

The role of attention in visual perception: a computational neuroscience model

Gustavo Deco(*) and Silvia Corchs(**)

(*) Institució Catalana de Recerca i Estudis Avançats (ICREA),
Universitat Pompeu Fabra, Dept. of Technology, Computational Neuroscience,
Passeig de Circumval.lació, 8 - 08003 Barcelona, Spain

(**) Siemens AG, Corporate Technology Information Communications 4,
Otto-Hahn-Ring 6, D-81739 Munich, Germany

Abstract

We review different functions involved in visual perception that have been integrated by a model based on the biased competition hypothesis. Attentional top-down bias guides the dynamics to concentrate at a given spatial location or on given features. The model integrates, in a unifying form, the explanation of several existing types of experimental data obtained at different levels of investigation. At the microscopic level, single cell recordings are simulated. At the mesoscopic level of cortical areas, results of functional magnetic resonance imaging (fMRI) studies are reproduced. Finally, at the macroscopic level, psychophysical experiments like visual search tasks are also described by the model.

Keywords: visual attention, computational neuroscience, biased competition, theoretical model

1 Introduction

To understand how the brain works, including how it functions in vision it is necessary to combine different approaches, including neural computation. Neurophysiology at the single neuron level is needed because this is the level at which information is exchanged between the computing elements of the brain. Evidence from neuropsychology is needed to help understand what different parts of the system do and what each part is necessary for. Neuroimaging is useful to indicate where in the human brain different processes take place, and to show which functions can be dissociated from each other. Knowledge of the biophysical and synaptic properties of neurons is essential to understand how the computing elements of the brain work, and therefore what the building blocks of biologically realistic computational models should be. Knowledge of the anatomical and functional architecture of the cortex is needed

to show what types of neuronal network actually perform the computation. And finally the approach of neural computation is needed, as this is required to link together all the empirical evidence to produce an understanding of how the system actually works. This review utilizes evidence from some of these disciplines to develop an understanding of how vision is implemented by processing in the brain, focusing on visual attentional mechanisms.

The dominant neurobiological hypothesis to account for attentional selection is that attention serves to enhance the responses of neurons representing stimuli at a single relevant location in the visual field. This enhancement model is related to the metaphor for focal attention in terms of a spotlight [1, 2]. This metaphor postulates a spotlight of attention which illuminates a portion of the field of view where stimuli are processed in higher detail while the information outside the spotlight is filtered out. According to this classical view, a relevant object in a cluttered scene is found by rapidly shifting the spotlight from one object in the scene to the next one, until the target is found. Therefore, according to this assumption the concept of attention is based on explicit serial mechanisms.

There exists an alternative mechanism for selective attention, the *biased competition* model [3, 4, 5, 6]. According to this model, the enhancement of attention on neuronal responses is understood in the context of competition among all of the stimuli in the visual field. The *biased competition* hypothesis states that the multiple stimuli in the visual field activate populations of neurons that engage in competitive mechanisms. Attending to a stimulus at a particular location or with a particular feature biases this competition in favor of neurons that respond to the location or the features of the attended stimulus. This attentional effect is produced by generating sig-

nals within areas outside the visual cortex which are then fed back to extrastriate areas, where they bias the competition such that when multiple stimuli appear in the visual field, the cells representing the attended stimulus "win", thereby suppressing cells representing distracting stimuli. According to this line of work, attention appears as an emergent property of competitive interactions that work in parallel across the visual field.

Visual attention can function in two distinct modes: spatial focal attention which can be visualized as a spotlight that *illuminates* a certain location of visual space for focused visual analysis, and object attention which is spatially dispersed and with which a target object can be searched for in parallel over a large area of visual space. Duncan [4] proposed that both modes of operation are manifestations of a top-down selection process. In spatial attention, the selection is focused in the spatial dimension and spread in the feature dimension. In object attention, the selection is focused in the feature dimension and spread in the spatial attention.

In the following section we review a number of experimental studies that provide insights into the neural basis of attention. In section 3 we present a neurodynamical model that addresses the issues of attentional mechanisms and in section 4 we show that the model can reproduce the findings of a number of attention-related neurophysiological experiments as well as the results of psychophysical experiments.

2 Experimental studies

2.1 Single cell experiments

Reynolds et al. [7] first examined the presence of competitive interactions in the absence of attentional effects, making the monkey attend to a location far outside the receptive field of the neuron they were recording. They compared the firing activity response of the neuron when a single reference stimulus was located within the receptive field to the response when a probe stimulus was added to the visual field. When the probe was added to the field, the activity of the neuron was shifted towards the activity level that would have been evoked had the probe appeared alone. When the reference is an effective stimulus (high response) and the probe is an ineffective stimulus (low response) the firing activity is suppressed after adding the probe. In contrast, the response of the cell increased when an effective probe stimulus was added to an ineffective reference stimulus. The study also tested attentional modulatory

effects independently by repeating the same experiment with the difference that the monkey attended to the reference stimulus within the receptive field of the recorded neuron. The effect of the attention on the response of the V2 neuron was to almost compensate the suppressive or excitatory effect of the probe. That is, if the probe caused a suppression of the activity response to the reference when the attention was outside the receptive field, then attending to the reference restored the neuron's activity to the level corresponding to the case of the reference stimulus alone. Similarly, if the probe stimulus had increased the neuron's level of activity, attending to the reference stimulus compensates the response by shifting the activity to the level that had been recorded when the reference was presented alone.

2.2 fMRI experiments

The experimental studies of Kastner et al. [8, 9] show 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. The authors also observed that directing attention to one of the stimuli counteracts the suppressive influence of nearby stimuli. These experimental results were obtained by applying the functional magnetic resonance imaging (fMRI) technique in humans. The authors designed an experiment and different conditions were examined. In the first experimental condition the authors tested the presence of suppressive interactions among stimuli presented simultaneously within the visual field in the absence of directed attention, in the second experimental condition they investigated the influence of spatially directed attention on the suppressive interactions, and in the third condition they analyzed the neural activity during directed attention but in the absence of visual stimulation. The authors observed that, because of the mutual suppression induced by competitively interacting stimuli, the fMRI signals were smaller during the simultaneous presentations than during the sequential presentations. In the second part of the experiment there were two main factors: presentation condition (sequential versus simultaneous) and directed attention condition (unattended versus attended). The average fMRI signals with attention increased more strongly for simultaneously presented stimuli than the corresponding signals for sequentially presented stimuli. Thus, the suppressive interactions were partially cancelled out by attention.

2.3 Psychophysical experiments: visual search

We now concentrate on the macroscopic level of psychophysics. Evidence for different temporal behaviours of attention in visual processing come 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. All other items in the display which are different from the target serve the role as distractors. The relevant variable typically measured is search time as a function of the number of items in the display. Much work has been based on two kinds of search paradigm: feature search, and conjunction search. 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, implying a serial process. On the other hand, search times in a feature search can be independent of the number of items in the display.

Quinlan and Humphreys [10] analyzed feature search and three different kinds of conjunction search, namely: standard conjunction search and two kinds of triple conjunction with the target differing from all distractors in one or two features respectively. Let us define the different kinds of search tasks by using a pair of numbers m and n , where m is the number of distinguishing feature dimensions between target and distractors, and n is the number of features by which each distractor group differs from the target. Using this terminology, feature search corresponds to a 1,1-search; a standard conjunction search corresponds to a 2,1-search; a triple conjunction search can be a 3,1 or a 3,2-search depending of whether the target differs from all distractor groups by one or two features respectively.

Quinlan and Humphreys [10] showed that in feature search (1,1), the target is detected in parallel across the visual field. They also show that the reaction time in both standard conjunction search and triple conjunction search conditions is a linear function of the display size. The slope of the function for the triple conjunction search task can be steeper or relatively flat, depending upon whether the target differs from the distractors in one (3,1) or two features (3,2), respectively.

3 The neurodynamical model

The overall systemic representation of the model is shown in Fig.1. The system is essentially composed of six modules (V1, V2-V4, IT, PP, v46, d46), structured such that they resemble the two known main visual paths of the mammalian visual cortex: the *what* and *where* paths [11, 12, 13]. These six modules represent the minimum number of components to be taken into account within this complex system in order to describe the desired visual attention mechanism.

Information from the retino-geniculo-striate pathway enters the visual cortex through areas V1-V2 in the occipital lobe and proceeds into two processing streams. The occipital-temporal stream (*what* pathway) leads ventrally through V4 and IT (inferotemporal cortex) and is mainly concerned with object recognition, independently of position and scaling. The occipito-parietal stream (*where* pathway) leads dorsally into PP (posterior parietal) and is concerned with the location of objects and the spatial relationships between objects. The model considers that feature attention biases intermodular competition between V4 and IT, whereas spatial attention biases intermodular competition between V1, V4 and PP.

The ventral stream consists of four modules: V1, V2-V4, IT, and a module v46 corresponding to the ventral area 46 of the prefrontal cortex, which maintains the short-term memory of the recognized object or generates the target object in a visual search task. The module V1 is concerned with the extraction of simple features (for example bars at different locations, orientations and size). It consists of pools of neurons with Gabor receptive fields tuned at different positions in the visual field, orientations and spatial frequency resolutions. The V1 module contains $P \times P$ hypercolumns that cover the $N \times N$ pixel scene. Each hypercolumn contains L orientation columns of complex cells with K octave levels corresponding to different spatial frequencies. This V1 module inputs spatial and feature information up to the dorsal and ventral streams. Also, there is one inhibitory pool interacting with the complex cells of all orientations at each scale. The inhibitory pool integrates information from all the excitatory pools within the module and feedbacks unspecific inhibition uniformly to each of the excitatory pools. It mediates normalizing lateral inhibition or competitive interactions among the excitatory cell pools within the module.

The module IT is concerned with the recognition of objects and consists of pools of neurons which are

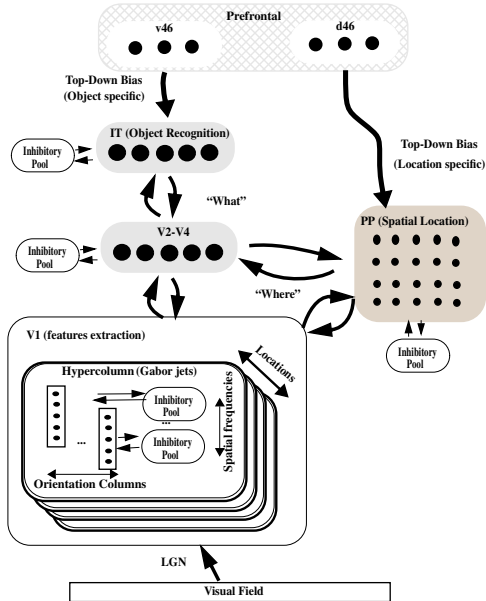


Figure 1: Architecture of the neurodynamical approach. The system is essentially composed of six modules structured such that they resemble the two known main visual paths of the visual cortex.

sensitive to the presence of a specific object in the visual field. It contains C pools, as the network is trained to search for or recognize C particular objects. The V2-V4 module serves primarily to pool and channel the responses of V1 neurons to IT to achieve a limited degree of translation invariance. It also mediates a certain degree of localized competitive interaction between different targets. A lattice is used to represent the V2-V4 module. Each node in this lattice has $L \times K$ assemblies as in a hypercolumn in V1. Each cell assembly receives convergent inputs from the cell assemblies of the same tuning from an $M \times M$ hypercolumn neighborhood in V1. The feedforward connections from V1 to V2-V4 are modeled with convergent Gaussian weight function, with symmetric recurrent connection.

The dorsal stream consists of three modules: V1, PP and d46. The module PP consists of pools coding the position of the stimuli. It is responsible for mediating spatial attention modulation and for updating the spatial position of the attended object. A lattice of $N \times N$ nodes represents the topographical organization of the module PP. Each node on the lattice corresponds to the spatial position of each pixel in the input image. The module d46 corresponds to the dorsal area 46 of the prefrontal cortex that maintains the short term spatial memory or generates the

attentional bias for spatial location.

The prefrontal areas 46 (modules v46 and d46) are not explicitly modeled. Feedback connections between these areas provide the external top-down bias that specifies the task. The feedback connection from area v46 to the IT module specifies the target object in a visual search task. The feedback connection from area d46 to the PP module generates the bias to a targeted spatial location.

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 learning during several presentations of a specific object. During the recognition mode there are two possibilities of running the system. First, an object can be localised in a scene (visual search) by biasing the system with an external top-down component at the IT module which drives the competition in favour of the pool associated with the specific object to be searched. Then, the intermodular attentional modulation V1-V4-IT will enhance the activity of the pools in V4 and V1 associated with the features of the specific object to be searched. Finally, the intermodular attentional modulation V4-PP and V1-PP will drive the competition in favour of the pool localising the specific object. Second, an object can be identified (object recognition) at a specific spatial location by biasing the system with an external top-down component at the PP module. This drives the competition in favour of the pool associated with the specific location such that the intermodular attentional modulation V4-PP and V1-PP will favour the pools in V1 and V4 associated with the features of the object at that location. Intermodular attentional modulation V1-V4-IT will favour the pool that recognized the object at that location.

Each pool of neurons will be described within the mean field approximation [14, 15, 16, 17] which consists of replacing the temporal averaged discharge rate of a cell with an equivalent activity of a neural population (ensemble average). The mathematical formulation of our model is given in the following section.

3.1 Mathematical formulation of the model

We consider a pixelized grey-scaled image given by a $N \times N$ matrix Γ_{ij}^{orig} . The subindex ij denotes the spatial position of the pixel. Each pixel value is given a grey value coded between 0 (black) and 255 (white). The first step in the preprocessing consists in removing the DC component of the image

which is probably done in the lateral geniculate nucleus (LGN) of the thalamus. We denote this LGN representation by the $n \times n$ matrix Γ_{ij} . In our experiments the images were pixelized in a 66×66 matrix ($N = 66$) for the single cell recordings simulations and in a 64×64 matrix for the fMRI simulations. Feedforward connections to a layer in V1 perform the extraction of simple features. Simple cells in the primary visual cortex are modeled by 2D-Gabor functions. The Gabor receptive fields have five degrees of freedom (2D-location of the receptive field's center, size of the receptive field, orientation and symmetry) and are given by the product of an elliptical Gaussian and a complex plane wave.

The neurons in the pools in V1 have receptive fields performing a Gabor wavelet transform. Let us denote by $I_{k p q l}^{V,E}$ the sensorial input activity to a pool in V1 which is sensitive to a determined spatial frequency given at octave k , to a preferred orientation defined by the rotation index l and to stimuli at the center location specified by the indices $p q$. The sensorial input activity to a pool in V1 is therefore defined by the module of the convolution between the corresponding receptive fields and the image. The large receptive fields of V2 and V4 can be approximately taken into account by including in V1 pools with receptive fields corresponding to several octaves of the 2D-Gabor wavelet transform (i.e. not only the typical narrow receptive fields of V1 but also larger receptive fields are included in V1). Therefore, for the fMRI simulations we compact the V1 and V2-V4 modules in just one module, namely V1. The reduced system connects all cell assemblies in V1 with all cell assemblies in IT. However, in order to simulate Reynolds *et al.*' experiment and to be able to compare with their data we include a V2 pool that is directly connected to V1.

Let us now define the neurodynamical equations that regulate the evolution of the whole system. The activity level of the input current in the V1 module is given by

$$\begin{aligned} \tau \frac{\partial}{\partial t} I_{k p q l}^{V1}(t) &= -I_{k p q l}^{V1}(t) + aF(I_{k p q l}^{V1}(t)) \\ &- bF(I_k^{V1,I}(t)) + I_{k p q l}^{V1,E}(t) I_{p q}^{V1-PP}(t) \\ &+ I_{k p q l}^{V1-IT}(t) + I_0 + \nu \end{aligned} \quad (1)$$

where the attentional biasing due to the intermodular *where* connections with the pools in the PP module, $I_{p q}^{V1-PP}$, is given by

$$I_{p q}^{V1-PP} = \sum_{i,j} W_{p q i j} F(I_{i j}^{PP}(t)) \quad (2)$$

and the coefficients $W_{p q i j}$ are evaluated as

$$W_{p q i j} = K_{V1PP} \cdot \exp[-dist^2/2S^2] \quad (3)$$

with K_{V1PP} being the coupling constant between both modules, $S = 2$ and $dist$ represents the distance from spatial localization (i, j) to the position of the receptive field (p, q) .

Also, in eq. 1 I_0 is a diffuse spontaneous background input, ν is the Gaussian noise and $I_{k p q l}^{V1-IT}$ represents the attentional biasing due to the intermodular *what* connections with the pools in the temporal module IT and is defined by

$$I_{k p q l}^{V1-IT}(t) = \sum_{c=1}^{c=C} w_{c k p q l} F(I_c^{IT}(t)) \quad (4)$$

where $w_{c k p q l}$ is the connection strength between pools V1 and IT, corresponding to the coding of a specific object category c . We assume that the IT module has C pools corresponding to different object categories.

Excitatory cell pools in each module are engaged in competition, mediated by an inhibitory pool which receives excitatory input from all the excitatory pools and provides uniform inhibitory feedback to each of the excitatory pools. The current activity of the inhibitory pools $I_k^{V1,I}$ obey the following equations

$$\begin{aligned} \tau \frac{\partial}{\partial t} I_k^{V1,I}(t) &= -I_k^{V1,I}(t) + c' \sum_{p,q,l} F(I_{k p q l}^{V1,I}(t)) \\ &- dF(I_k^{V1,I}(t)) \end{aligned} \quad (5)$$

Similarly, the current activity of the excitatory pools in the PP module are given by

$$\begin{aligned} \tau \frac{\partial}{\partial t} I_{i j}^{PP}(t) &= -I_{i j}^{PP}(t) + aF(I_{i j}^{PP}(t)) \\ &- bF(I^{PP,I}(t)) + I_{i j}^{PP-V1}(t) \\ &+ I_{i j}^{PP,A} + I_0 + \nu \end{aligned} \quad (6)$$

where $I_{i j}^{PP,A}$ is an external attentional spatial-specific top-down bias, the intermodular attentional biasing $I_{i j}^{PP-V1}$ through the connections with the pools in the module V1 is

$$I_{i j}^{PP-V1}(t) = \sum_{k,p,q,l} W_{p q i j} F(I_{k p q l}^{V1}(t)) \quad (7)$$

and the activity current of the common PP inhibitory pool evolves according to

$$\begin{aligned} \tau \frac{\partial}{\partial t} I^{PP,I}(t) &= -I^{PP,I}(t) + c' \sum_{i,j} F(I_{i j}^{PP}(t)) \\ &- dF(I^{PP,I}(t)) \end{aligned} \quad (8)$$

The dynamics of the inferotemporal module IT is given by

$$\begin{aligned} \tau \frac{\partial}{\partial t} I_c^{IT}(t) &= -I_c^{IT}(t) + aF(I_c^{IT}(t)) \\ &\quad - bF(I^{IT,I}(t)) + I_c^{IT-V1}(t) \\ &\quad + I_c^{IT,E} + I_0 + \nu \end{aligned} \quad (9)$$

where $I_c^{IT,E}$ denotes an external attentional spatial-specific top-down bias and the intermodular attentional biasing between IT and V1 pools is

$$I_c^{IT-V1}(t) = \sum_{k,p,q,l} w_{ckpql} F(I_{kpl}^{V1})(t) \quad (10)$$

where the weights w_{ckpql} are trained by Hebbian learning.

Finally, the activity current of the common PP inhibitory pool evolves according to

$$\begin{aligned} \tau \frac{\partial}{\partial t} I^{IT,I}(t) &= -I^{IT,I}(t) + c' \sum_c F(I_c^{IT}(t)) \\ &\quad - dF(I^{IT,I}(t)) \end{aligned} \quad (11)$$

4 The numerical simulations

4.1 Simulation of the single-cell experiments

In this section we present the simulations [19, 20] corresponding to the experiments by Reynolds *et al.* on single cell recording in V2 neurons in monkeys. We study the dynamical behavior of the cortical architecture presented in the previous section by numerically solving the system of coupled differential equations in a computer simulation. We introduce for this experiment a module of V2 neurons. The input system processed an image of 66 x 66 pixels (N=66). The V1 hypercolumns covered the entire image uniformly. They were distributed in 33 x 33 locations (P=33) and each hypercolumn was sensitive to two spatial frequencies and to eight different orientations (K=2 and L=8). The V2 module has 2 x 8 pools receiving convergent input from the pools of the same tuning from a 10 x 10 (i.e., M=10) hypercolumn neighborhood in V1. The feedforward connection from V1 to V2 are modeled with a convergent Gaussian weight function, having a symmetric recurrent connection. We analyzed the firing activity of a single pool in the V2 module which was highly sensitive to a vertical bar presented in its receptive field (effective stimulus) and poorly sensitive to a 75 degrees oriented bar presented in its receptive field (ineffective stimulus). The size of the bars were 2

x 4 pixels. Following the experimental setup of the work by Reynolds *et al.*, we calculate the evolution of the firing activity of a V2 pool under four different conditions: 1) single reference stimulus within the receptive field; 2) single probe stimulus within the receptive field; 3) reference and probe stimulus within the receptive field without attention; 4) reference and probe stimuli within the receptive field and attention directed to the spatial location of the reference stimulus. In the simulations, the attention was directed to the reference location by setting the top-down attentional bias $I_{ij}^{PP,A}$ in PP equal to 0.07 if i and j correspond to the location of the reference stimulus and to zero otherwise. In the unattended condition, the external top-down bias was set equal to zero everywhere. Comparing with the experimental results, the same qualitative behavior is observed for all experimental conditions analyzed. The competitive interactions in the absence of attention are due to the intramodular competitive dynamics at the level of V1 (i.e. the suppressive and excitatory effects of the probe). The modulatory biasing corrections in the attended condition are caused by the intermodular interactions between V1 and PP pools, and PP pools and prefrontal top-down modulation.

4.2 Simulations of the fMRI data

The dynamical evolution of activity at the cortical area level, as evidenced in the behaviour of fMRI signals in experiments with humans, can be simulated in the framework of the present model by integrating the pool activity in a given area over space and time. The integration over space yields an average activity of the considered brain area at a given time. With respect to the integration over time, it is performed in order to simulate the temporal resolution of fMRI experiments. In this section we simulate fMRI signals from V4 under the experimental conditions defined by Kastner *et al.* [9]. We use the same parameters as in the last section but the V1 hypercolumns include now three levels of spatial resolution ($K = 3$). Let us remark that for this particular case, the IT module as well as the v46 area are not explicitly needed for the computational simulations.

In order to simulate the data by Kastner *et al.*, we also use four complex images similar to the ones these authors used in their work. These images were presented as input images in four nearby locations in the upper right quadrant. The neurodynamics is solved through an interactive process where we choose 200 iterations to represent 1s (that corresponds to the time resolution of the fMRI measurements).

Stimuli were shown in the two above mentioned

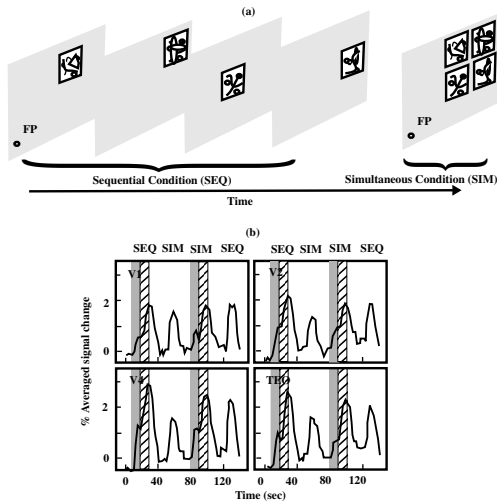


Figure 2: (a) Experimental design of Kastner et al [9]. (b) Computer simulations of fMRI signals in visual cortex. Grey shade areas indicate the expectation period, striped areas the attended presentations and blocks without shading correspond to unattended condition.

conditions: sequential and simultaneous (see Fig 2a). In the SEQ condition, stimuli were presented separately in one of the four locations for 250 ms. In the SIM condition, the four stimuli appeared simultaneously for 250 ms and with equal blank intervals between each other. The order of the stimuli and location was randomized. Two attentional conditions were simulated: an unattended condition, during which no external top-down bias from prefrontal areas was present (i.e. $I_{ij}^{PP,A}$ is zero everywhere) and an attended condition that was defined 10s before the onset of visual presentations (expectation period EXP) and continued during the subsequent 10s block.

The attended condition was implemented by setting $I_{ij}^{PP,A}$ equal to 0.07 for the locations associated with the lowest left stimulus and zero elsewhere. Fig.2b shows the results of the computational simulations [20] for a sequential simulation block: BLK-EXP-SEQ(attended)-BLK-SIM-BLK-EXP-SIM(attended). As in the experiments of Kastner *et al.* [9], these simulations show that the fMRI signals were smaller in magnitude during the SIM than during the SEQ presentations in the unattended conditions because of the mutual suppression induced by competitively interacting stimuli. On the other hand, the average fMRI signals with attention increased more strongly for simultaneously presented

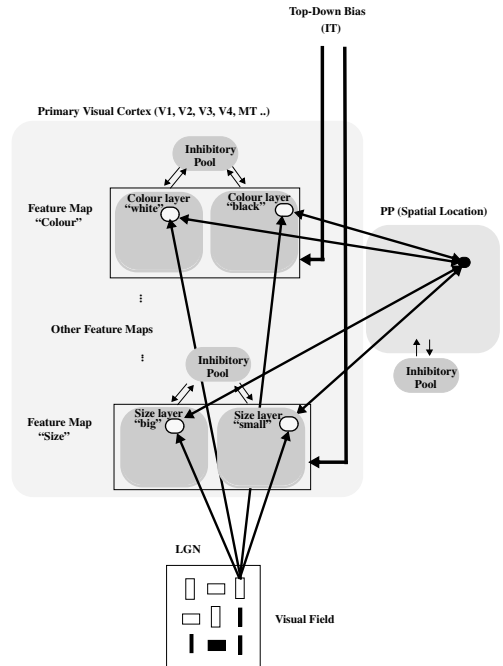


Figure 3: Extended cortical architecture for visual attention and binding of multiple feature components (shape, colour, etc.).

stimuli than the corresponding ones for sequentially presented stimuli. Thus, the suppressive interactions were partially cancelled out by attention. Finally, during the expectation period activity increased in the absence of visual presentations and further increased after the onset of visual stimuli. The theoretical data describe quite well the qualitative behaviour of the experiments.

4.3 Simulation of visual search tasks

Deco and Zihl [13] and Rolls and Deco [18] extended the neurodynamical model to account for the different slopes observed experimentally in complex conjunction visual search tasks. Figure 3 shows the overall architecture of the extended model. The input retina is given as a matrix of visual items. The location of each item on the retina is specified by two indices (ij), describing the position in row i and column j . The dimension of this matrix is $S \times S$, i.e. the number of items in the display is S^2 . Information is processed across the different spatial locations in parallel. The authors assume that selective attention results from independent competition mechanisms operating within each feature dimension. Each visual item can be defined by M features. Each feature m can adopt $N(m)$ values, for example the feature

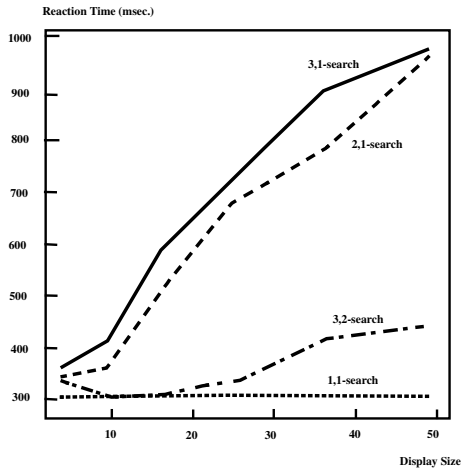


Figure 4: Search times for feature and conjunction searches obtained utilizing the extended computational cortical model.

colour can have the values *black* or *white* (in this case $N(\text{colour}) = 2$). For each feature map m , there are $N(m)$ layers of neurons characterizing the presence of each feature value. A cell assembly consisting of a population of fully connected excitatory integrate-and-fire spiking neurons is allocated to every location in each layer, in order to encode the presence of a specific feature value (e.g. *colour white*) at the corresponding feature position. The feature maps are topographically ordered; i.e. the receptive fields of the neurons belonging to cell assembly ij in one of these maps are limited to the location ij in the retinal input. It is also assumed that the cell assemblies in layers corresponding to one feature dimension are mutually inhibitory. The posterior parietal module is bidirectionally coupled with the different feature maps and serves to bind the different feature dimensions at each item location, in order to implement local conjunction detectors. The inferotemporal connections provide top-down information, consisting of the feature values for each feature dimension of the target item. This information is fed into the system by including an extra excitatory input to the corresponding feature layer. For example, if the target is defined as small, vertical and black, then all excitatory pools at each location in the layer coding *small* in the feature map dimension *size*, in the layer coding *vertical* in the feature map *orientation* and in the layer coding *black* in the feature map *colour*, receive an extra excitatory input from the IT module.

In Fig. 4, the computational results obtained by Deco and Zihl [13] for 1,1; 2,1; 3,1 and 3,2-searches are presented. The items are defined by three feature dimensions ($M = 3$, e.g. size, orientation and

colour), each having two values ($N(m) = 2$ for $m = 1,2,3$, e.g. size: big/small, orientation: horizontal/vertical, colour: white/black). For each display size, the experiment is repeated 100 times, each time with different randomly generated targets at random positions and randomly generated distractors. The mean value T of the 100 simulated search times is plotted as a function of the display size S . The slopes for all simulations are consistent with existing experimental results [10]

5 Conclusions

We followed a computational neuroscience approach in order to study the role of attention in visual perception. The aim of this review was to attempt to provide a mathematical formulation that unifies microscopic, mesoscopic and macroscopic mechanisms involved in the brain functions, allowing the description of the existing experimental data (and the prediction of new results as well) at all neuroscience levels (psychophysics, functional brain imaging and single neural cells measurements).

We have focused on the analysis of the microscopic neurodynamical mechanisms that underlie visual attention. We presented a computational system that consists of interconnected populations of cortical neurons distributed in different brain modules which in turn can be related to the different areas of the dorsal and ventral paths of the cortex. Competitive mechanisms were implemented by connecting the pools of a given module with a common inhibitory pool. In this way, the more pools of the module are active, the more active will the common inhibitory pool be and, consequently the pools in the module will experience more feedback inhibition, such that only the most excited group of pools will survive and win the competition. On the other hand, external top-down bias could shift the competition in favor of a specific group of pools. Therefore, this basic computational model implements the biased competition hypothesis. Taking into account the computational role of individual brain areas and their mutual interactions, the macroscopic phenomenological behavior arises as the result of the global dynamical interactions between the different modules.

In summary, computational neuroscience provides a mathematical framework for studying the mechanisms involved in brain function, like visual attentional mechanisms, that we have reviewed in the present work. The neurodynamical model here analyzed is based on evidence from functional, neurophysiological and psychological findings. The sim-

ulations obtained with this theoretical model successfully reproduce the experimental results of neurophysiological and fMRI experiments on spatial attention, as well as studies on serial and parallel search.

References

- [1] Treisman, A. Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8:194–214, 1982.
- [2] Treisman, A. Features and objects: The fourteenth Barlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, 40A:201–237, 1988.
- [3] Desimone, R., and J. Duncan. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18:193–222, 1995.
- [4] Duncan, J. The locus of interference in the perception of simultaneous stimuli. *Psychological Review* 87:272–300, 1980.
- [5] Duncan, J. Cooperating brain systems in selective perception and action. In *Attention and performance XVI*, T. Inui and J.L. McClelland (Eds.), Cambridge: MIT Press, 549–578. 1996.
- [6] Duncan, J., and G. Humphreys. Visual search and stimulus similarity. *Psychological Review* 96:433–458, 1989.
- [7] Reynolds, J., Chelazzi, L., and R. Desimone. Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, 19:1736–1753, 1999.
- [8] Kastner, S., De Weerd, P., Desimone, R., and L. Ungerleider. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, 282:108–111, 1998.
- [9] Kastner, S., Pinsk, M., De Weerd, P., Desimone, R., and L. Ungerleider. Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, 22:751–761, 1999.
- [10] Quinlan, P. T. and G. W. Humphreys. Visual search for targets defined by combination of color, shape and size: An examination of the task constraints on feature and conjunction searches. *Perception and Psychophysics*, 41:455–472, 1987.
- [11] Deco, G., and J. Zihl. A neural model of binding and selective attention for visual search. Proc. of the 5th Neural Computation and Psychology Workshop (NCPW'98). Hrgs von D. Heinke, G. W. Humphreys, A. Olson. University of Birmingham, England, September 1998. London, Springer Verlag, 262–271, 1998.
- [12] Hamker, F. The role of feedback connections in task-driven visual search. Proc. of the 5th Neural Computation and Psychology Workshop (NCPW'98). Hrgs von D. Heinke, G. W. Humphreys, A. Olson. University of Birmingham, England, September 1998. London: Springer Verlag 1999, 252–261, 1998.
- [13] Deco, G., and J. Zihl. Top-down selective visual attention, a neurodynamical approach. *Visual Cognition* 8:118–139, 2001.
- [14] Wilson, H., and S. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biological Cybernetics*, 12:1–24, 1972.
- [15] Amit, D., and M. Tsodyks. Quantitative study of attractor neural network retrieving at low spike rates: I. Substrate spikes, rates and neuronal gain. *Network* 2, 259–273, 1991.
- [16] Abbot, L. Firing rate models for neural populations. *Neural Networks: From Biology to High Energy Physics*. Pisa: ETS Editrice. In O. Benhar, C. Bosio, P. Giudice and E. Tabet (Eds.), 1992.
- [17] Usher, M., and E. Niebur. Modeling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience*, 8:311–327, 1996.
- [18] Rolls, T., and G. Deco. Computational Neuroscience of Vision. *Oxford University Press Inc. New York*, 2002.
- [19] Deco, G., and T.S. Lee. An unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing* 44-46:769–774, 2002.
- [20] Corchs, S., and G. Deco. Large-scale Neural Model for Visual Attention: Integration of Experimental Single-cell and fMRI Data. *Cerebral Cortex* 12:339–348, 2002.