Integrating fMRI and Single-Cell Data of Visual Working Memory

Gustavo Deco[a] and Edmund T. Rolls[b] and Barry Horwitz[c]

[a] *Institució Catalana de Recerca i Estudis Avançats, Universitat Pompeu Fabra, Barcelona, Spain*

[b] *University of Oxford, Dept. of Experimental Psychology, Oxford, England*

[c] *National Institutes of Health, Bethesda, USA*

Based on single-cell and fMRI data, two different current models of the topographical and functional organization of the prefrontal cortex has been proposed: organization-by-stimulus-domain model, and organization-by-process model. The present work utilizes a computational neuroscience based approach in order to integrate, via a detailed large-scale microscopic neurodynamical computational model, single-cells and fMRI measurements of the prefrontal cortex (PFC) associated with working memory processing, in order to delve into the question on the topographical structure of the prefrontal cortex. For this purpose, we formulate an explicit model of the spiking and synaptic dynamical mechanisms that underly working memory-related activity during the execution of delay tasks with a 'what'-then-'where' design (object and spatial delayed response within the same trial).

1. Introduction

Cognitive neuroscience is mainly intended to establish a link between cognitive functions and the underlying neural substrate. In the case of the frontal lobe, experimental cognitive neuroscience methods suggested that at least some types of working memory and related processes such as planning are strongly related with the prefrontal cortex (PFC). Empirical observations have let to two different hypothesis of the topographical and functional organization of the prefrontal cortex (see [9] for a review), namely: a 'what'/'where' mnemonic oriented organization of the PFC (i.e. organization-by-stimulus-domain), and a higher order functional (e.g. maintenance and manipulation) non-mnemonic oriented organization of the PFC (i.e. organization-by-process). The first 'what'/'where' mnemonic oriented 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. Event-related MRI studies in humans [8] and single-cell data in primates [6] support this specific organization-by-stimulus-domain. The 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 whereby the mnemonic maintenance functions are allocated to inferior prefrontal areas. Event related fMRI studies [10] and single-cell data in primates [11] with a 'what'-then-'where' design provide evidence against the organization-by-stimulus-domain hypothesis.

The topographical microscopic and functional macroscopic organization of the PFC can be better understood if different kinds of cognitive neuroscience data taken at different levels, are unified and integrated via a computational neuroscience approach. The approach of computational neuroscience leads to a precise definition of how the computation is performed, and to precise and quantitative tests of the theories produced. The present work is a review of a model [3] we proposed to help to understand 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, object and spatial delayed response within the same trial.

2. Neurodynamical Model of the Prefrontal Cortex

We begin by describing the basic neuronal units comprising the excitatory and inhibitory neurons of our model of prefrontal cortex [1]. The basic circuit is an integrate-and-fire model, that consists of the cell membrane capacitance C_m in parallel with the cell membrane resistance R_m driven by a synaptic current (excitatory or inhibitory post-synaptic potential, EPSP or IPSP, respectively). If the voltage across the capacitor reaches a threshold θ the circuit is shunted and a δ-pulse (spike) is generated and transmitted to other neurons. We use biologically realistic parameters. We assume for both kinds of neuron a resting potential $V_L = -70$ mV, a firing threshold $\theta = -50$ mV, and a reset potential $V_{\rm reset} = -55$ mV. The membrane capacitance $C_{\rm m}$ is 0.5 nF for the excitatory pyramidal neurons and 0.2 nF for the inhibitory interneurons. The membrane leak conductance $g_{\rm m}$ is 25 nS for pyramidal cells, and 20 nS for interneurons. The refractory period $\tau_{\rm ref}$ is 2 ms for pyramidal cells, and 1 ms for interneurons. More specific, the subthreshold membrane potential $V(t)$ of each neuron evolves according following the equation:

$$C_{\rm m}\frac{dV(t)}{dt} = -g_{\rm m}(V(t) - V_L) - I_{\rm syn}(t) \quad (1)$$

where $I_{\rm syn}(t)$ is the total synaptic current flow into the cell.

The synaptic current flows into the cells are mediated by three different families of receptors. The total synaptic current is given by the sum of glutamatergic excitatory components (NMDA and AMPA) and inhibitory components (GABA, $I_{\rm G}$). We consider that external excitatory contributions are produced through AMPA receptors ($I_{\rm Ae}$), while the excitatory recurrent synapses are produced through AMPA and NMDA receptors ($I_{\rm Ar}$ and $I_{\rm Nr}$). The total synaptic current is therefore given by:

$$I_{\rm syn}(t) = I_{\rm Ae}(t) + I_{\rm Ar}(t) + I_{\rm Nr}(t) + I_{\rm G}(t) \quad (2)$$

where

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

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

$$I_{\rm Nr}(t) = \frac{g_{\rm N}(V(t) - V_E)}{(1 + C_{Mg}e^{(\frac{-0.062V(t)}{3.57})})}\sum_{j=1}^{N_E} w_j s_j^{\rm N}(t) \quad (5)$$

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

In the preceding equations $V_E = 0$ mV and $V_I = -70$ mV. The fractions of open channels s are given by:

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

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

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

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

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

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

The values of the conductances (in nS) for pyramidal neurons were: $g_{\rm Ae} = 2.08$, $g_{\rm Ar} = 0.052$, $g_{\rm N} = 0.164$ and $g_{\rm G} = 0.67$; and for interneurons: $g_{\rm Ae} = 1.62$, $g_{\rm Ar} = 0.0405$, $g_{\rm N} = 0.129$ and $g_{\rm G} = 0.49$. We consider that the NMDA currents have a voltage dependence that is controlled by the extracellular magnesium concentration, $[{\rm Mg}^{++}] = 1$ mM. We neglect the rise time of both AMPA and GABA synaptic currents, because they are typically extremely short (< 1 ms). The rise time for NMDA synapses is $\tau_{\rm N,r} = 2$ ms. The decay time for AMPA synapses is $\tau_{\rm AMPA} = 2$ ms, for NMDA synapses $\tau_{\rm N,d} = 100$ ms, and for GABA synapses $\tau_{\rm G} = 10$ ms.

Figure 1 shows schematically the synaptic structure assumed in the prefrontal cortical network we use to simulate the what-then-where

experimental design. The network is composed of N_E (excitatory) pyramidal cells and N_I inhibitory interneurons. In our simulations, we use $N_E = 1600$ and $N_I = 400$. The neurons are fully connected. There are different populations or pools of neurons in the prefrontal cortical network. Each pool of excitatory cells contains fN_E neurons, where f, the fraction of the neurons in any one pool, was set to be 0.05. There are four subtypes of excitatory pools, namely: Object-tuned (what-pools), space-tuned (where-pools), object-and-space-tuned (what-and-where-pools), and nonselective. Object pools are feature-specific. The spatial pools are location-specific and encode the spatial position of a stimulus. The integrated object-and-space tuned pools encode both specific feature and location information. The remaining excitatory neurons are in a nonselective pool. All the inhibitory neurons are clustered into a common inhibitory pool, so that there is global competition throughout the network.

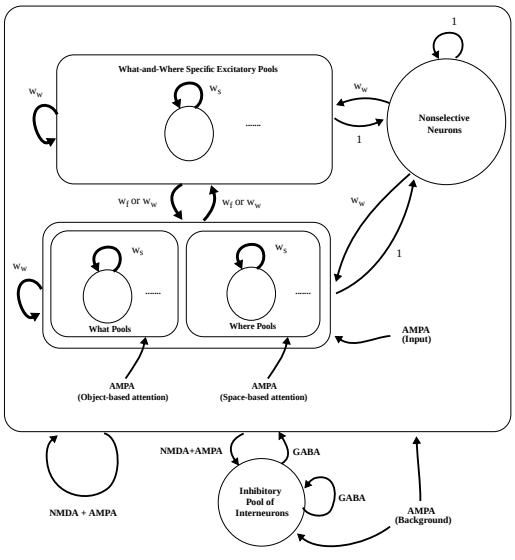

Figure 1: Prefrontal Cortical Module.

Assuming Hebbian learning, neurons within a specific excitatory pool are mutually coupled with a strong weight $w_s = 2.1$. Neurons in the inhibitory pool are mutually connected with an intermediate weight $w = 1$. They are also connected with all excitatory neurons with the same intermediate weight $w = 1$. The connection strength between two neurons in two different specific excitatory pools is weak and given by $w_w = 1-2f(w_s-1)/(1-2f)$ (=0.8778) unless otherwise specified. Neurons in a specific excitatory pool are connected to neurons in the nonselective pool with a feedforward synaptic weight $w = 1$ and a feedback synaptic connection of weight w_w. The connections between the different pools are set up so that specific integrated what-and-where pools are connected with the corresponding specific what-tuned pools and where-tuned pools, as if they were Hebbian learning based on the activity of individual pools while the different tasks are being performed. The forward connections (input to integrated what-and-where pools) are $w_f = 1.65$. The corresponding feedback synaptic connections are symmetric.

Each neuron (pyramidal cells and interneurons) receives $N_{\text{ext}} = 800$ excitatory AMPA synaptic connections from outside the network. These connections provide three different type of external interactions: 1) a background noise due to the spontaneous firing activity of neurons outside the network; 2) a sensory related input; and 3) an attentional bias that specifies the task (what or where delayed response). The external inputs are given by a Poisson train of spikes. We assume that Poisson spikes arrive at each external synapse with a rate of 3 Hz, consistent with the spontaneous activity observed in the cerebral cortex.The sensory input is encoded by increasing the external input Poisson rate ν_{ext} to $\nu_{\text{ext}}+\lambda_{\text{input}}$ to the neurons in the appropriate specific sensory pools. We used $\lambda_{\text{input}} = 85$ Hz. Finally, the attentional biasing specification of the task (i.e. which dimension is relevant) is modelled by assuming that each neuron in each of the pools associated with the relevant stimulus-domain (object or space) receives external Poisson spikes with an increased rate from ν_{ext} to $\nu_{\text{ext}} + \lambda_{\text{att}}$ throughout the trial. We use $\lambda_{\text{att}} = 85$ Hz.

We simulate the temporal evolution of fMRI signals (i.e. event-related) by convolving the total synaptic activity with the standard hemodynamics response of $h(t)$ [5], i.e.:

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

where

$$h(t) = t^{n_1}e^{-t\backslash t_1}/c_1 - a_2 t^{n_2}e^{-t\backslash t_2}/c_2$$

$$c_i = max(t^{n_i}e^{-t\backslash t_i})$$

In our simulation we calculated numerically the convolution by sampling the total synaptic activity every 0.1 seconds and introducing a cut-off at a delay of 25 seconds. The parameters utilized for the hemodynamic standard response $h(t)$ were taken from the paper of [5], and were: $n_1 = 6.0$, $t_1 = 0.9s$, $n_2 = 12.0$, $t_2 = 0.9s$, and $a_2 = 0.2$. Figure 3 plots the hemodynamic standard response $h(t)$ for this set of parameters.

3. Results

3.1. Electrophysiological Single-Cell Recordings

In this subsection, we present a theoretical analysis of neuronal activity in the primate prefrontal cortex underlying the execution of a 'What'-Then-'Where' working memory task. The microscopic neural recordings of [11] demonstrated the existence of neurons showing domain-specific-sensitivity, i.e. object-tuned activity in the what delay and location-tuned activity in the where delay, but they found also a large proportion of neurons showing integrated what-and-where-tuned activity during both what and where delays. During each trial of their experiment, while the monkey maintained fixation on a center spot, a sample object was briefly presented on the screen. After a first delay ('What'-delay), two test objects were briefly presented at two of four possible extrafoveal locations. one of the test objects matched the sample, the other was a non-match. After a second delay ('Where'-delay), the monkey had to respond with a saccade to the remembered location of the match. Figure 2A illustrates the experimental recordings of [11].

We performed [3] numerical simulations of the experiment of [11] and we evaluated in the model the neural spiking activity of specific pools. The simulation starts with a pre-cue period of 1000 ms. A target stimulus with a feature characteristic F_i and at a location S_j is presented next during the first cue period of 1500 ms. After the first cue period, the stimulus is removed, and only the feature characteristics of the target object has to be encoded and retained during a 'what'-delay period of 6500 ms. We modeled the attentional 'what'-bias by assuming that all feature-specific pools receive Poisson spikes with an increased rate $(\nu_{\mathrm{ext}} + \lambda_{\mathrm{att}})$. This is followed by a second cue period of 1500 ms, where a matched object reappeared at another new location different from the one originally cued during the first cue period. After that, only the location of the matched target has to be encoded, and the feature information can be ignored during this second 'where'-delay period of 6500 ms. Again, we modeled the attentional 'where'-bias by assuming that all location-specific pools receive Poisson spikes with an increased rate $(\nu_{\mathrm{ext}} + \lambda_{\mathrm{att}})$. This second delay is followed by a period of 1500 ms where the final probe is presented and a response has to be elicited.

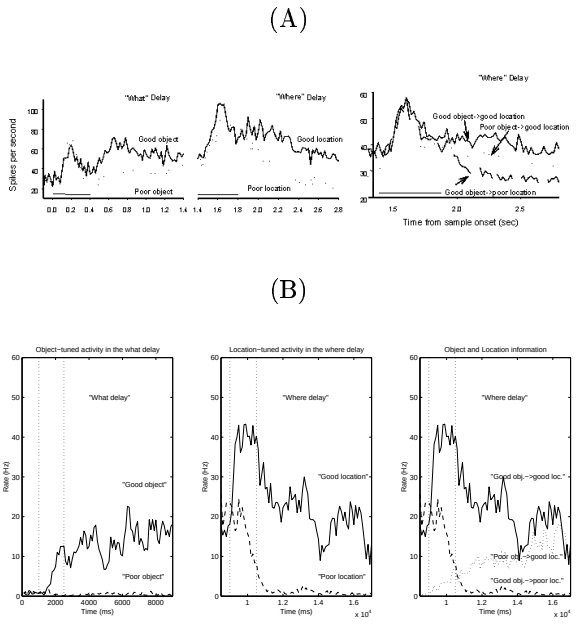

Figure 2: Single-Cell experiments and simulations.

Figure 2A and 2B plot the response of a single

prefrontal neuron (or averaged pool activity for the numerical simulations of Fig. 2B) showing object-tuned activity in the first what-delay and location-tuned activity during the second where-delay (middle and central pannel). The right pannel show object and location information during the second where-delay. For the generation of the simulations presented in Figure 2B, we plot the averaged pool activity of the different integrated what-and-where pools. As evidenced by the experiments, cueing a good location with a good object elicited more activity than cueing a good location with a poor object. A poor location elicited less activity than a good location, regardless of which object cued it. These specific global attractors corresponding to a specific stimulus-domain-attention condition, incorporates several single pool attractors, from the group of sensory pools (object or space specifics) and from the group of integrated what-and-where pools. The cue stimulus, and the biasing attentional top-down information applied to the sensory neurons, drive the system into the corresponding global attractor according to the biased competition mechanism. The numerical simulations show that the assumed microcircuits in the PFC are consistent with the empirical microscopic measurements of [11], and therefore offer a concrete microscopic organization of the PFC by an extended stimulus-domain-specific criterion that combines sensory pools with what-specific and where-specific sensitivity with integrating what-and-where pools.

3.2. Event-Related fMRI Data: 'What'-Then-'Where'

The work of [10] presents an event-related fMRI design with which the authors investigated the macroscopic organization of the prefrontal cortex. They used the behavioral task 'What-Then-Where' and 'Where-Then-What'. Figures 3A and 3C plot the temporal evolution the trial-averaged measured fMRI signal extracted from the ventrolateral and dorsolateral PFC, respectively. Activity in the ventrolateral and dorsolateral prefrontal cortex is observed during both what- and where- delay periods. Because of the similarity of the observed fMRI signal evolution under both the 'What-Then-Where' and 'Where-Then-What'conditions, especially during both periods of delay, the hypothesis that ventrolateral and dorsolateral PFC regions may differentially support working memory for object and spatial stimuli, respectively, could not be confirmed, suggesting a more functional organization of the PFC. Note in the figure, the 5-6 sec. delay of the fMRI signal due to the hemodynamical response.

We ran our model for the setup of [10] and simulated the temporal evolution of the fMRI signal [3,7]. We simulated with our model both the 'What-Then-Where' and 'Where-Then-What' conditions. We found that the experimentally observed event-related fMRI data could be obtained only if we assumed that the network associated with the dorsolateral PFC had a higher level of inhibition than the network associated with the ventrolateral PFC. The level of inhibition was increased by multiplying the maximal GABA conductivity constants by a factor of 1.025.

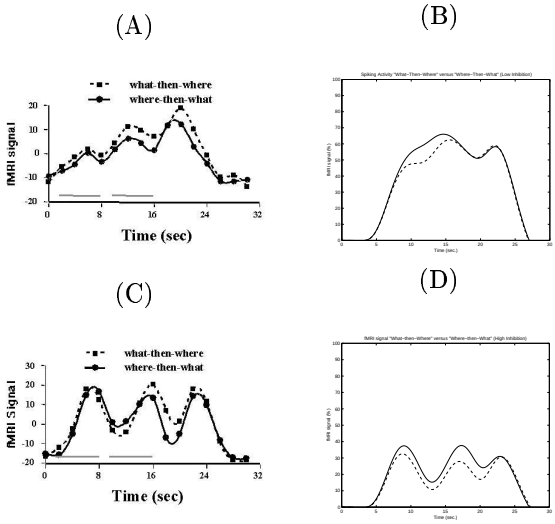

Figure 3: Event-related fMRI experiments and simulations.

Figures 3B and 3D present the simulated fMRI signal for both conditions and for both ventrolateral (low inhibition, top-right figure) and dorsolateral (high inhibition, bottom-right figure) PFC. The simulations compare favorably with the results of [10]. No differences is macroscopically

detected between the 'What-Then-Where' and 'Where-Then-What' conditions for the ventrolateral network-model with low inhibition or for the dorsolateral network-model with high inhibition. However, this fact does not mean that both ventrolateral and dorsolateral dynamic behaviors are identical during both 'What-Then-Where' and 'Where-Then-What' conditions, because the underlying microscopic activity is different. The spatio-temporal spiking activity shows that at the microscopic level, topographically organized by what, where, and what-and-where specific neurons, strong differences in the evolution and the structure of the succesivelly elicited attractors for each temporal period are evident. This detailed microscopic structure is lost at the macroscopic level of coarser spatial resolution measured by MRI. In fact, during the short term memory delay period associated with a 'what' ('where') task, only the neurons representing the feature characteristics (spatial location) of the cue maintain persistent activity and build up a stable global attractor in the network that maintains the firing during the delay period. These specific global attractors, corresponding to a specific stimulus-domain-attention condition, incorporate several single pool attractors, from the group of sensory pools (object or space specifics) and from the group of integrated what-and-where pools. The cue stimulus, and the biasing attentional top-down information applied to the sensory neurons, drives the system into the corresponding global attractor according to the biased competition mechanism [4,2].

4. Discussion

We presented a computational neuroscience based approach that integrates, via a detailed large-scale microscopic neurodynamical computational model, single-cells and fMRI measurements of the prefrontal cortex (PFC) associated with working memory processing. We demonstrated that this type of computational biophysical realistic modeling can be used for a thourough investigation of the topographical structure of the prefrontal cortex, providing a specifc way for contrasting different hypotheses with experimental evidence at different levels (single-cell and fMRI). For this purpose, we formulated an explicit model of the mechanisms that underlies working memory-related activity during the execution of delay tasks with a 'what'-then-'where' design (object and spatial delayed response within the same trial).

REFERENCES

1. Brunel, N. and Wang, X., J. of Comp. Neurosc. 11 (2001) 63.
2. Rolls, E. T. and Deco, G., *Computational Neuroscience of Vision*, Oxford University Press (2002)
3. Deco, G., Rolls, E. T. and Horwitz, B., J. of Cog. Neurosc., in press (2003).
4. Desimone, R. and Duncan, J., Annual Review of Neuroscience 18 (1995) 193.
5. Glover, G., Neuroimage 9 (1999) 416.
6. Goldman-Rakic, P., Handbook of physiology (1987) 373.
7. Tagamets, M. and Horwitz, B., Cerebral Cortex 8 (1998) 310.
8. Leung, H., Gore, J. and Goldman-Rakic, P., J. of Cog. Neurosc. 14 (2002) 659.
9. Miller, E., Nature Rev. Neurosc. 1 (2000) 59.
10. Postle, B. and D'Esposito, M., Psychobiology 28 (2000) 132.
11. Rao, S., Rainer, G. and Miller, E., Science 276 (1997) 821.