{NMDAn}}(t)
\]

\[
I_{\text{GABA}}(t) = g_{\text{GABA}}(V(t) - V_I) \sum_{j=1}^{N_I} s_j^{\text{GABA}}(t)
\]

where \( V_E = 0 \) mV and \( V_L = -70 \) mV and \( w_j \) are the synaptic weights. The voltage dependence of the NMDA currents is controlled by the extracellular magnesium concentration \([\text{Mg}^{++}]^c = 1\) mM.

The values for the synaptic conductances (in nS) for pyramidal neurons are: \( g_{\text{AMPAn,ext}} = 2.08, g_{\text{AMPAn,rec}} = 0.104, g_{\text{NMDAn,rec}} = 0.327 \) and \( g_{\text{GABA}} = 1.25 \) and for interneurons: \( g_{\text{AMPAn,ext}} = 1.62, g_{\text{AMPAn,rec}} = 0.081, g_{\text{NMDAn,rec}} = 0.258 \) and \( g_{\text{GABA}} = 0.973 \), as chosen for the 1000 neurons network in Brunel & Wang’s (2001) simulation.

The fractions of open channels \( s \) are given by:

\[
\frac{d s_j^{\text{AMPAn,ext}}(t)}{dt} = - \frac{s_j^{\text{AMPAn,ext}}(t)}{\tau_{\text{AMPAn,ext}}} + \sum_k \delta(t - t_k)
\]

\[
\frac{d s_j^{\text{AMPAn,rec}}(t)}{dt} = - \frac{s_j^{\text{AMPAn,rec}}(t)}{\tau_{\text{AMPAn,rec}}} + \sum_k \delta(t - t_k)
\]

\[
\frac{d s_j^{\text{NMDAn}}(t)}{dt} = - \frac{s_j^{\text{NMDAn}}(t)}{\tau_{\text{NMDAn}}} + \alpha x_j(t)[1 - s_j^{\text{NMDAn}}(t)]
\]

\[
\frac{dx_j(t)}{dt} = - \frac{x_j(t)}{\tau_{\text{NMDA, rise}}} + \sum_k \delta(t - t_j^k)
\]
\[
\frac{dx_j(t)}{dt} = - \frac{x_j(t)}{\tau_{\text{GABA}}} + \sum_k \delta(t - t_j^k)
\]

where the rise time for NMDA synapses is \(\tau_{\text{NMDA, rise}} = 2\) ms (the rise time of AMPA and GABA is neglected being smaller than 1 ms); the decay time is \(\tau_{\text{NMDA, decay}} = 100\) ms for NMDA synapses, \(\tau_{\text{AMPA}} = 2\) ms for AMPA synapses and \(\tau_{\text{GABA}} = 10\) ms for GABA synapses; \(\alpha = 0.5\) ms\(^{-1}\). The sums over \(k\) represent a sum over spikes formulated as Kronecker \(\delta\)-peaks \([\delta(t)]\) emitted by presynaptic neuron \(j\) at time \(t_j^k\).

**Appendix 2**

The mean-field approximation used in this work was derived by Brunel and Wang (Brunel & Wang, 2001) assuming that the network of neurons is in a stationary state. The potential of the ‘integrate and fire’ neuron is calculated as:

\[
\tau_x \frac{dV(t)}{dt} = -V(t) + \mu_x + \sigma_x \sqrt{\tau_x} \eta(t)
\]

where \(V(t)\) is the membrane potential at time \(t\), \(x\) labels the populations, \(\tau_x\) is the effective membrane time constant, \(\mu_x\) is the mean value the membrane potential would have in the absence of spiking and fluctuations, \(\sigma_x\) measures the magnitude of the fluctuations and \(\eta\) is a Gaussian process with absolute exponentially decaying correlation function with time constant \(\tau_{\text{AMPA}}\).

\[
\mu_x = \left( T_{\text{ext}} v_{\text{ext}} + T_{\text{AMPA}} n_x + (\rho_1 N_x V_E + \rho_2 N_x \langle V \rangle + T_1 w_{x,x} V_1 + V_L \right) / S_x
\]

\[
\sigma_x^2 = \frac{g_{\text{AMPA, ext}}^2 \langle (V) - V_E \rangle^2 N_x v_{\text{ext}} \tau_{\text{AMPA}}^2 \tau_x}{g_m^2 \tau_m^2}
\]

\[
S_x = 1 + T_{\text{ext}} v_{\text{ext}} + T_{\text{AMPA}} n_x + (\rho_1 + \rho_2) N_x + T_1 w_{x,x} v_1
\]

\[
\tau_x = \frac{C_m}{g_m S_x}
\]

where \(n_x\) represents the average firing rate for AMPA components and \(N_x\) represents the average firing rate for NMDA components; \(N_{\text{ext}}\) is the number of connections the network receives from outside.

\[
n_x = \sum_{j=1}^{p} f_j w_{j,x} v_j
\]

\[
N_x = \sum_{j=1}^{p} f_j w_{j,x} \psi(v_j)
\]

\[
\psi(v) = \frac{v \tau_{\text{NMDA}}}{1 + v \tau_{\text{NMDA}}} \left( 1 + \frac{1}{1 + v \tau_{\text{NMDA}}} \sum_{n=1}^{\infty} \frac{(-\alpha \tau_{\text{NMDA, rise}})^n}{(n+1)!} T_n(v) \right)
\]

\[
T_n(v) = \sum_{k=0}^{n} (-1)^k \frac{\gamma^n}{k!} \frac{\tau_{\text{NMDA, rise}} (1 + v \tau_{\text{NMDA}})}{\tau_{\text{NMDA, rise}} (1 + v \tau_{\text{NMDA}}) + k \tau_{\text{NMDA, decay}}}
\]

\[
\tau_{\text{NMDA}} = \alpha \tau_{\text{NMDA, rise}} \tau_{\text{NMDA, decay}}
\]

\(p\) is the number of excitatory pools; \(f_j\) is the fraction of the neurons in the \(x\) excitatory pool; \(w_{j,x}\) is the incoming connection weight from pool \(j\).

The time coefficients for the AMPA, NMDA and GABA values are:

\[
T_{\text{ext}} = \frac{g_{\text{AMPA, ext}} N_{\text{ext}} v_{\text{AMPA}}}{g_m}
\]

\[
T_{\text{AMPA}} = \frac{g_{\text{AMPA, rest}} N_E v_{\text{AMPA}}}{g_m}
\]

\[
\rho_1 = \frac{g_{\text{NMDA}} N_E}{g_m J}
\]

\[
\rho_2 = \beta \frac{g_{\text{NMDA}} N_E (\langle V \rangle - V_E) (J - 1)}{g_m J^2}
\]

\[
J = 1 + \gamma \exp(-\beta \langle V \rangle)
\]

\[
T_1 = \frac{g_{\text{GABA}} N_I v_{\text{GABA}}}{g_m}
\]

where \(N_{\text{ext}}\) is the number of excitatory neurons; \(N_I\) is the number of inhibitory neurons. The discharge rate of the pool as a function of the parameters that define its statistical properties is given by:

\[
\nu_x = \phi(\mu_x, \sigma_x)
\]

where

\[
\phi(\mu_x, \sigma_x) = \left( \tau_p + \tau_x \int_{\beta(\mu_x, \sigma_x)}^{\infty} \sqrt{\pi} \exp(-u^2) \text{erf}(u) \right)^{-1}
\]

\[
\alpha(\mu_x, \sigma_x) = \left( \frac{\nu_{\text{rest}} - \mu_x}{\sigma_x} \right) \left( 1 + 0.5 \frac{\tau_{\text{AMPA}}}{\tau_x} \right) + 1.03 \sqrt{\frac{\tau_{\text{AMPA}}}{\tau_x}}
\]

\[
\beta(\mu_x, \sigma_x) = \frac{\nu_{\text{reset}} - \mu_x}{\sigma_x}
\]

where \(\text{erf}(u)\) is the error function.

The average membrane potential \(\langle V \rangle\) is given by

\[
\langle V \rangle = \mu_x - (\nu_{\text{rest}} - \nu_{\text{reset}}) \nu_x \tau_x
\]

should always stay between \(-55\) mV and \(-50\) mV and be initialized with a value in this range.