"What" and "Where" in Visual Working Memory: A Computational Neurodynamical Perspective for Integrating fMRI and Single-Neuron Data

Gustavo Deco[1], Edmund T. Rolls[2], and Barry Horwitz[3]

Abstract

■ Single-neuron recordings, functional magnetic resonance imaging (fMRI) data, and the effects of lesions indicate that the prefrontal cortex (PFC) is involved in some types of working memory and related cognitive processes. Based on these data, two different models of the topographical and functional organization of the PFC have been proposed: organization-by-stimulus-domain, and organization-by-process. In this article, we utilize an integrate-and-fire network to model both single-neuron and fMRI data on short-term memory in order to understand data obtained in topologically different parts of the PFC during working memory tasks. We explicitly model the mechanisms that underlie working-memory-related activity during the execution of delay tasks that have a "what"-then-"where" design (with both object and spatial delayed responses within the same trial). The model contains different populations of neurons (as found experimentally) in attractor networks that respond in the delay period to the stimulus object, the stimulus position, and to combinations of both object and position information. The pools are arranged hierarchically and have global inhibition through inhibitory interneurons to implement competition. It is shown that a biasing attentional input to define the current relevant information (object or location) enables the system to select the correct neuronal populations during the delay period in what is a biased competition model of attention. The processes occurring at the AMPA and NMDA synapses are dynamically modeled in the integrate-and-fire implementation to produce realistic spiking dynamics. It is shown that the fMRI data characteristic of the dorsal PFC and linked to spatial processing and manipulation of items can be reproduced in the model by a high level of inhibition, whereas the fMRI data characteristic of the ventral PFC and linked to object processing can be produced by a lower level of inhibition, even though the network is itself topographically homogeneous with no spatial topology of the neurons. This article, thus, not only presents a model for how spatial versus object short-term memory could be implemented in the PFC, but also shows that the fMRI BOLD signal measured during such tasks from different parts of the PFC could reflect a higher level of inhibition dorsally, without this dorsal region necessarily being primarily spatial and the ventral region object-related. ■

INTRODUCTION

The aim of this article is to integrate, via a large-scale neuronal network model that generates the dynamics and synaptic processing of neurons in an integrate-and-fire implementation, single-neuron and functional magnetic resonance imaging (fMRI) measurements of the prefrontal cortex (PFC) associated with visual working memory processing. One of the aims of this integration is to help interpret the fMRI signals recorded in topographically separate parts of the PFC and, more generally, to provide a fundamental approach to understanding fMRI signals.

The PFC is involved in at least some types of working memory and related processes such as planning (Fuster, 2000; Goel & Grafman, 1995; Goldman-Rakic, 1995; Goldman-Rakic, 1996), as shown by single-neuron (Funahashi, Bruce, & Goldman-Rakic, 1989), neuroimaging (Ungerleider, Courtney, & Haxby, 1998), and lesion studies (Levy & Goldman-Rakic, 1999; Goldman-Rakic, 1987). Working memory refers to an active system for maintaining and manipulating information in mind, held during a short period of time (usually seconds) (Baddeley, 1986).

Two models of the topographical and functional organization of the PFC have been proposed (see Miller, 2000, for a review). The first model proposes organization-by-stimulus-domain, with spatial ("where") working memory supported by the dorsolateral PFC in the neighborhood of the principal sulcus (Brodmann's area [BA] 46/9 in the middle frontal gyrus [MFG]); and object ("what") working memory supported by the ventrolateral PFC on the lateral convexity (BA 45 in the inferior frontal gyrus [IFG]). Some, but not all, fMRI studies in humans and single-cell data in primates

[1]Institucion Catalana de Recerca; Estudis Avançats (ICREA) and Universitat Pompeu Fabra, [2]University of Oxford, [3]National Institutes of Health

support this topographical organization (but see below) (Leung, Gore, & Goldman-Rakic, 2002; Wilson, O'Scalaidhe, & Goldman-Rakic, 1993; Goldman-Rakic, 1987; Fuster, Bauer, & Jervey, 1982). A second model proposes a hierarchical organization of the PFC, with nonmnemonic higher order functions (e.g., manipulation of items in memory) ascribed to dorsolateral prefrontal areas, and short-term memory maintenance functions ascribed to inferior prefrontal areas (Owen et al., 1999; D'Esposito et al., 1998; Petrides, 1994). Consistent with this second model, some event-related fMRI studies (Postle & D'Esposito, 1999, 2000) with a what-then-where design failed to find evidence for anatomical segregation of spatial and object visual working memory in the PFC of humans.

Analysis of how the PFC implements working memory has been extended by analyzing neuronal activity when a monkey performs a delay task with a what-then-where design (i.e., the monkey performs an object and spatial delayed response within the same trial). With this kind of experimental paradigm, Rao, Rainer, and Miller (1997) found some neurons that showed either object-tuned (what) or location-tuned (where) delay activity. However, a large percentage of the measured PFC neurons (52%) showed both what and where tuning. Similarly, in delay conditional response and delayed spatial response tasks, some neurons respond in the delay period to the stimulus object, some to the stimulus position, some to combinations of the stimulus object and its position, and some to the response required, with no clear topographical separation reported (White & Wise, 1999; Asaad, Rainer, & Miller, 1998, 2000; Hoshi, Shima, & Tanji, 1998).

Taken together, all these results imply that the topographical and functional organization of the PFC in relation to different types of working memory is still insufficiently understood. To elucidate the underlying mechanisms and the fMRI signals that can result in humans, we describe here a computational model of the PFC that can perform what-then-where and where-then-what memory tasks of the type illustrated in Figure 1 (see Methods for details), and has spiking neurons whose simulated activities are directly comparable to the single-neuron data recorded in the PFC (Rao et al., 1997). The integrate-and-fire neurons are organized into a hierarchical set of attractor or autoassociation networks each implementing a short-term memory as illustrated in Figure 2 (see Methods for details and Rolls & Treves, 1998; Rolls & Deco, 2002). There are separate attractors for ''what'' stimulus was shown, for ''where'' the stimulus was shown, and a combined ''what–where'' attractor pool of neurons. The relative activity of the different attractor pools required to perform the what-then-where and where-then-what short-term memory tasks is provided by an external attentional signal that biases the different pools of neurons within the general framework of the biased competition model of attention (Rolls & Deco, 2002; Reynolds & Desimone, 1999; Chelazzi, Miller, Duncan, & Desimone, 1993; Chelazzi, 1998; Miller, Gochin, and Cross, 1993; Motter, 1993; Spitzer, Desimone, & Moran, 1988; Moran & Desimone, 1985). Rolls and Deco (2002) added to this framework by introducing a neurodynamical theoretical framework for biased competition, which enables the whole framework to operate from the level of integrate-and-fire neurons up to global phenomena such as short-term memory and attention. Because the model is implemented at the integrate-and-fire single-neuron level with the biophysical properties of the synapses part of the simulation, we are then able, as described in this article, to integrate over the total synaptic activity within an area and convolve with the hemodynamic response function illustrated in Figure 3 (see Methods for details and Rolls and Deco, 2002; Horwitz, Tagamets, & McIntosh, 1999; Horwitz, Friston, & Taylor, 2000) in order to model the fMRI blood oxygen level-dependent (BOLD) signal and simulate human neuroimaging experiments. Basing the link to the BOLD signal on the total synaptic activity (in fact, current) within a region is appropriate because channels are opened by synaptic activity that require ions to be pumped back against electrochemical gradients, and this is the very energy-intensive aspect of neuronal activity. Experimental support for this idea has been provided by Logothetis, Pauls, Augath, Trinath, and Oeltermann (2001) (for reviews, see Lauritzen, 2001; Jueptner & Weiller, 1995).

RESULTS

Single-Neuron Recordings

In this subsection, we present a theoretical analysis of neuronal activity in the primate PFC underlying the execution of a what-then-where working memory task. The neuronal recordings of Rao et al. (1997) demonstrated the existence of neurons showing stimulus-domain-specific sensitivity, that is, 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, and the other was a nonmatch. After a second delay (where-delay), the monkey had to respond with a saccade to the remembered location of the match. Figures 4A and 5A illustrate the neuronal recordings of Rao et al.

The simulation of this experiment starts with a precue period of 1000 msec, in which the network exhibits spontaneous activity (3 Hz for the excitatory pools and

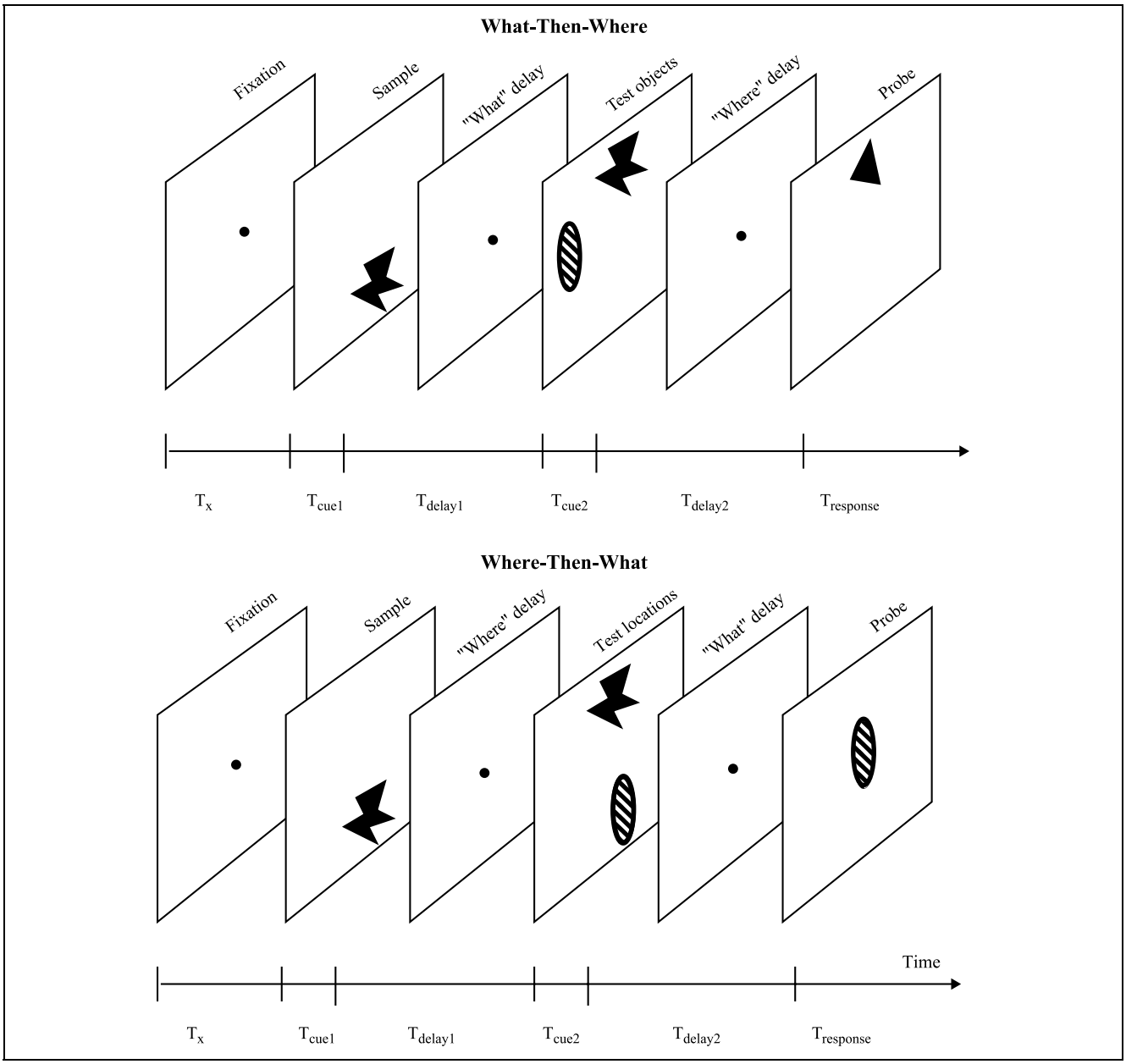

Figure 1. Schematic diagram of a typical what-then-where and where-then-what trial. Following an instructional cue, the behavioral task begins with a fixation period (T_{fix}), followed by an initial cueing stimulus presentation (T_{cue1}), followed by a first delay period (T_{delay1}), followed by the presentation of a matching stimulus and a distractor (T_{cue2}), followed by a second delay period (T_{delay2}), followed by a probe stimulus ($T_{response}$) that elicited a response. In what-then-where trials, an object delay task is followed by a spatial delayed response task. In where-then-what trials, a spatial delay task is followed by an object delayed response task.

9 Hz for the inhibitory pools, as in the experimental recordings of Wilson, Scalaidhe, & Goldman-Rakic, 1994). 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 msec (i.e., during this period the object pool F_i and the spatial pool S_j receive external Poisson spikes with an increased rate from ν_{ext} to $\nu_{ext} + \lambda_{input}$). After the first cue period, the stimulus is removed, and only the feature characteristics of the target object have to be encoded and retained during a what-delay period of 6500 msec. We modeled the attentional what-bias by assuming that all feature-specific pools (i.e., all pools F_n for all n) receive Poisson spikes with an increased rate ($\nu_{ext} + \lambda_{att}$). This is followed by a second cue period of 1500 msec, where a matched object (identical to the cued 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 msec. Again, we modeled the attentional where-bias by assuming that all location-specific pools (i.e., all

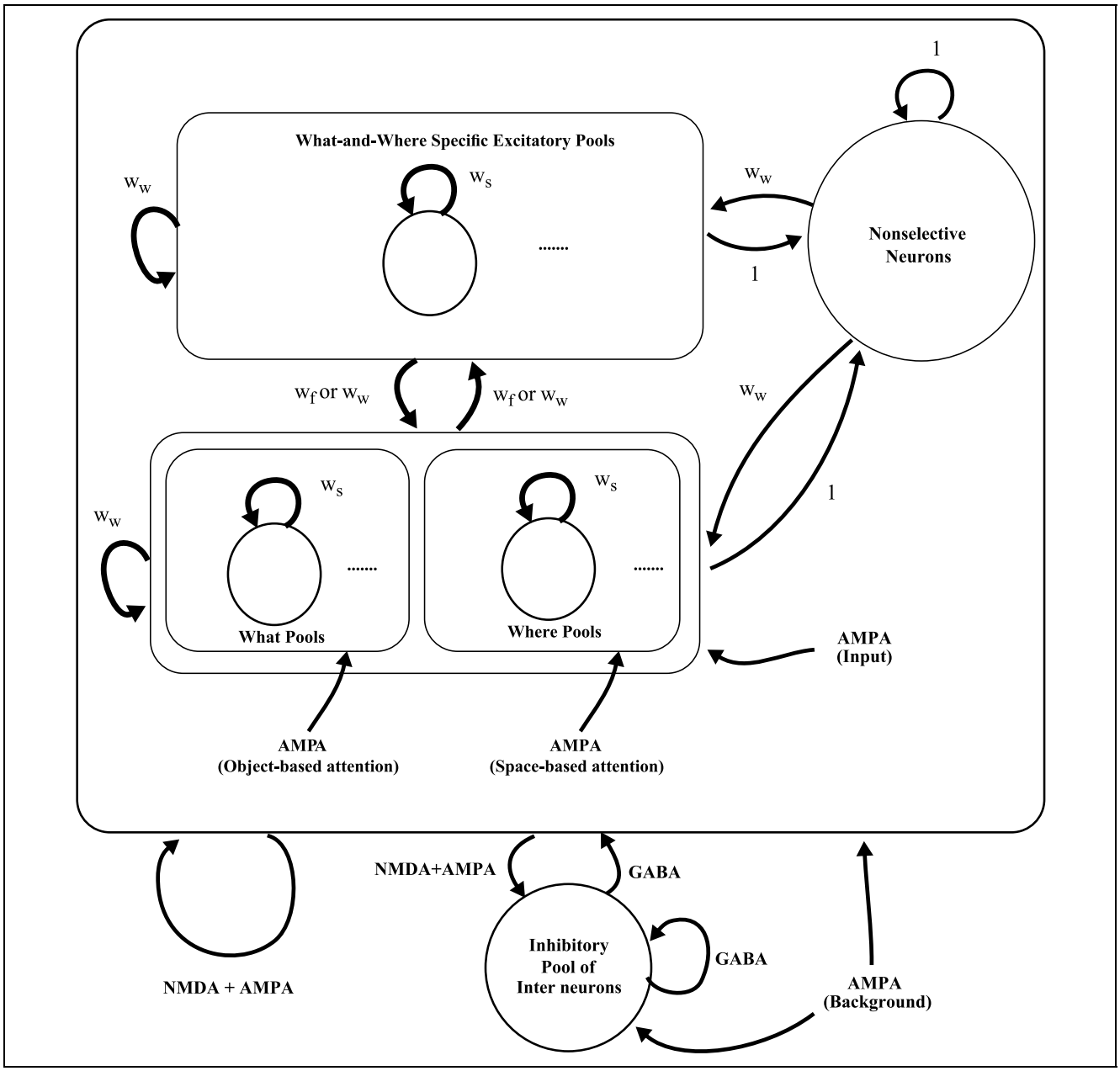

Figure 2. Prefrontal cortical module. The synaptic current flows into the cell are mediated by four different families of receptors. The recurrent excitatory postsynaptic currents are given by two different types of EPSP, respectively, mediated by AMPA and NMDA receptors. These two glutamatergic excitatory synapses are on the pyramidal cells and interneurons. The external inputs (background, sensory input, or external top-down interactions from other areas) are mediated by AMPA synapses on pyramidal cells and interneurons. Inhibitory GABAergic synapses on pyramidal cells and interneurons yield corresponding IPSPs. Each neuron receives also N_{ext} excitatory AMPA synaptic connections from outside the network. These connections provide three different type of external interactions: (1) a background noise due to the external spontaneous firing activity; (2) a sensory input; and (3) an attentional stimulus-domain-specific bias.

pools S_n for all n) receive Poisson spikes with an increased rate ($\nu_{ext} + \lambda_{att}$). This second delay is followed by a period of 1500 msec where the final probe is presented and a response has to be made.

Figures 4B and 5B show the results of the simulations, which can be directly compared with the corresponding experimental observations of Rao et al. (1997) shown in Figures 4A and 5A. Figures 4A and 4B plot the responses of prefrontal neurons showing either object-tuned (top part of A and B) or location-tuned (bottom part of A and B) delayed activity. ''Good object'' and ''poor object'' refer to the objects used as samples (''good'' means the preferred object or feature for that neuron, and ''poor'' refers to the nonpreferred object or feature for that neuron). ''Good locations'' and ''poor locations'' refer to the locations cued by the matching object during the second cue (''good'' means the preferred spatial location for that neuron, and

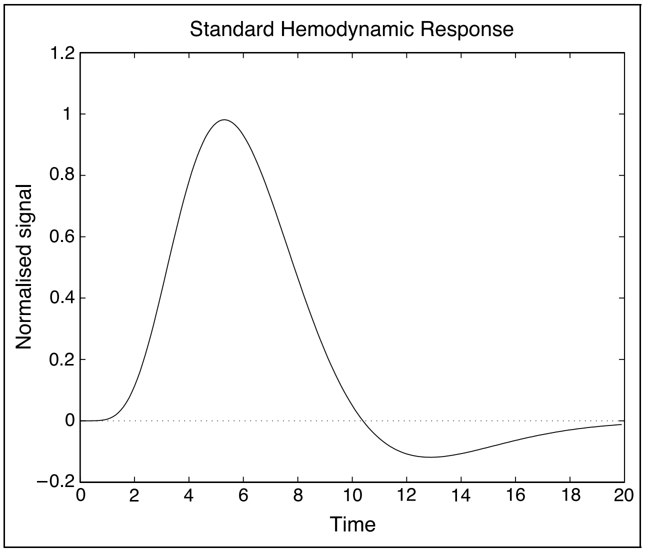

Figure 3. Standard hemodynamic response function utilized for the convolution with synaptic activity to produce simulated event-related fMRI activity from the neuronal network model of the PFC.

''poor'' refers to the nonpreferred spatial location for that neuron). The experimental and numerical bin widths were 20 msec. In the case of the simulations (Figure 4B) we present the calculated temporal evolution of the averaged population activity (over all neurons in each specific pool during the temporal bin of 20-msec period) of specific pools. The spatiotemporal spiking activity shows that during the first short-term memory what-delay period, only the what-specific neurons representing the feature characteristics of the first cue maintain persistent activity and build up a stable global attractor in the network that maintains the firing during the delay period (Figure 4B, top). On the other hand, during the second short-term memory where-delay period, only the where-specific neurons representing the location of the matched cue during the second stimulus presentation maintain persistent activity and build up a stable global attractor in the network that maintains the firing during the second delay period (Figure 4B, bottom).

Figure 5 shows the responses of prefrontal neurons with best responses to a combination of a particular object and a particular response in the second delay period. These neurons show object-tuned activity in the first what-delay (left panel), and location-tuned activity during the second where-delay (middle panel). The right panel shows activity that is tuned to both object and location during the second where-delay. Figure 5A shows neuronal data from the recordings of Rao et al. (1997). Figure 5B shows the averaged activity of the population of neurons in the simulation that respond best to a combination of what and where information. In the simulation as well as in the neurophysiological 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, which correspond to a specific condition of attention to a particular location triggered by a particular object condition in the task, incorporate several single-condition attractors, including object-specific, location-specific, and object- and location-specific attractor populations of neurons. 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 thus show that the assumed microcircuits in the PFC shown in Figure 2 are consistent with the empirical microscopic measurements (single-neuron recording) of Rao et al. (1997), and instantiate, therefore, a concrete microscopic (neuron level) organization of the PFC (which is stimulus-domain-specific) that incorporates sensory pools with what-specific and where-specific sensitivity with combination what-and-where pools. The specific underlying wiring could be established by Hebbian learning.

Event-Related fMRI Data: What-Then-Where

In this section we describe simulations of the event-related fMRI investigation of Postle and D'Esposito (1999) in which they investigated the time course of the activations in the dorsolateral PFC (Areas 9 and 46) and ventrolateral PFC (Areas 44, 45, 47) during the what-then-where and where-then-what tasks described above (see Figure 1). Figure 6A and C plot the temporal evolution of the trial-averaged fMRI signal extracted from the ventrolateral and dorsolateral PFC, respectively. Delay-period activity in the ventrolateral and dorsolateral PFC was 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 delay periods, Postle and D'Esposito concluded that 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 Figure 6A and B the 5- to 6-sec delay of the fMRI signal due to the hemodynamical response.

In order to understand better the neurodynamical substrate underlying these fMRI results, which suggest a more functional organization of the PFC than a stimulus-domain-specific organization (object vs. spatial), and especially to harmonize these facts with the topographical stimulus-domain-specific organization suggested by some single- neuron recording experiments (Rao et al., 1997; Goldman-Rakic, 1987), we ran our model for the setup of Postle and D'Esposito (1999) and

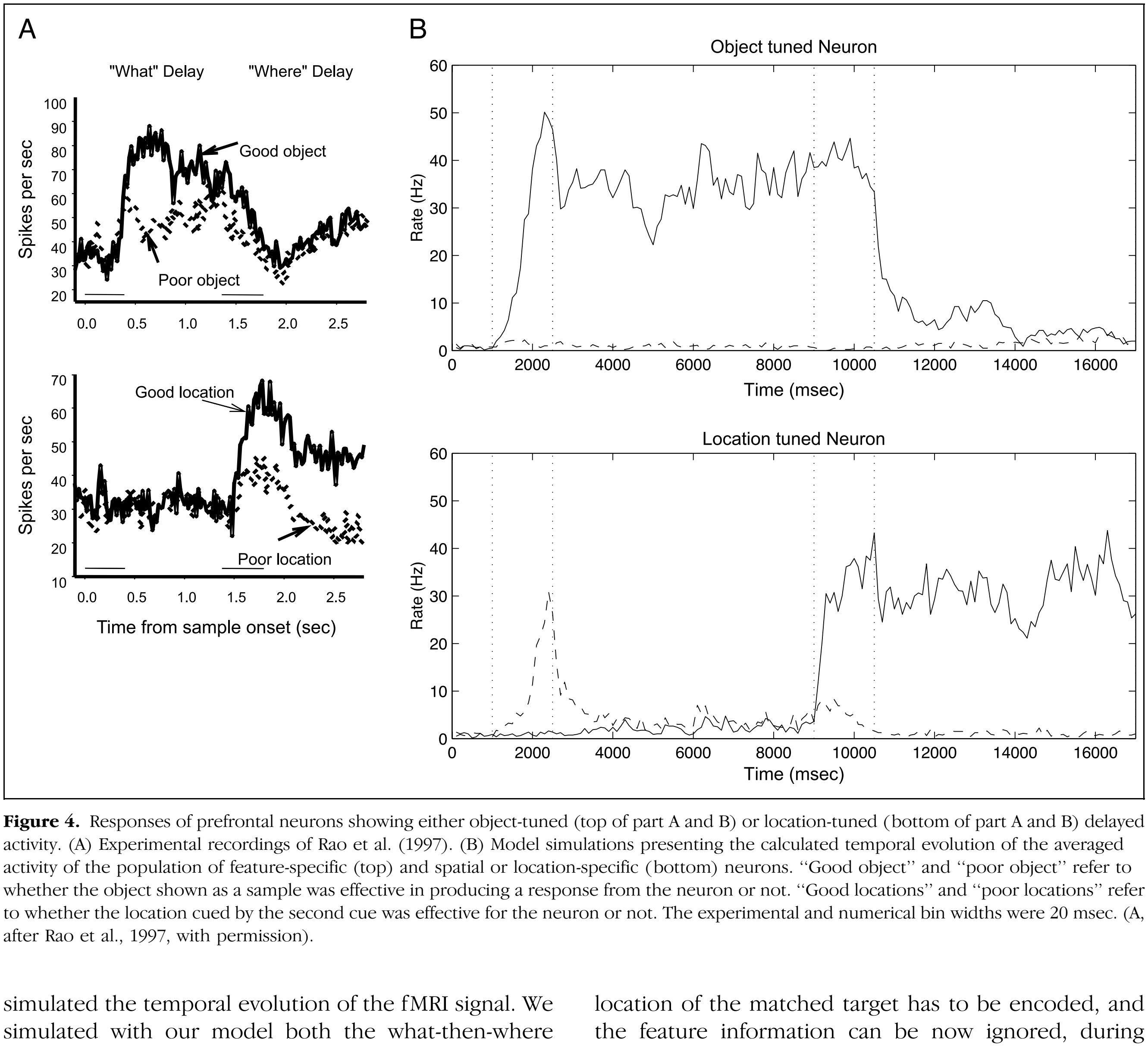

Figure 4. Responses of prefrontal neurons showing either object-tuned (top of part A and B) or location-tuned (bottom of part A and B) delayed activity. (A) Experimental recordings of Rao et al. (1997). (B) Model simulations presenting the calculated temporal evolution of the averaged activity of the population of feature-specific (top) and spatial or location-specific (bottom) neurons. "Good object" and "poor object" refer to whether the object shown as a sample was effective in producing a response from the neuron or not. "Good locations" and "poor locations" refer to whether the location cued by the second cue was effective for the neuron or not. The experimental and numerical bin widths were 20 msec. (A, after Rao et al., 1997, with permission).

simulated the temporal evolution of the fMRI signal. We simulated with our model both the what-then-where and where-then-what conditions.

For the what-then-where condition, the simulation starts with a precue period of 1000 msec during which the network exhibits spontaneous activity. Next, a target stimulus with a feature characteristic F_i and at a location S_j is presented during the first cue period of 1500 msec (i.e., during this period, object pool F_i and spatial pool S_j receive external Poisson spikes with an increased rate from ν_{ext} to $\nu_{ext} + \lambda_{input}$). After the first cue period, the stimulus is removed, and only the feature characteristics of the target object have to be encoded and retained during a what-delay period of 6500 msec. We model the attentional what-bias by assuming that all feature-specific pools receive Poisson spikes with an increased rate ($\nu_{ext} + \lambda_{att}$). This is followed by a second cue period of 1500 msec, where a matched object (identical to the cued object) reappears 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 now ignored, during this second where-delay period of 6500 msec. Again, we model the attentional where-bias by assuming that all location-specific pools (i.e., all pools S_n for all n) receive Poisson spikes with an increased rate ($\nu_{ext} + \lambda_{att}$). This second delay is followed by a period of 1500 msec where the final probe is presented and a response has to be performed.

For the where-then-what condition, after the precue period of 1000 msec during which the network exhibits spontaneous activity, a target stimulus with a feature characteristic F_i and at a location S_j is presented during the first cue period of 1500 msec. (During this period object pool F_i and spatial pool S_j receive external Poisson spikes with an increased rate from ν_{ext} to $\nu_{ext} + \lambda_{input}$). After the first cue period, the stimulus is removed, and only the spatial location of the target object has to be encoded and retained during a where-delay period of 6500 msec. We model the attentional where-bias by assuming that all spatial-specific pools

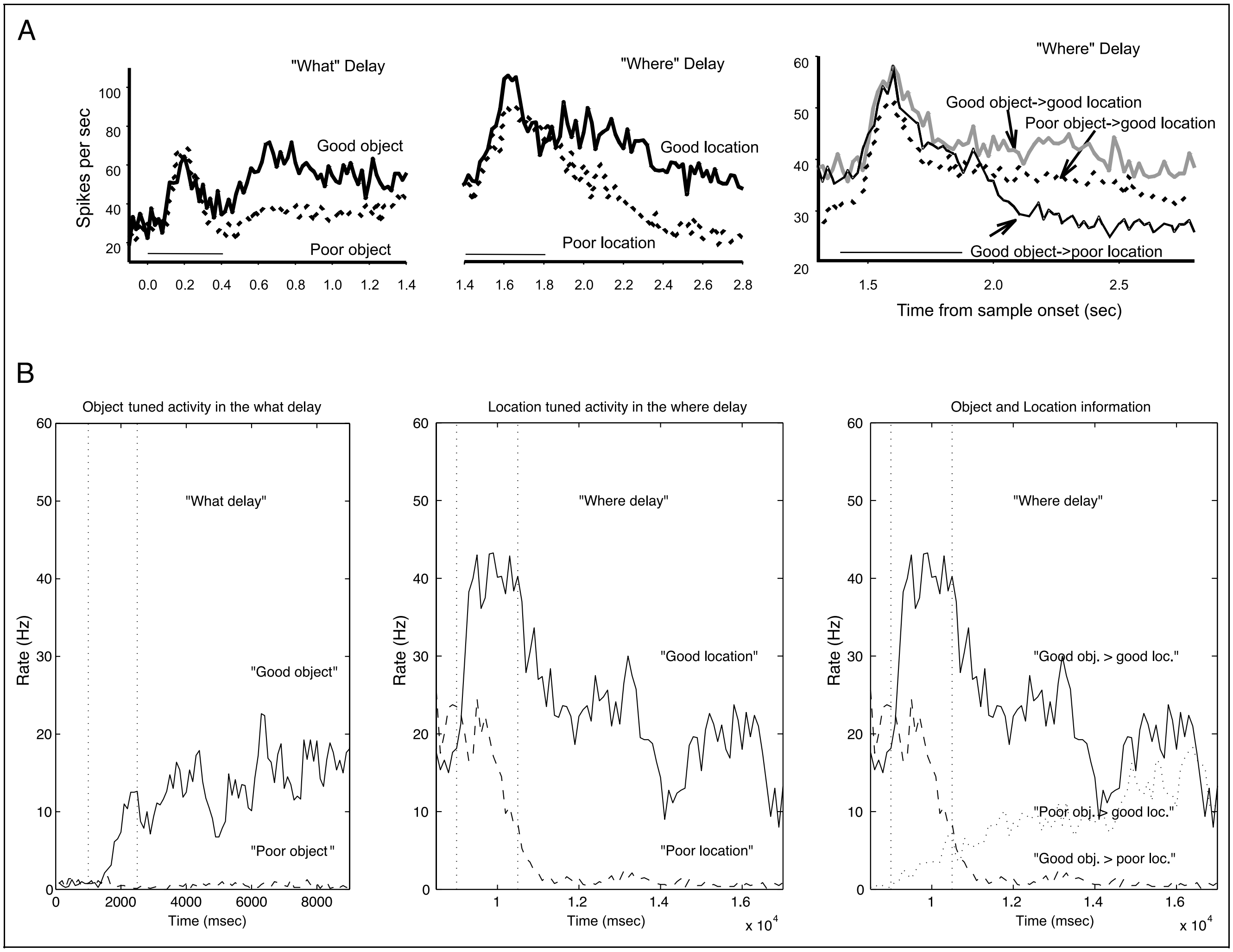

Figure 5. Responses of prefrontal neurons with best responses to a particular combination in the second delay period of a particular object and a particular response. These neurons show object-tuned activity in the first what-delay (left panel) and location-tuned activity during the second where-delay (middle panel). The right panel shows activity that is tuned to both object and location during the second where-delay. (A) Experimental single-neuron recordings of Rao et al. (1997). (B) Model simulations presenting the calculated temporal evolution of the averaged activity of the population of neurons responding best to a combination of what and where information. In the case illustrated, cueing a good location with a good object elicited more activity than cueing a good location with a poor object; and poor location elicited less activity than a good location, regardless of which object cued it. (After Rao et al., 1997, with permission.)

receive Poisson spikes with an increased rate ($\nu_{\text{ext}} + \lambda_{\text{att}}$). This is followed by a second cue period of 1500 msec, where an object different from the target reappeared at the first cued location. After that, only the feature characteristics of the object presented during the second cue period have to be encoded, and the spatial information can now be ignored during this second what-delay period of 6500 msec. Again, we model the attentional what-bias by assuming that all feature-specific pools (i.e., all pools F_n for all n) receive Poisson spikes with an increased rate ($\nu_{\text{ext}} + \lambda_{\text{att}}$). This second delay period is followed by a period of 1500 msec where the final probe is presented and a response has to be performed.

We found that the event-related fMRI data of Postle and D'Esposito (1999) could be modeled by varying the parameters that regulate the dynamics of the network, and that no spatial topology had to be introduced into the prefrontal network. Specifically, we had to assume that the network associated with the dorsolateral PFC has a higher level of inhibition than the network associated with the ventrolateral PFC. The level of inhibition was increased by increasing by a factor of 1.025 the maximal GABA conductivity constants g_{GABA} specified in Appendix A. The evidence for this finding is described next.

Figure 6B and D presents 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 Postle and D'Esposito (1999) shown in Figure 6A and C. The important result of the

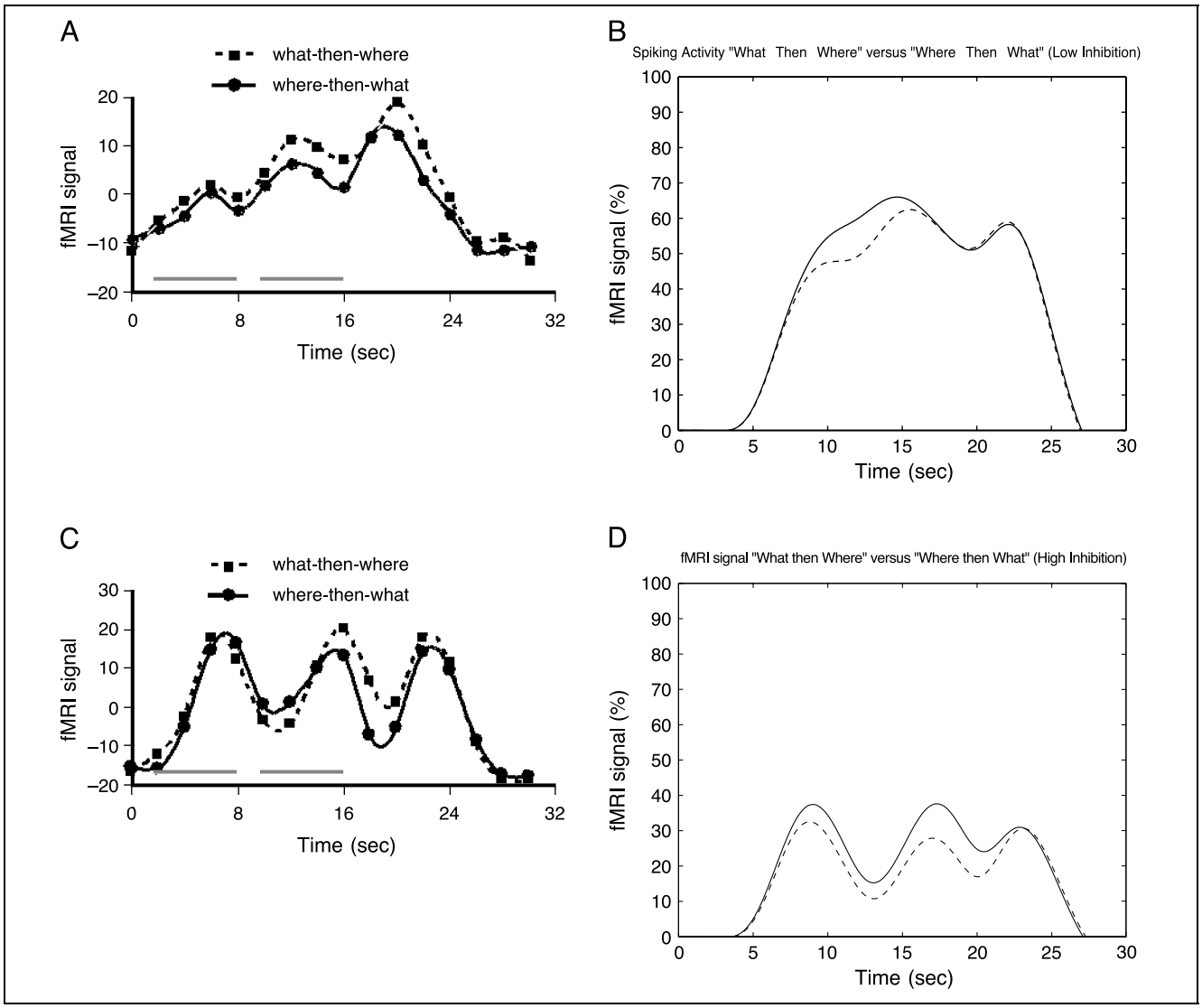

Figure 6. Temporal evolution of the trial-averaged measured fMRI signal extracted from the ventrolateral (A) and dorsolateral (C) PFC (after Postle and D'Esposito, 1999); and simulated fMRI signals for both the what-then-where and where-then-what conditions and for both the low-inhibition (ventrolateral) (B) and high-inhibition (dorsolateral) models (D) PFC. Delay period activity in the ventrolateral and dorsolateral PFC is observed during both what- and where-delay periods. Two conditions were compared: what-then-where and where-then-what. Because of the similarity of the development of the fMRI signal under both conditions, and specially during both delay periods, the hypothesis that the ventrolateral PFC region may differentially support working memory for objects, and the dorsolateral PFC for spatial stimuli, could not be confirmed, suggesting a more functional organization of the PFC. Note in the figure the 5- to 6-sec delay of the fMRI signal due to the hemodynamical response.

simulations is that fMRI activations of the type described by Postle and D'Esposito for the ventrolateral PFC (Figure 6A) can be obtained by using a low level of inhibition (Figure 6B); and the fMRI activations of the types described for the dorsolateral PFC (Figure 6C) can be obtained by using a high level of inhibition (Figure 6D).

Unsurprisingly, because there was no topology of object versus spatial neurons in the simulated network, no differences are detected macroscopically (i.e., at the fMRI level) between the what-then-where and where-then-what conditions for the low-inhibition (ventrolateral) network model (top right). Nor are any differences detected macroscopically between the what-then-where and where-then-what conditions for the high-inhibition (dorsolateral) network model (bottom right). The similarities between the results of the simulations at the macroscopic level in the what-then-where and where-then-what task conditions are as found experimentally in the fMRI signal. However, this fact does not mean of course that the underlying neuronal responses are identical during the what-then-where and where-then-what conditions. Figures 7 and 8 plot rastergrams for a population of single neurons in the simulations for both conditions for the ventrolateral network model with low inhibition (Figure 7), and for the dorsolateral network model with high inhibition (Figure 8). The spatiotempo-

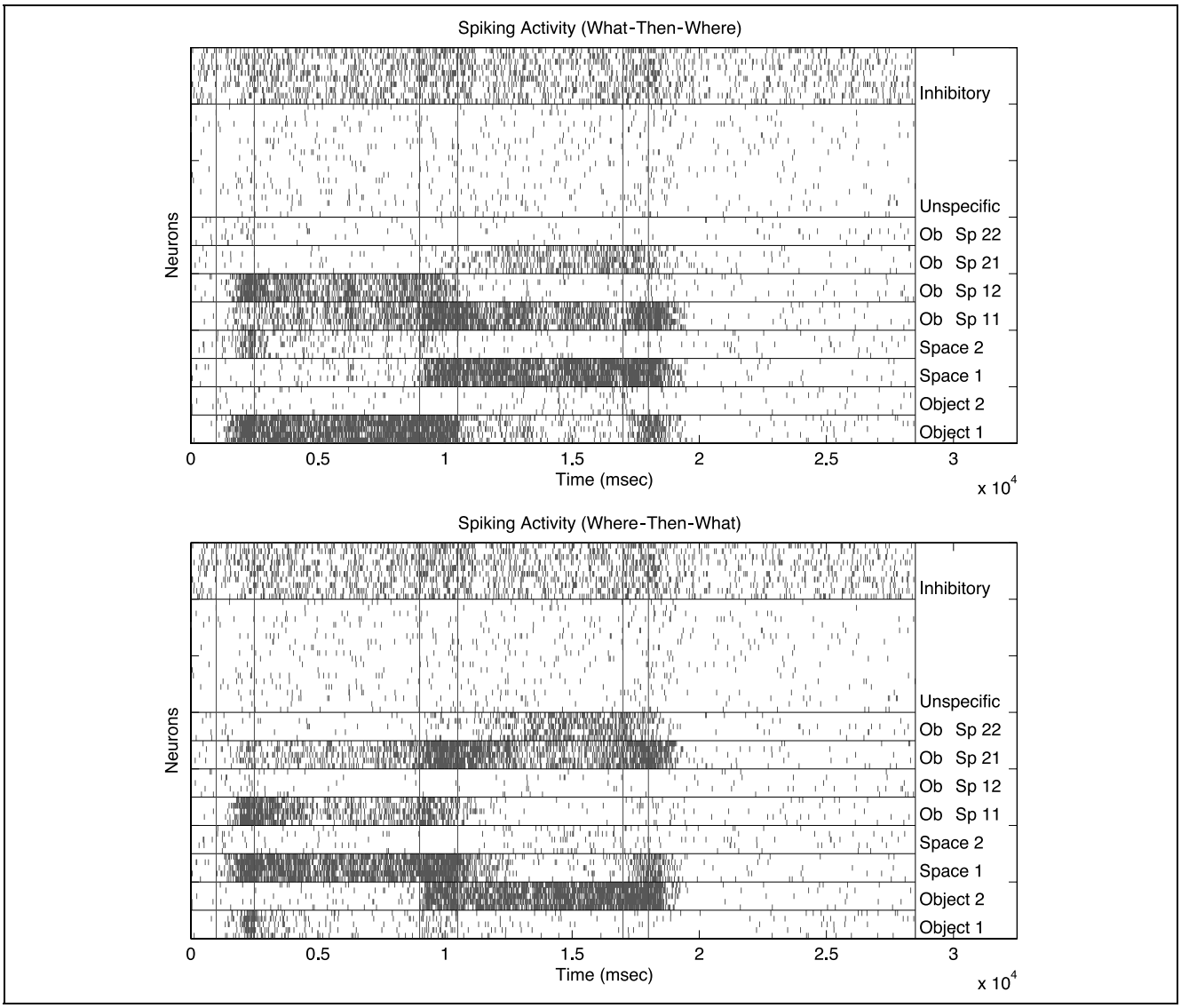

Figure 7. Rastergrams of simulations of the what-then-where (top) and where-then-what (bottom) task after the experimental paradigm of Postle and D'Esposito (1999), for the case of a network model with a low level of inhibition, which produces results like those from the ventrolateral region of the PFC. The spiking activity of different populations of neurons is shown, some with "what" tuning (to Object 1 or 2), some with "where" tuning to spatial location (labeled Space 1 or 2), and some with what-and-where tuning to different combinations of objects and locations (labeled Ob-Sp).

ral spiking activity shows that at the neuronal (microscopic) level, topographically organized in Figures 7 and 8 by what (Object 1 or 2), where (Space 1 or 2), and what-and-where (Ob-Sp) specific neurons, strong differences in the evolution and the structure of the successively elicited attractors for each temporal period are evident. This fine-grain microscopic (neuron-level) 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 (or where) task, only the neurons representing the object feature characteristics (or 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, each corresponding to a specific stimulus-domain-attention condition, incorporate several single-pool attractors, formed from the group of sensory pools or neuronal population (object-specific or space-specific neuronal populations) and from the group of combination tuned what-and-where neuronal 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.

To test the alternative hypothesis that the dorsolateral PFC is more associated with spatial working memoryand the ventrolateral PFC is more associated with object working memory, we ran simulations for the same tasks, but now assuming that in the dorsolateral PFC there are more spatial sensory neurons (a factor

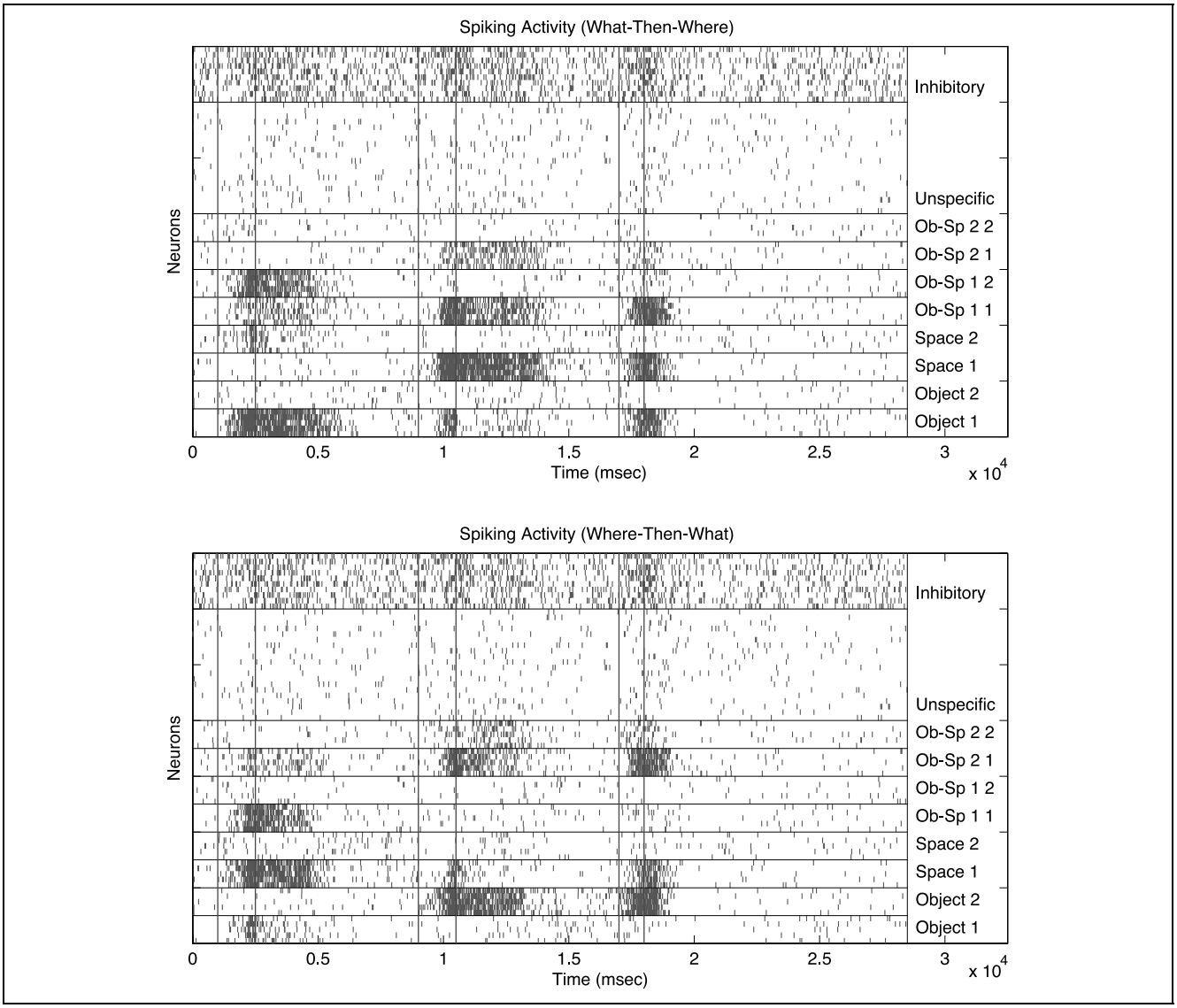

Figure 8. Rastergrams of simulations of the what-then-where (top) and where-then-what (bottom) task after the experimental paradigm of Postle and D'Esposito (1999), for the case of a network model with high level of inhibition, which produces results like those from the dorsolateral region of the PFC. The spiking activity of different populations of neurons is shown, some with "what" tuning (to Object 1 or 2), some with "where" tuning to spatial location (labeled Space 1 or 2), and some with what-and-where tuning to different combinations of objects and locations (labeled Ob-Sp).

1.05 more than object sensory neurons), and in the ventrolateral PFC there are more object sensory neurons (also a factor 1.05). For this set of simulations, we set the level of inhibition to be low for both the ventrolateral and dorsolateral PFC. Figure 9 shows the simulation results. An asymmetric behavior of the fMRI signal response is observed between the what-then-where (dashed line) and where-then-what (continuous line) conditions. In the dorsolateral PFC (bottom of the figure) more activity is observed during the first where-delay period during the where-then-what condition, whereas in the what-then-where condition more activity is observed during the second where-delay period. In the ventrolateral PFC (top of the figure) more activity is observed during the second what-delay period during the where-then-what condition, whereas in the what-then-where condition more activity is observed during the first what-delay period. These simulated fMRI signals are not consistent with the empirical findings of Postle and D'Esposito (1999). We emphasize that Figure 9 is the only figure in this article in which the spatial and object neurons are treated as being topographically organized into separable populations, realized in the simulations performed by running the simulations separately with more object or more spatial neurons to represent the ventral and dorsolateral PFC.

In summary, our simulations show that single-cell and fMRI data are consistent with the hypothesis that differences between the dorsal and ventral PFC in the

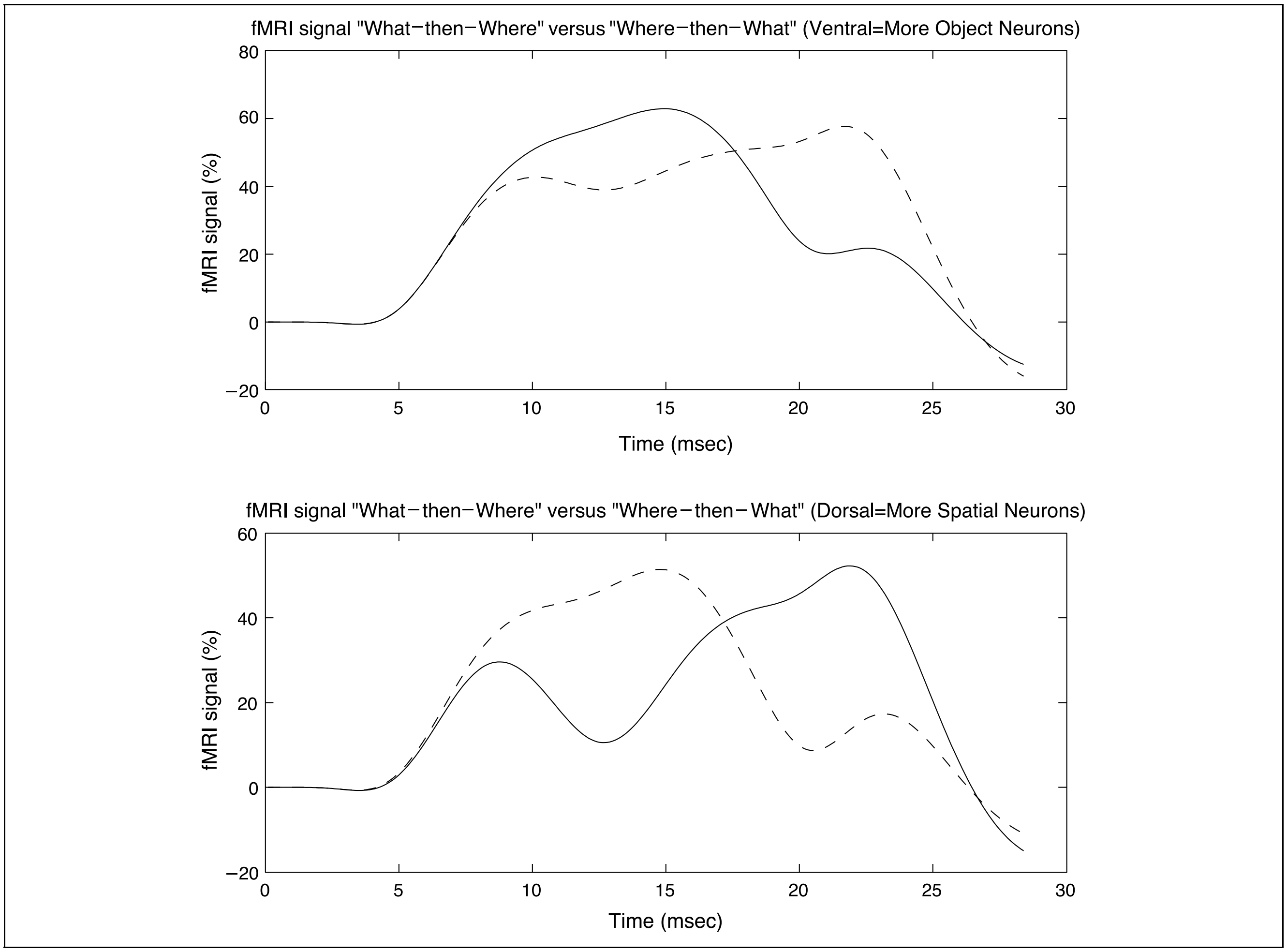

Figure 9. Simulations with spatial topology in the network to simulate data according to the hypothesis that the dorsolateral PFC is more associated with spatial working memory, and the ventrolateral PFC is more associated with object working memory. The simulations for the dorsolateral PFC model have more spatial than object neurons (by a factor of 1.05). The simulations for the ventrolateral PFC model have more object than spatial neurons (by a factor of 1.05). An asymmetry in the behavior of the fMRI signal was observed between the what-then-where (dashed line) and where-then-what (continuous line) conditions. In the dorsolateral PFC condition (bottom of the figure) more activity is observed during the where (first) delay period during the where-then-what condition, and in the what-then-where condition more activity is observed during the where (second) delay period. Consistently, in the ventrolateral PFC (top of the figure), more activity is observed during the what (second) delay period during the where-then-what condition, and in the what-then-where condition more activity is also observed during the what (first) delay period.

fMRI signal can be accounted for by a higher level of inhibition in the dorsal PFC relative to the ventral PFC. The simulations also show that the imaging results are consistent with an architecture that is stimulus-domain-specific at the microscopic or neuronal level (with different what, where, and what-and-where sensitive neurons), but with the different neurons intermixed so that there is no separate topography at the macroscopic level with separate regions for objects (ventral) versus space (dorsal). Indeed, the simulations shown in Figure 9 with networks for objects and locations that are even minimally spatially segregated (by 10%) do not reproduce the fMRI results of Postle and D'Esposito (1999). Of course, this latter point does not mean that empirical evidence for object versus location topology in the PFC will not be revealed in future.

DISCUSSION

In the present study, we investigated two different hypotheses concerning the functional organization of the PFC during the delay period of a working memory task. One, the organization-by-stimulus-domain hypothesis posits that the ventrolateral PFC contains a large number of neurons that maintain active representations of the visual features of objects during working memory delay periods, and that the dorsolateral PFC contains a large number of neurons that maintain representations of the spatial locations of objects during such delays. Second, the organization-by-process hypothesis asserts that the main functional difference between the ventrolateral and dorsolateral PFC is that the ventrolateral PFC is concerned with maintenance of

the information during the delay period of a working memory task, whereas the dorsolateral PFC is involved in manipulation of stored information. Our main finding was that our model could account for the neurophysiological activity seen in both the ventrolateral and dorsolateral PFC during the delay periods of working memory tasks (as shown by the empirical results of Rao et al., 1997; see Figure 4), and at the same time could provide simulated fMRI patterns that matched experimental findings during a what-then-where short-term memory task for both PFC sectors (as shown by the fMRI findings of Postle & D'Esposito, 1999; see Figure 6). However, we could not do this if we assumed that the difference between ventrolateral and dorsolateral PFC followed the organization-by-stimulus-domain hypothesis (see Figure 9). Rather, we had to assume that the differences between these two prefrontal regions resulted from assigning a greater amount of inhibition to the dorsolateral portion of the PFC. Our modeling thus suggests that different levels of competition of the networks associated with the ventrolateral and dorsolateral PFC could be the neural basis of the different fMRI signals associated with these brain regions. In addition, the network model suggests that one important functional difference between the dorsolateral PFC and the ventrolateral PFC is related to greater inhibition in the dorsolateral PFC than in the ventrolateral PFC. Both brain areas show maintenance capabilities related to their capacities to maintain stable attractors during delay periods, but the increased level of inhibition assumed in the dorsolateral PFC may be associated with the capacity of this brain region to support more complex functions. Exactly what those more complex functions may be is not revealed by the present studies, but higher inhibition in the dorsolateral PFC might be useful for maintaining several separate representations and preventing the formation of a global attractor, which could be useful if several items must be held in memory for manipulation. Another possibility is that the information coming into the dorsolateral PFC might be more distributed (which might be consistent with a spatial as compared to an object representation; cf. Rolls, Stringer, & Trappenberg, 2002) and thus might require more inhibition to prevent an overdistributed representation, which does have disadvantages (Rolls & Treves, 1998; Rolls & Deco, 2002).

Previous studies using such large-scale models to relate neural activity to functional brain imaging data have employed leaky integrator-type neuronal units (e.g., Tagamets & Horwitz, 1998, employed Wilson–Cowan units). One novel aspect of our model and of the work described in this article is that our architecture, which has multiple attractor networks each composed of a separate population of neurons interconnected to form a separate or local attractor, is implemented with integrate-and-fire neurons (in the theoretical framework of Brunel and Wang, 2001), so that the details of the spiking and synaptic mechanisms involved can be understood, and so that predictions can be made about the effects, for example, of neurotransmitters and pharmacological agents that have particular effects on synaptic transmission. The processes occurring at the AMPA and *N*-methyl-D-aspartate (NMDA) synapses are dynamically modeled in the integrate-and-fire implementation to produce realistic spiking dynamics. In addition, the architecture described is an extension beyond attractor architectures in that different neuronal pools or populations, connected hierarchically, are simulated, and in that an attentional bias is applied that allows the network to select the correct response given the current stimulus and attentional bias. Specifically, in this article we have formulated the biased competition hypothesis for the first time in the framework of a very detailed and biophysically realistic spiking neuronal and synaptic dynamics. This has enabled us to integrate the effects of recurrent maintenance associated with short-term memory with the biasing attentional effects associated with a cognitive task. As a result, we have been able to provide not only a quantitative and qualitative description of different experiments, but also are able to make concrete predictions that are testable experimentally, for example, about the effects of dopamine receptor blockade on the operation of the working memory systems described here (in preparation).

It was pointed out in the Introduction that among experimentalists, there are two major hypotheses concerning the organization of PFC: the organization-by-stimulus-processing-domain hypothesis (Goldman Rakic, 1987), and the organization-by-functional-processing hypothesis (Owen et al., 1999; D'Esposito et al., 1998). For the latter, the main distinction is centered on the difference between a manipulation function (dorsolateral PFC) and a maintenance function (ventrolateral PFC) during the delay period of a working memory task. Our finding—that the distinction between dorsolateral and ventrolateral PFC is that the former has a greater level of inhibition than does the latter—is more consistent with the organization-by-functional-processing hypothesis than it is with the organization-by-stimulus-domain hypothesis. A test for the future would be to explicitly employ a task requiring active manipulation of the items in working memory and compare simulated and experimental fMRI activity. However, we should mention that our results do not rule out entirely the organization-by-stimulus-domain hypothesis. It may be that spatial processing of visual information itself requires a greater level of inhibition in the PFC. A way to test this would be to expand our model to explicitly model the preprocessing steps in posterior cortex required for object and spatial vision (cf. Corchs and Deco, 2002; Rolls and Deco, 2002; Tagamets & Horwitz, 1998).

Overall, the results described here demonstrate how the use of large-scale neural modeling that allows one to relate microscopic and macroscopic neurophysiological activity enables one to make explicit hypotheses about the neural substrates for high-level cognitive concepts that can be tested empirically in both humans and nonhumans.

METHODS

What-Then-Where and Where-Then-What Delayed Response Experiments

What-then-where- and where-then-what delayed response tasks have been used to study the organization of the PFC at the single-neuron (Rao et al., 1997) and fMRI levels (Postle & D'Esposito, 1999, 2000). Figure 1 shows two examples of these designs. Following an instructional cue, the behavioral task begins with a fixation period (T_{fix}), followed by an initial cueing stimulus presentation (T_{cue1}), followed by a first delay period (T_{delay1}), followed by the presentation of a matching stimulus and a distractor (T_{cue2}), followed by a second delay period (T_{delay2}), followed by a probe stimulus ($T_{response}$), which indicates that a response can be made. In what-then-where trials, one has to encode the featural (object-based) characteristics of the initial cue, ignoring the spatial location on the screen, and to retain in short-term memory this what-based information during the first delay. Following this first delay, during an intermediate period, two stimuli appeared on the screen, a target (identical to the cued object) and a distracting stimulus with featural characteristics that distinguish it from the target. The location of the target stimulus to be matched is different from the position at which the initial cueing target appeared. After that, only the location of the matched target has to be encoded, and the feature information can now be ignored. During the second delay period only the where-based information (location) has to be retained. Upon presentation of the final probe, one has to judge whether it occupied the same location as the location of the retained matched target during the intermediate stimulus presentation. In where-then-what trials, one has to encode the spatial location of the initial cue, ignoring the featural characteristics of the object presented on the screen, and to retain in short-term memory only the where-based information (location) during the first delay. Following this first delay, during an intermediate period, two stimuli appear on the screen, one at the initially cued location (target location) and the other at another location. Both stimuli have different featural characteristics from each other. The featural characteristics of the stimulus at the target location have to be encoded and retained during the second delay period; that is, only the what-based information (object) has to be retained. Upon presentation of the final probe, one has to judge whether the probe has the same featural characteristics of the retained matched target during the intermediate stimulus presentation.

We simulated this task. In all our simulations, we used the following parameters for the task (following Postle and D'Esposito, 1999, 2000) T_{fix} = 1000 msec, T_{cue1} = 1500 msec, T_{delay1} = 6500 msec, T_{cue2} = 1500 msec, T_{delay2} = 6500 msec, and $T_{response}$ = 1500 msec.

The Neurodynamical Model of the Prefrontal Cortex

We follow the theoretical framework for modeling a single integrate-and-fire attractor network introduced and studied by Brunel and Wang (2001), and extend it to multiple hierarchically organized networks organized into a biased competition architecture introduced by the authors (Corchs & Deco, 2002; Rolls & Deco, 2002; Deco & Lee, 2002; Deco & Zihl, 2001). We incorporate shunting inhibition (Battaglia & Treves, 1998; Rolls & Treves, 1998) and inhibitory-to-inhibitory cell synaptic connections (Brunel & Wang, 2001), which are useful in maintaining stability of the dynamical system, and incorporate appropriate currents to achieve low firing rates (Brunel & Wang, 2001; Amit & Brunel, 1997). According to the experimental neurophysiological evidence of Rao et al. (1997), we assume the existence of different types of neuronal populations or pools that show either object-tuned (what), or location-tuned (where), or both what and where tuned activity in the delay period. We show that local synaptic connections (which could be set up by associative learning) between these neuronal pools are sufficient for operation of the model. In this section, we describe the architecture and operation of the model, and the neuronal parameters and equations used are given in Appendix A.

The Neurons

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

These potentials are integrated by the cell, and if a threshold θ is reached, a δ-pulse (spike) is fired and transmitted to other neurons, and the potential of the neuron is reset. The incoming presynaptic δ-pulse current from another neuron is first low-pass filtered by the synaptic and membrane time constants, and is then realized as an EPSP or IPSP in the one-compart-

ment neuronal model. We use biologically realistic parameters (McCormick, Connors, Lighthall, & Prince, 1985). We take for both excitatory and inhibitory neurons a resting potential $V_L = -70$ mV, a firing threshold $\theta = -50$ mV, and a reset potential $V_{reset} = -55$ mV. The membrane capacitance C_m is 0.5 nF for the pyramidal neurons and 0.2 nF for the inhibitory interneurons. The membrane leak conductance g_m is 25 nS for pyramidal cells and 20 nS for interneurons. The refractory period τ_{ref} is 2 msec for pyramidal cells and 1 msec for interneurons. Hence, the membrane time constant $\tau_m = C_m/g_m$ is 20 msec for pyramidal cells and 10 msec for interneurons, respectively.

The synaptic current flows into the cells are mediated by three different families of receptors. The recurrent excitatory postsynaptic EPSPs are mediated by AMPA and NMDA receptors. These two glutamatergic excitatory synapses are on the pyramidal cells and on the interneurons. The external inputs (background, sensory input, or external top-down interaction from other areas) are mediated by AMPA synapses on pyramidal cells and interneurons. Inhibitory GABAergic synapses on pyramidal cells and interneurons yield the corresponding IPSPs. The mathematical descriptions of each synaptic channel are provided in Appendix A, and the corresponding parameters are also specified there. We consider that the NMDA currents have a voltage dependence that is controlled by the extracellular magnesium concentration (Jahr & Stevens, 1990), $C_{Mg^{2+}} = 1$ mM. We neglect the rise time of both AMPA and GABA synaptic currents because they are typically extremely short (<1 msec). The rise time for NMDA synapses is $\tau_{NMDA,rise} = 2$ msec (Spruston, Jonas, & Sakmann, 1995; Hestrin, Sah, & Nicoll, 1990). All synapses have a latency (time delay) of 0.5 msec. The decay time for AMPA synapses is $\tau_{AMPA} = 2$ msec (Spruston et al., 1995; Hestrin et al., 1990), for NMDA synapses $\tau_{NMDA,decay} = 100$ msec (Spruston et al., 1995; Hestrin et al., 1990), and for GABA synapses $\tau_{GABA} = 10$ msec (Xiang, Huguenard, & Prince, 1998; Salin & Prince, 1996). The synaptic conductivities for each receptor type were taken from Brunel and Wang (2001) and were adjusted using a mean field analysis to be approximately 1 nS in magnitude; these were consistent with experimen- tally observed values (Destexhe et al., 1998) (see Appendix A). As was noted by Brunel and Wang (2001), Wang (1999), and Lisman, Fellous, and Wang (1998), the recurrent excitation was assumed to be largely mediated by the NMDA receptors, in order to provide more robust persistent activity during the short-term memory related delay period, and the amplitude of recurrent excitation was smaller than that of inhibition; therefore, the net recurrent input to a neuron was hyperpolarizing during spontaneous activity (i.e., without external inputs) (Brunel & Wang, 2001; Amit & Brunel, 1997). Figure 2 shows schematically the synaptic structure assumed in the prefrontal cortical network.

The Network Architecture

The network is composed of N_E (excitatory) pyramidal cells and N_I inhibitory interneurons. In our simulations, we use $N_E = 1600$ and $N_I = 400$, consistent with the neurophysiologically observed proportion of 80% pyramidal cells versus 20% interneurons (Abeles, 1991). The neurons are fully connected (with synaptic strengths as specified below). Neurons in the prefrontal cortical network shown in Figure 2 are clustered into populations or pools. Each pool of excitatory cells contains fN_E neurons (in our simulations $f = 0.05$). There are two different types of pool: excitatory and inhibitory. There are four subtypes of excitatory pool, namely, object-tuned (what pools), space-tuned (where pools), object-and-space-tuned (what-and-where pools), and nonselective. Object pools are feature specific, encoding, for example, the identity of an object (e.g., form, color, etc.). 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 do not have specific sensory, response or biasing inputs, and are in a nonselective pool. (The neurons in the nonselective pool have some spontaneous firing, and help to introduce some noise into the simulation, which aids in generating the almost Poisson spike firing patterns of neurons in the simulation that are a property of many neurons recorded in the brain (Brunel & Wang, 2001). All the inhibitory neurons are clustered into a common inhibitory pool, so that there is global competition throughout the network.

We assume that the synaptic coupling strengths between any two neurons in the network act as if they were established by Hebbian learning; that is, the coupling will be strong if the pair of neurons have correlated activity, and weak if they are activated in an uncorrelated way. Because of this, neurons within a specific excitatory pool are mutually coupled with a strong weight $w_s = 2.1$. Neurons in the inhibitory pool are mutually connected with an intermediate weight $w = 1$ (forming the inhibitory to inhibitory connections that are useful in achieving nonoscillatory firing). They are also connected with all excitatory neurons with the same intermediate weight $w = 1$. The connection strength between two neurons in two different specific excitatory pools is weak and given by $w_w = 1 - 2f(w_s - 1)/(1 - 2f)$ unless otherwise specified (see next paragraph). 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 and where-tuned pools, as if they were based on Hebbian learning of the activity of individual pools while the

Table 1. Neuronal Connectivity Between Different Neuronal Pools in the Architecture of Figure 2

Pools	F_1	F_2	S_1	S_2	FS_{11}	FS_{12}	FS_{21}	FS_{22}	*Unsp.*	*Inh.*
F_1	w_s	w_w	w_w	w_w	w_f	w_f	w_w	w_w	1	1
F_2	w_w	w_s	w_w	w_w	w_w	w_w	w_f	w_f	1	1
S_1	w_w	w_w	w_s	w_w	w_f	w_w	w_f	w_w	1	1
S_2	w_w	w_w	w_w	w_s	w_w	w_f	w_w	w_f	1	1
FS_{11}	w_f	w_w	w_f	w_w	w_s	w_w	w_s	w_w	1	1
FS_{12}	w_f	w_w	w_w	w_f	w_w	w_s	w_w	w_w	1	1
FS_{21}	w_w	w_f	w_f	w_w	w_w	w_w	w_s	w_w	1	1
FS_{22}	w_w	w_f	w_w	w_f	w_w	w_w	w_w	w_s	1	1
Unsp.	w_w	w_w	w_w	w_w	w_w	w_w	w_w	w_w	1	1
Inh.	1	1	1	1	1	1	1	1	1	1

F_1 and F_2: object-tuned pools (corresponding to two possible objects, e.g., Object A and B); S_1 and S_2: space-tuned pools (corresponding to two locations, e.g., left and right); FS_{ij}: combination-tuned what-and-here neuronal pools for each of the possible combinations of "what" (F_i) and "where" (S_j) sensory information coming from the domain-specific pools. Unsp. = nonspecific neuronal pool; Inh. = inhibitory neuron pool.

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 (i.e., identical to the corresponding feedforward synaptic connections).

The strengths of the feedforward and feedback connections between different pools are indicated in Table 1. The neurons in the stimulus-domain-specific sensory pools have forward and backward strong weight connections with neurons with which they are associated in the what-and-where–tuned pools. In our simulations, we consider two feature-tuned pools F_1 and F_2 (corresponding to two possible objects, e.g., object A and B), two space-tuned pools S_1 and S_2 (corresponding to two locations, e.g., left and right), and the corresponding four integrated what-and-where pools (FS_{ij}) for each of the possible combinations of what (F_i) and where (S_j) sensory information coming from the domain-specific pools.

Each neuron (pyramidal cells and interneurons) receives $N_{ext} = 800$ excitatory AMPA synaptic connections from outside the network. These connections provide three different types of external interactions: (1) a background noise due to the spontaneous firing activity of neurons outside the network; (2) a sensory-related input (object- or space-specific); 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. In order to model the background spontaneous activity of neurons in the network (Brunel & Wang, 2001), 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 (Rolls & Treves, 1998; Wilson et al., 1994). In other words, the effective external spontaneous background input rate of spikes to each cell is $\nu_{ext} = N_{ext} \times 3$ Hz $= 2.4$ kHz. The sensory input is encoded by increasing the external input Poisson rate ν_{ext} to $\nu_{ext} + \lambda_{input}$ to the neurons in the appropriate specific sensory pools (Brunel & Wang, 2001). For example, if the stimulus is defined as an object with a feature characteristic F_i and at a spatial location S_j, then the neurons in the sensory object pool F_i and in the spatial sensory pool S_j will receive the increased external Poisson input just defined. We used $\lambda_{input} = 85$ Hz. Finally, the attentional biasing specification of the task (i.e., which dimension is relevant) is modeled 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 $\nu_{ext} + \lambda_{att}$ throughout the trial. We used $\lambda_{att} = 85$ Hz. This external top-down domain-specific input probably comes from the external prefrontal neurons that directly encode abstract rules (Wallis, Anderson, & Miller, 2001), which in turn are influenced by the reward system (in the orbitofrontal cortex and amygdala [Rolls, 1999]), which enables the correct context to be selected. During the last 100 msec of the response period, the external rate to all neurons is increased by a factor 1.5 in order to take into account the increase in afferent inputs due to behavioral responses and reward signals (Brunel & Wang, 2001).

The cortical architecture introduced above has the characteristic that its different global attractors, corresponding to the different domain-specific relevant sensory situations, are each composed of a set of single-pool attractors, where the single pools (each pool is a population of neurons) that are active represent a particular combination of what-tuned, where-tuned, and what-and-where–tuned pools. The cue stimulus and the biasing top-down attentional information drive the system into the corresponding attractor. In fact, the system is dynamically driven according to the biased competition hypothesis (Reynolds & Desimone, 1999; Chelazzi et al., 1993; Chelazzi, 1998; Miller et al., 1993; Motter, 1993; Spitzer et al., 1988; Moran & Desimone, 1985). Multiple excitatory pools of neurons activated by the sensory cue stimulus engage in competitive interactions using the interneurons to implement the global competition. The external, top-down interactions bias this competition in favor of specific pools, resulting in the buildup of the global attractor that corresponds to the stimulus-domain-specific situation required. In this way, irrelevant sensory information will be suppressed by the underlying neurodynamics, implementing a form of internal prefrontal attentional system, which is the basis of the attentional top-down bias transmitted to posterior perceptual areas (Rolls & Deco, 2002).

All neuronal and synaptic equations were integrated using the second-order Runge–Kutta method, with an integration step of $dt = 0.1$ msec. Checks were performed to show that this was sufficiently small. For the neural membrane potential equations, interpolation of the spike time and their use in the synaptic currents and potentials were taken into account following the prescription of Hansel, Mato, Meunier, and Neltner (1998), in order to avoid numerical problems due to the discontinuity of the membrane potential and its derivative at the spike firing time. The external trains of Poisson spikes were generated randomly and independently.

Simulation of the Event-related fMRI Response: Hemodynamic Convolution of Synaptic Activity

The links between neural and synaptic activity, and fMRI measurements, are still not fully understood. The fMRI signal is unfortunately strongly filtered and perturbed by the hemodynamic delay inherent in the BOLD contrast mechanism (Buxton & Frank, 1997). The fMRI signal is only a secondary consequence of neuronal activity, and yields therefore a blurred distortion of the temporal development of the underlying brain processes. Regionally, increased oxidative metabolism causes a transient decrease in oxyhemoglobin and increase in deoxyhemoglobin, as well as an increase in CO_2 and NO. This provokes, over several seconds, a local dilatation and increased blood flow in the affected regions that leads by overcompensation to a relative decrease in the concentration of deoxyhemoglobin in the venules draining the activated region; the alteration of deoxyhemoglobin, which is paramagnetic, can be detected by changes in $T2$ or $T2^*$ in the MRI signal as a result of the decreased susceptibility and thus decreased local inhomogeneity, which increases the MR intensity value (Glover, 1999; Buxton & Frank, 1997; Buxton, Wong, & Frank, 1998). Glover (1999) demonstrated that a good fit of the hemodynamical response can be achieved by the following analytic function:

$$h(t) = c_1 t^{n_i} e^{-\frac{t}{t_1}} - a_2 c_2 t^{n_2} e^{-\frac{t}{t_2}}$$

$$c_i = \max(t^{n_i} e^{-\frac{t}{t_1}})$$

where t is the time, and c_1, c_2, a_2, n_1, and n_2 are parameters that are adjusted to fit the experimental measured hemodynamical response.

Recently, the putative neural basis of the BOLD signal has been explored by Logothetis et al. (2001). They analyzed simultaneously recorded neuronal and fMRI responses from the visual cortex of monkeys, and observed that the largest magnitude changes were observed in local field potentials (LFPs), which at recording sites characterized by transient responses were the only signal that significantly correlated with the hemodynamic responses. These findings suggest that the BOLD contrast mechanism reflects the synaptic input and intracortical processing of a given area rather than its spiking output.

Based on these results, and similar to Horwitz and Tagamets (1999), we simulate the temporal evolution of fMRI signals by convolving the total synaptic activity with the standard hemodynamic response formulation of Glover (1999) presented above. The total synaptic current (I_{syn}) is given by the sum of the absolute values of the glutamatergic excitatory components (implemented through NMDA and AMPA receptors) and inhibitory components (GABA) (Tagamets & Horwitz, 1998). As described above, we consider that external excitatory contributions are produced through AMPA receptors ($I_{AMPA,ext}$), while the excitatory recurrent synapses are produced through AMPA and NMDA receptors ($I_{AMPA,rec}$ and $I_{NMDA,rec}$). The GABA inhibitory currents are denoted by I_{GABA} (see Appendix A for details). Consequently, the fMRI signal activity is calculated by the following convolution equation:

$$S_{\text{f MRI}}(t) = \int_0^\infty h(t - t') I_{\text{syn}}(t') dt'$$

In our simulation, we calculated numerically the convolution by sampling the total synaptic activity every 0.1 sec and introducing a cutoff at a delay of 25 sec. The parameters utilized for the hemodynamic standard response $h(t)$ were taken from the article of Glover (1999), and were $n_1 = 6.0$, $t_1 = 0.9$ sec, $n_2 = 12.0$, $t_2 = 0.9$ sec, and $a_2 = 0.2$. Figure 3 plots the hemodynamic standard response $h(t)$ for this set of parameters.

APPENDIX A

In this appendix we give the mathematical equations that describe the spiking activity and synapse dynamics in the network, following in general the formulation described by Brunel and Wang (2001). Each neuron is described by an integrate-and-fire model. The subthreshold membrane $V(t)$ potential of each neuron evolves according to the following equation:

$$C_m \frac{dV(t)}{dt} = -g_m(V(t) - V_L) - I_{\text{syn}}(t) \qquad (1)$$

where $I_{syn}(t)$ is the total synaptic current flow into the cell. When the membrane potential $V(t)$ reaches the threshold θ a spike is generated, and the membrane potential is reset to V_{reset}. The neuron is unable to

spike during the first τ_{ref}, which is the absolute refractory period.

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

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

where

$$I_{\text{AMPA,ext}}(t) = g_{\text{AMPA,ext}}(V(t) - V_{\text{E}}) \sum_{j=1}^{N_{ext}} s_j^{\text{AMPA,ext}}(t) \tag{3}$$

$$I_{\text{AMPA,rec}}(t) = g_{\text{AMPA,rec}}(V(t) - V_{\text{E}}) \sum_{j=1}^{\text{N}_{\text{E}}} w_j \, s_j^{\text{AMPA,rec}}(t) \tag{4}$$

$$I_{\text{NMDA,rec}}(t) = \frac{g_{\text{NMDA,rec}}(V(t) - V_{\text{E}})}{(1 + C_{\text{Mg}^{++}} \exp(-0.062V(t))/3.57)} \times \sum_{j=1}^{\text{N}_{\text{E}}} w_j s_j^{\text{NMDA,rec}}(t) \tag{5}$$

$$I_{\text{GABA}}(t) = g_{\text{GABA}}(V(t) - V_{\text{I}}) \sum_{j=1}^{N_{\text{I}}} s_j^{\text{GABA}}(t) \tag{6}$$

In the preceding equations $V_{\text{E}} = 0$ mV and $V_{\text{T}} = -70$ mV. The synaptic strengths w_j are specified in Methods and in Table 1. The fractions of open channels are given by:

$$\frac{ds_j^{\text{AMPA,ext}}(t)}{dt} = -\frac{s_j^{\text{AMPA,ext}}(t)}{\tau_{\text{AMPA}}} + \sum_k \delta(t - t_j^k) \tag{7}$$

$$\frac{ds_j^{\text{AMPA,rec}}(t)}{dt} = -\frac{s_j^{\text{AMPA,rec}}(t)}{\tau_{\text{AMPA}}} + \sum_k \delta(t - t_j^k) \tag{8}$$

$$\frac{ds_j^{\text{NMDA,rec}}(t)}{dt} = -\frac{s_j^{\text{NMDA,rec}}(t)}{\tau_{\text{NMDA,decay}}} + \alpha x_j(t) \times (1 - s_j^{\text{NMDA,rec}}(t)) \tag{9}$$

$$\frac{dx_j(t)}{dt} = -\frac{x_j(t)}{\tau_{\text{NMDA,rise}}} + \sum_k \delta(t - t_j^k) \tag{10}$$

$$\frac{ds_j^{\text{GABA}}(t)}{dt} = -\frac{s_j^{\text{GABA}}(t)}{\tau_{\text{GABA}}} + \sum_k \delta(t - t_j^k) \tag{11}$$

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

The values of the conductances for pyramidal neurons were $g_{\text{AMPA,ext}} = 2.08$, $g_{\text{AMPA,rec}} = 0.052$, $g_{\text{NMDA,rec}} = 0.164$, and $g_{\text{GABA}} = 0.67$ and for interneurons were $g_{\text{AMPA,ext}} = 1.62$, $g_{\text{AMPA,rec}} = 0.0405$, $g_{\text{NMDA,rec}} = 0.129$, and $g_{\text{GABA}} = 0.49$.

Acknowledgments

Gustavo Deco acknowledges support through the German Ministry for Research, BMBF Grant 01IBC01A; and through the European Union, grant IST-2001-38099.

Reprint requests should be sent to Edmund Rolls via e-mail (Edmund.Rolls@psy.ox.ac.uk, http://www.cns.ox.ac.uk) or to Barry Horwitz, Brain Imaging and Modeling Section, National Institute on Deafness and other Communication Disorders, National Institutes of Health, Bethesda, MD 20892, USA.

REFERENCES

Abeles, A. (1991). *Corticonics.* New York: Cambridge University Press.

Amit, D., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex, 7,* 237–252.

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

Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology, 84,* 451–459.

Baddeley, A. (1986). *Working memory.* New York: Oxford University Press.

Battaglia, F., & Treves, A. (1998). Stable and rapid recurrent processing in realistic autoassociative memories. *Neural Computation, 10,* 431–450.

Brunel, N., & Wang, X. (2001). Effects of neuromodulation in a cortical networks model of object working memory dominated by recurrent inhibition. *Journal of Computational Neuroscience, 11,* 63–85.

Buxton, R. B., & Frank, L. R. (1997). A model for the coupling

between cerebral blood flow and oxygen metabolism during neural stimulation. *Journal of Cerebral Blood Flow and Metabolism, 17,* 64–72.

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine, 39,* 855–864.

Chelazzi, L. (1998). Serial attention mechanisms in visual search: A critical look at the evidence. *Psychological Research, 62,* 195–219.

Chelazzi, L., Miller, E., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature (London), 363,* 345–347.

Corchs, S., & Deco, G. (2002). Large-scale neural model for visual attention: Integration of experimental single cell and fMRI data. *Cerebral Cortex, 12,* 339–348.

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

Deco, G., & Zihl, J. (2001). Top-down selective visual attention: A neurodynamical approach. *Visual Cognition, 8,* 119–140.

D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., & Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Cognitive Brain Research, 7,* 1–13.

Destexhe, A., Mainen, Z., & Sejnowski, T. (1998). Kinetic models of synaptic transmission. In C. Koch & I. Segev (Eds.), *Methods in neural modeling: From ions to networks* (2nd ed., pp. 1–25). Cambridge: MIT Press.

Funahashi, S., Bruce, C., & Goldman-Rakic, P. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61,* 331–349.

Fuster, J. (2000). Executive frontal functions. *Experimental Brain Research, 133,* 66–70.

Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1982). Cellular discharge in the dorsolateral prefrontal cortex of the monkey in cognitive tasks. *Experimental Neurology, 77,* 679–694.

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

Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions? Interpreting data from the Tower of Hanoi. *Neuropsychologia, 33,* 632–642.

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

Goldman-Rakic, P. (1995). Cellular basis of working memory. *Neuron, 14,* 477–485.

Goldman-Rakic, P. (1996). Regional and cellular fractionation of working memory. *Proceedings of the National Academy of Sciences, U.S.A. 93,* 13473–13480.

Hansel, D., Mato, G., Meunier, C., & Neltner, L. (1998). On numerical simulations of integrate-and-fire neural networks. *Neural Computation, 10,* 467–483.

Hestrin, S., Sah, P., & Nicoll, R. (1990). Mechanisms generating the time course of dual component excitatory synaptic currents recorded in hippocampal slices. *Neuron, 5,* 247–253.

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

Horwitz, B., Friston, K. J., & Taylor, J. G. (2000). Neural modeling and functional brain imaging: An overview. *Neural Networks, 13,* 829–846.

Horwitz, B., Tagamets, M.-A., & McIntosh, A. R. (1999). Neural modeling, functional brain imaging, and cognition. *Trends in Cognitive Sciences, 3,* 85–122.

Hoshi, E., Shima, K., & Tanji, J. (1998). Task-dependent selectivity of movement-related neuronal activity in the primate prefrontal cortex. *Journal of Neurophysiology, 80,* 3392–3397.

Jahr, C., & Stevens, C. (1990). Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics. *Journal of Neuroscience, 10,* 3178–3182.

Jueptner, M., & Weiller, C. (1995). Does measurement of regional cerebral blood flow reflect synaptic activity?—Implications for PET and fMRI. *Neuroimage, 2,* 148–156.

Lauritzen, M. (2001). Relationship of spikes, synaptic activity, and local changes of cerebral blood flow. *Journal of Cerebral Blood Flow and Metabolism, 21,* 1367–1383.

Leung, H., Gore, J., & Goldman-Rakic, P. (2002). Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *Journal of Cognitive Neuroscience, 14,* 659–671.

Levy, R., & Goldman-Rakic, P. (1999). Executive frontal functions. *Journal of Neuroscience, 19,* 5149–5158.

Lisman, J. E., Fellous, J. M., & Wang, X. J. (1998). A role for NMDA-receptor channels in working memory. *Nature Neuroscience, 1,* 273–275.

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

McCormick, D., Connors, B., Lighthall, J., & Prince, D. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons in the neocortex. *Journal of Neurophysiology, 54,* 782–806.

Miller, E., Gochin, P., & Gross, C. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Research, 616,* 25–29.

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience, 1,* 59–65.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229,* 782–784.

Motter, B. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology, 70,* 909–919.

Owen, A. M., Herrod, N. J., Menon, D. K., Clark, C. J., Downey, S. P. M. J., Carpenter, T. A., Minhas, P. S., Turkheimer, F. E., Williams, E. J., Robbins, T. W., Sahakian, B. J., Petrides, M., & Pickard, J. (1999). Redefining the functional organization of working memory processes within human lateral prefrontal cortex. *European Journal of Neuroscience, 11,* 567–574.

Petrides, M. (1994). Frontal lobes and behaviour. *Current Opinion in Neurobiology, 4,* 207–211.

Postle, B. R., & D'Esposito, M. (1999). "What"-then-"Where" in visual working memory: An event-related fMRI study. *Journal of Cognitive Neuroscience, 11,* 585–597.

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

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

Reynolds, J., & Desimone, R. (1999). The role of neural

mechanisms of attention in solving the binding problem. *Neuron, 24,* 19–29.

Rolls, E. T. (1999). *The brain and emotion.* Oxford: Oxford University Press.

Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision.* Oxford: Oxford University Press.

Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function.* Oxford: Oxford University Press.

Rolls, E. T., Stringer, S. M., & Trappenberg, T. P. (2002). A unified model of spatial and episodic memory. *Proceedings of the Royal Society of London. Series B, 269,* 1087–1093.

Salin, P., & Prince, D. (1996). Spontaneous GABA-A receptor mediated inhibitory currents in adult rat somatosensory cortex. *Journal of Neurophysiology, 75,* 1573–1588.

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

Spruston, N., Jonas, P., & Sakmann, B. (1995). Dendritic glutamate receptor channel in rat hippocampal CA3 and CA1 pyramidal neurons. *Journal of Physiology, 482,* 325–352.

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

Ungerleider, L., Courtney, S., & Haxby, J. (1998). A neural system for human visual working memory. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 883–890.

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

Wang, X. (1999). Synaptic basis of cortical persistent activity: The importance of NMDA receptors to working memory. *Journal of Neuroscience, 19,* 9587–9603.

White, I., & Wise, S. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Experimental Brain Research, 126,* 315–335.

Wilson, F. A. W., O'Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science, 260,* 1955–1958.

Wilson, F. A. W., O'Scalaidhe, S. P., & Goldman-Rakic, P. (1994). Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 91,* 4009–4013.

Xiang, Z., Huguenard, J., & Prince, D. (1998). GABA-A receptor mediated currents in interneurons and pyramidal cells of rat visual cortex. *Journal of Physiology, 506,* 715–730.