

Attention in natural scenes: Neurophysiological and computational bases

Edmund T. Rolls^{a,*}, Gustavo Deco^b

^a University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, United Kingdom

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

Received 11 May 2006; accepted 1 August 2006

Abstract

How does attention operate in natural scenes? We show that the receptive fields of inferior temporal cortex neurons that implement object representations become small and located at the fovea in complex natural scenes. This facilitates the readout of information about an object that may be reward or punishment associated, and may be the target for action. Top-down biased competition to implement attention has a much weaker effect in complex natural scenes than in otherwise blank scenes with two objects. Part of the solution to the binding problem is thus that competition and the foveal cortical magnification factor emphasize what is present at the fovea, and limit the binding problem. Part of the solution to the binding problem is that neurons respond to combinations of features present in the correct relative spatial positions. Stimulus-dependent neuronal synchrony does not appear to be quantitatively important in feature binding, and in attention, in natural visual scenes, at least in the inferior temporal visual cortex, as shown by information theoretic analyses. The perception of multiple objects in a scene is facilitated by the fact that inferior temporal visual cortex neurons have asymmetrical receptive fields with respect to the fovea in complex scenes. Computational models of this processing are described.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Attention; Biased competition; Natural scenes; Receptive field size; Inferior temporal visual cortex

We have produced a unified theory of visual attention and working memory, and how these processes are influenced by rewards to influence decision making (Deco & Rolls, 2005a, 2006; Rolls & Deco, 2002). The theory is based on the 'Biased Competition Hypothesis' (Chelazzi, 1998; Chelazzi, Miller, Duncan, & Desimone, 1993; Miller, Gochin, & Gross, 1993; Moran & Desimone, 1985; Motter, 1993; Reynolds & Desimone, 1999; Rolls & Deco, 2002; Spitzer, Desimone, & Moran, 1988). In this approach (Deco & Rolls, 2005a; Rolls & Deco, 2002) multiple activated populations of neurons which may be hierarchically organized engage in competitive interactions, and external top-down effects bias this competition in favor of specific neurons (see Section 4). This approach leaves open however how object and spatial attention operate in complex natural scenes, the neural encoding of information about objects, and how multiple objects and feature binding are implemented in complex scenes. These are the main subjects of this paper.

1. Object-based attention and object selection in complex natural scenes

1.1. Neurophysiology of object selection in the inferior temporal visual cortex

Object-based attention refers to attention to an object. For example, in a visual search task the object might be specified as what should be searched for, and its location must be found. In spatial attention, a particular location in a scene is pre-cued, and the object at that location may need to be identified.

Much of the neurophysiology, psychophysics, and modelling of attention has been with a small number, typically two, of objects in an otherwise blank scene. In this section, we consider how attention operates in complex natural scenes, and in particular describe how the inferior temporal visual cortex operates to enable the selection of an object in a complex natural scene. The inferior temporal visual cortex contains distributed and invariant representations of objects and faces (Booth & Rolls, 1998; Hasselmo, Rolls, & Baylis, 1989; Rolls, 2000, in press; Rolls & Baylis, 1986; Rolls & Deco, 2002; Rolls & Tovee, 1995; Rolls, Treves, & Tovee, 1997; Tovee, Rolls, & Azzopardi, 1994).

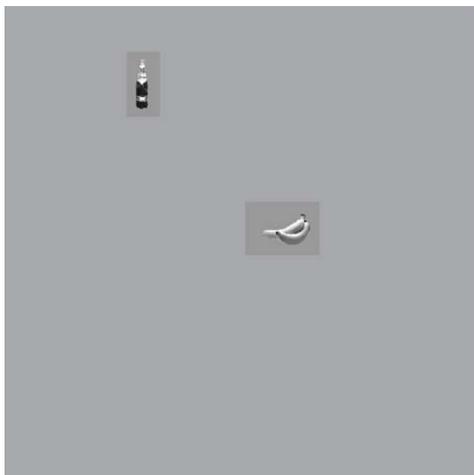
* Corresponding author. Tel.: +44 1865 271348; fax: +44 1865 310447.
E-mail address: Edmund.Rolls@psy.ox.ac.uk (E.T. Rolls).
URL: <http://www.cns.ox.ac.uk> (E.T. Rolls).

To investigate how attention operates in complex natural scenes, and how information is passed from the inferior temporal cortex (IT) to other brain regions to enable stimuli to be selected from natural scenes for action, [Rolls, Aggelopoulos, and Zheng \(2003\)](#) analyzed the responses of inferior temporal cortex neurons to stimuli presented in complex natural backgrounds. The monkey had to search for two objects on a screen, and a touch of one object was rewarded with juice, and of another object was punished with saline (see [Fig. 1](#) for a schematic illustration and [Fig. 2](#) for a version of the display with examples of the stimuli shown to scale). Neuronal responses to the effective stimuli for the neurons were compared when the objects were presented in the natural scene or on a plain background. It was found that the overall response of the neuron to objects was hardly reduced when they were presented in natural scenes, and the selectivity of the neurons remained. However, the main finding was that the magnitudes of the responses of the neurons typically became much less in the real scene the further the monkey fixated in the scene away from the object (see [Fig. 3](#)).

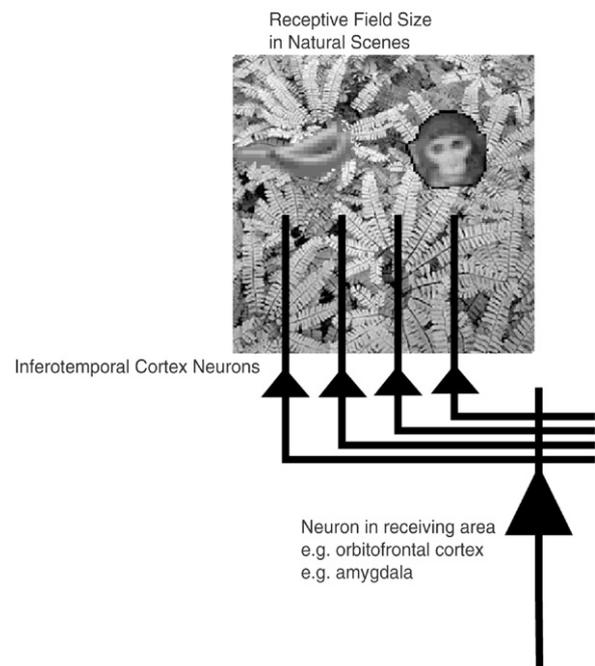
It is proposed that this reduced translation invariance in natural scenes helps an unambiguous representation of an object which may be the target for action to be passed to the brain regions which receive from the primate inferior temporal visual cortex. It helps with the binding problem, by reducing in natural scenes the effective receptive field of at least some inferior temporal cortex neurons to approximately the size of an object in the scene.

It is also found that in natural scenes, the effect of object-based attention on the response properties of inferior temporal cortex neurons is relatively small, as illustrated in [Fig. 4](#) ([Rolls, Aggelopoulos et al., 2003](#)).

The results summarized in [Fig. 4](#) for 5° stimuli show that the receptive fields were large (77.6°) with a single stimulus in a blank background (top left), and were greatly reduced in size (to 22.0°) when presented in a complex natural scene (top right). The results also show that there was little difference in receptive field size or firing rate in the complex background



[Fig. 2](#). The visual search task. The monkey had to search for and touch an object (in this case a banana) when shown in a complex natural scene, or when shown on a plain background. In each case a second object is present (a bottle) which the monkey must not touch. The stimuli are shown to scale. The screen subtended $70^\circ \times 55^\circ$.



[Fig. 1](#). Objects shown in a natural scene, in which the task was to search for and touch one of the stimuli. The objects in the task as run were smaller. The diagram shows that if the receptive fields of inferior temporal cortex neurons are large in natural scenes with multiple objects, then any receiving neuron in structures such as the orbitofrontal cortex and amygdala would receive information from many stimuli in the field of view, and would not be able to provide evidence about each of the stimuli separately.

when the effective stimulus was selected for action (bottom right, 19.2°), and when it was not (middle right, 15.6°) ([Rolls, Aggelopoulos et al., 2003](#)). (For comparison, the effects of attention against a blank background were much larger, with the receptive field increasing from 17.2° to 47.0° as a result of object-based attention, as shown in [Fig. 4](#), left middle and bottom.) The computational basis for these relatively minor effects of object-based attention when objects are viewed in natural scenes is considered in [Section 1.2](#).

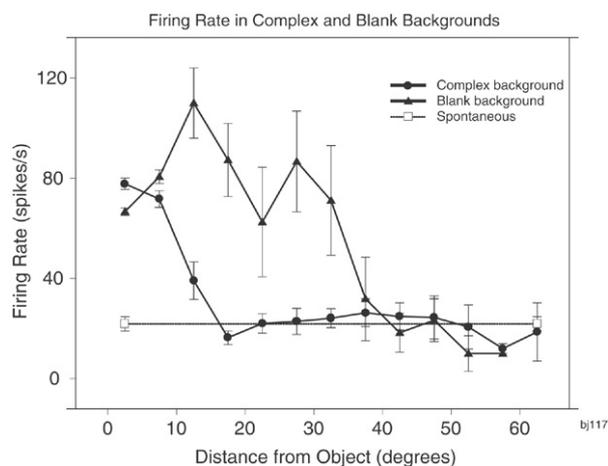


Fig. 3. Firing of a temporal cortex cell to an effective stimulus presented either in a blank background or in a natural scene, as a function of the angle in degrees at which the monkey was fixating away from the effective stimulus. The task was to search for and touch the stimulus. (After Rolls, Aggelopoulos et al. (2003)).

1.2. Attention in natural scenes — a computational account

The results described in Section 1.1 and summarized in Fig. 4 show that the receptive fields of inferior temporal cortex neurons were large (77.6°) with a single stimulus in a blank background (top left), and were greatly reduced in size (to 22.0°) when presented in a complex natural scene (top right). The results also show that there was little difference in receptive field size or firing rate in the complex background when the effective stimulus was selected for action (bottom right), and when it was not (middle right) (Rolls, Aggelopoulos et al., 2003).

Trappenberg, Rolls, and Stringer (2002) have suggested what underlying mechanisms could account for these findings, and simulated a model to test the ideas. The model utilizes an attractor network representing the inferior temporal visual cortex (implemented by the recurrent excitatory connections between inferior temporal cortex neurons), and a neural input layer with several retinotopically organized modules representing the visual scene in an earlier visual cortical area such as V4 (see Fig. 5). The attractor network aspect of the model produces the property that receptive fields of IT neurons can be large in blank scenes by enabling a weak input in the periphery of the visual field to act as a retrieval cue for the object attractor. On the other hand, when the object is shown in a complex background, the object closest to the fovea tends to act as the retrieval cue for the attractor, because the fovea is given increased weight in activating the IT module because the magnitude of the input activity from objects at the fovea is greatest due to the cortical higher magnification factor of the fovea incorporated into the model. (The cortical magnification factor can be expressed as the number of mm of cortex representing 1° of visual field. The cortical magnification factor decreases rapidly with increasing eccentricity from the fovea (Cowey & Rolls, 1975; Rolls & Cowey, 1970).) This results in smaller receptive fields of IT neurons in complex scenes, because the object tends to need to be close to the fovea

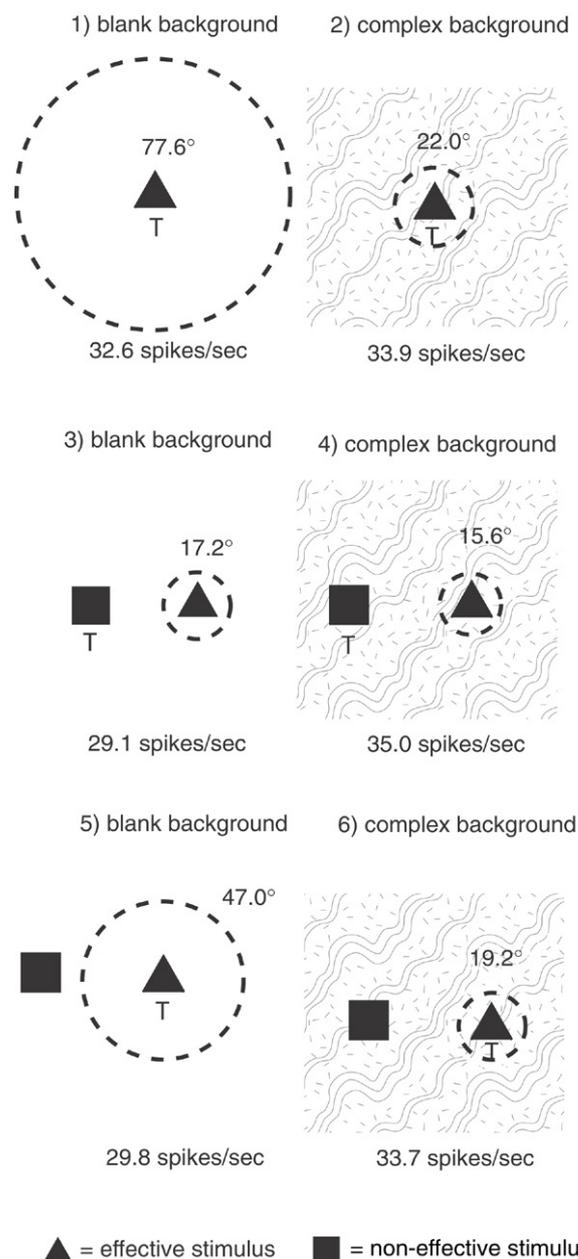


Fig. 4. Summary of the receptive field sizes of inferior temporal cortex neurons to a 5° effective stimulus presented in either a blank background (blank screen) or in a natural scene (complex background). The stimulus that was a target for action in the different experimental conditions is marked by T. When the target stimulus was touched, a reward was obtained. The mean receptive field diameter of the population of neurons analyzed, and the mean firing rate in spikes/s, is shown. The stimuli subtended $5^\circ \times 3.5^\circ$ at the retina, and occurred on each trial in a random position in the $70^\circ \times 55^\circ$ screen. The dashed circle is proportional to the receptive field size. Top row: responses with one visual stimulus in a blank (left) or complex (right) background. Middle row: responses with two stimuli, when the effective stimulus was not the target of the visual search. Bottom row: responses with two stimuli, when the effective stimulus was the target of the visual search. (After Rolls, Aggelopoulos et al. (2003).)

to trigger the attractor into the state representing that object. (In other words, if the object is far from the fovea, then it will not trigger neurons in IT which represent it, because neurons in IT are preferentially being activated by another object at the fovea.) This may be described as an attractor model in which

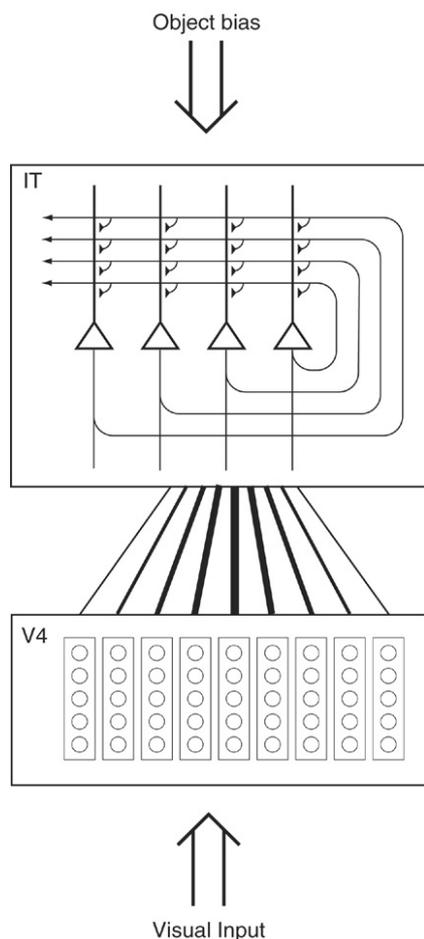


Fig. 5. The architecture of the inferior temporal cortex (IT) model of Trappenberg et al. (2002) operating as an attractor network with inputs from the fovea given preferential weighting by the greater magnification factor of the fovea. The model also has a top-down object-selective bias input. The model was used to analyze how object vision and recognition operate in complex natural scenes.

the competition for which attractor state is retrieved is weighted towards objects at the fovea.

Attentional top-down object-based inputs can bias the competition implemented in this attractor model, but have relatively minor effects (in for example increasing receptive field size) when they are applied in a complex natural scene, as then as usual the stronger forward inputs dominate the states reached. In this network, the recurrent collateral connections may be thought of as implementing constraints between the different inputs present, to help arrive at firing in the network which best meets the constraints. In this scenario, the preferential weighting of objects close to the fovea because of the increased magnification factor at the fovea is a useful principle in enabling the system to provide useful output. The attentional object biasing effect is much more marked in a blank scene, or a scene with only two objects present at similar distances from the fovea, which are conditions in which attentional effects have frequently been examined. The results of the investigation (Trappenberg et al., 2002) thus suggest that attention may be a much more limited phenomenon in complex, natural, scenes than in reduced displays with one or two objects

present. The results also suggest that the alternative principle, of providing strong weight to whatever is close to the fovea, is an important principle governing the operation of the inferior temporal visual cortex, and in general of the output of the ventral visual system in natural environments. This principle of operation is very important in interfacing the visual system to action systems, because the effective stimulus in making inferior temporal cortex neurons fire is in natural scenes usually on or close to the fovea. This means that the spatial coordinates of where the object is in the scene do not have to be represented in the inferior temporal visual cortex, nor passed from it to the action selection system, as the latter can assume that the object making IT neurons fire is close to the fovea in natural scenes (see Rolls and Deco (2002) and Rolls, Aggelopoulos et al. (2003)).

There may of course be in addition a mechanism for object selection that takes into account the locus of covert attention when actions are made to locations not being looked at. However, the simulations described in this section suggest that in any case covert attention is likely to be a much less significant influence on visual processing in natural scenes than in reduced scenes with one or two objects present.

Given these points, one might question why inferior temporal cortex neurons can have such large receptive fields, which show translation invariance (Rolls, 2000; Rolls, Aggelopoulos et al., 2003). At least part of the answer to this may be that inferior temporal cortex neurons must have the capability to be large if they are to deal with large objects (Rolls & Deco, 2002). A V1 neuron, with its small receptive field, simply could not receive input from all the features necessary to define an object. On the other hand, inferior temporal cortex neurons may be able to adjust their size to approximately the size of objects, using in part the interactive attentional bottom-up and top-down effects described in Section 4.

The implementation of the simulations is described by Trappenberg et al. (2002), and some of the results obtained with the architecture shown in Fig. 5 follow. In one simulation only one object was present in the visual scene in a plain background at different eccentricities from the fovea. As shown in Fig. 6(A) by the line labelled 'simple background', the receptive fields of the neurons were very large. The value of the object bias k^{ITBIAS} was set to 0 in these simulations. Good object retrieval (indicated by large correlations) was found even when the object was far from the fovea, indicating large IT receptive fields with a blank background. The reason that any drop is seen in performance as a function of eccentricity is because some noise was present in the recall process. This demonstrates that the attractor dynamics can support translation invariant object recognition even though the translation invariant weight vectors between V4 and IT are explicitly modulated by the modulation factor k^{IT-V4} derived from the cortical magnification factor.

In a second simulation individual objects were placed at all possible locations in a natural and cluttered visual scene. The resulting correlations between the target pattern and the asymptotic IT state are shown in Fig. 6(A) with the line labelled 'natural background'. Many objects in the visual scene are now competing for recognition by the attractor network, and

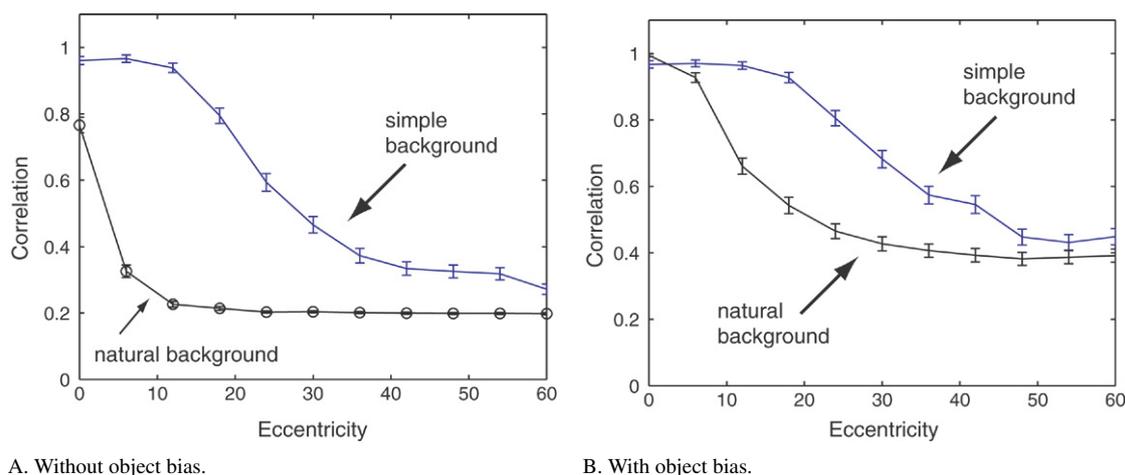


Fig. 6. Correlations as measured by the normalized dot product between the object vector used to train IT and the state of the IT network after settling into a stable state with a single object in the visual scene (blank background) or with other trained objects at all possible locations in the visual scene (natural background). There is no object bias included in the results shown in graph A, whereas an object bias is included in the results shown in B with $k^{\text{ITBIAS}} = 0.7$ in the experiments with a natural background and $k^{\text{ITBIAS}} = 0.1$ in the experiments with a blank background.

the objects around the foveal position are enhanced through the modulation factor derived from the cortical magnification factor. This results in a much smaller size of the receptive field of IT neurons when measured with objects in natural backgrounds.

In addition to this major effect of the background on the size of the receptive field, which parallels and we suggest may account for the physiological findings outlined above, there is also a dependence of the size of the receptive fields on the level of object bias provided to the IT network. Examples are shown in Fig. 6(B) where an object bias was used. The object bias biases the IT network towards the expected object with a strength determined by the value of k^{ITBIAS} , and has the effect of increasing the size of the receptive fields in both blank and natural backgrounds (see Fig. 6(B) compared to (A)). This models the effect found neurophysiologically (Rolls, Aggelopoulos et al., 2003).

Some of the conclusions are as follows. When single objects are shown in a scene with a blank background, the attractor network helps neurons to respond to an object with large eccentricities of this object relative to the fovea. When the object is presented in a natural scene, other neurons in the inferior temporal cortex become activated by the other effective stimuli present in the visual field, and these forward inputs decrease the response of the network to the target stimulus by a competitive process. The results found fit well with the neurophysiological data, in that IT operates with almost complete translation invariance when there is only one object in the scene, and reduces the receptive field size of its neurons when the object is presented in a cluttered environment. The model described here provides an explanation of the responses of real IT neurons in natural scenes.

In natural scenes, the model is able to account for the neurophysiological data that the IT neuronal responses are larger when the object is close to the fovea, by virtue of the fact that objects close to the fovea are weighted by the cortical magnification factor related modulation $k^{\text{IT-V4}}$.

The model accounts for the larger receptive field sizes from the fovea of IT neurons in natural backgrounds if the target is the object being selected compared to when it is not selected (Rolls, Aggelopoulos et al., 2003). The model accounts for this by an effect of top-down bias which simply biases the neurons towards particular objects compensating for their decreasing inputs produced by the decreasing magnification factor modulation with increasing distance from the fovea. Such object-based attention signals could originate in the prefrontal cortex and could provide the object bias for the inferotemporal cortex (Renart, Parga, & Rolls, 2000).

Important properties of the architecture for obtaining the results just described are the high magnification factor at the fovea and the competition between the effects of different inputs, implemented in the above simulation by the competition inherent in an attractor network.

We have also been able to obtain similar results in a hierarchical feedforward network where each layer operates as a competitive network (Deco & Rolls, 2004). This network thus captures many of the properties of our hierarchical model of invariant object recognition (Elliffe, Rolls, & Stringer, 2002; Rolls, 1992; Rolls & Deco, 2002; Rolls & Milward, 2000; Rolls & Stringer, 2001, in press-a, in press-b; Stringer, Perry, Rolls, & Proske, 2006; Stringer & Rolls, 2000, 2002; Wallis & Rolls, 1997), but incorporates in addition a foveal magnification factor and top-down projections with a dorsal visual stream so that attentional effects can be studied, as shown in Fig. 7.

Deco and Rolls (2004) trained the network described shown in Fig. 7 with two objects, and used the trace learning rule (Rolls & Milward, 2000; Wallis & Rolls, 1997) in order to achieve translation invariance. In a first experiment we placed only one object on the retina at different distances from the fovea (i.e. different eccentricities relative to the fovea). This corresponds to the blank background condition. In a second experiment, we also placed the object at different eccentricities relative to the fovea, but on a cluttered natural background.

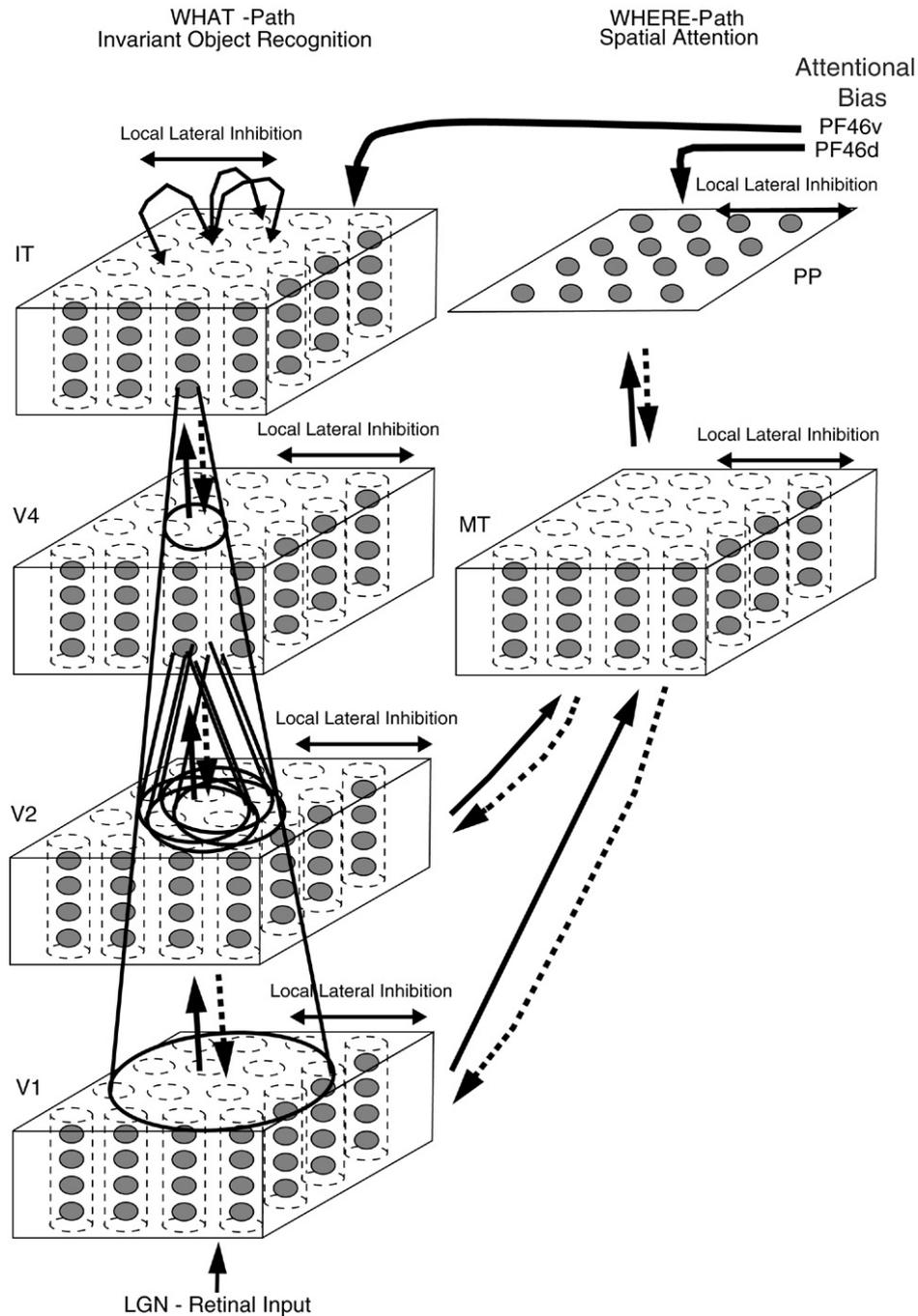


Fig. 7. Cortical architecture for hierarchical and attention-based visual perception (Deco & Rolls, 2004). The system is essentially composed of five modules structured such that they resemble the two known main visual paths of the mammalian visual cortex. Information from the retino-geniculo-striate pathway enters the visual cortex through area V1 in the occipital lobe and proceeds into two processing streams. The occipital-temporal stream leads ventrally through V2–V4 and IT (inferior temporal visual cortex), and is mainly concerned with object recognition. The occipito-parietal stream leads dorsally into PP (posterior parietal complex), and is responsible for maintaining a spatial map of an object's location. The solid lines with arrows between levels show the forward connections, and the dashed lines the top-down backprojections. Short term memory systems in the prefrontal cortex (PF46) apply top-down attentional bias to the object or spatial processing streams.

Fig. 8 shows the average firing activity of the inferior temporal cortex neuron specific for the test object as a function of the position of the object on the retina relative to the fovea (eccentricity). In both cases relatively large receptive fields are observed, because of the translation invariance obtained with the trace learning rule and the competition mechanisms implemented within each layer of the ventral stream. (The

receptive field size is defined as the width of the receptive field at the point where there is a half-maximal response.) However, when the object was in a blank background, larger receptive fields were observed. The decrease in neuronal response as a function of distance from the fovea is mainly due to the effect of the magnification factor implemented in V1. On the other hand, when the object was in a complex cluttered background, the

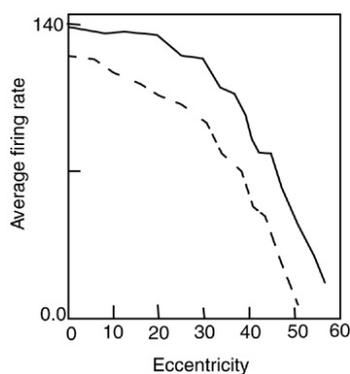
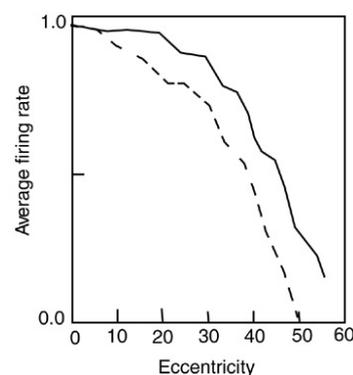


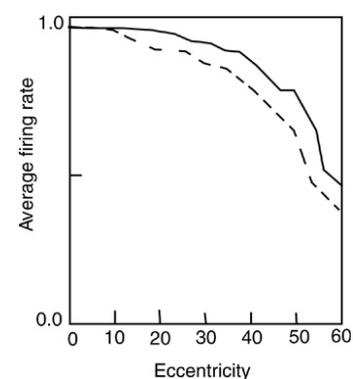
Fig. 8. Average firing activity of an inferior temporal cortex neuron as a function of eccentricity from the fovea, in the simulation of Deco and Rolls (2004). When the object was in a blank background (solid line), large receptive fields are observed because of the translation invariance of inferior temporal neurons. The decay is mainly due to the magnification factor implemented in V1. When the object was presented in a complex cluttered natural background (dashed line), the effective size of the receptive field of the same inferior temporal neuron was reduced because of competitive effect between the object features and the background features within each layer of the ventral stream.

effective size of the receptive field of the same inferior temporal cortex neuron shrinks because of competitive effects between the object features and the background features in each layer of the ventral stream. In particular, the global character of the competition expressed in the inferior temporal cortex module (due to the large receptive fields and the local character of the inhibition, in our simulations, between the two object specific pools) is the main cause of the reduction of the receptive field sizes in the complex scene.

Deco and Rolls (2004) also studied the influence of object-based attentional top-down bias on the effective size of an inferior temporal cortex neuron for the case of an object in a blank or a cluttered background. To do this, we repeated the two simulations but now considered a non-zero top-down bias coming from prefrontal area 46v and impinging on the inferior temporal cortex neuron specific for the object tested. Fig. 9 shows the results. We plot the average firing activity normalized to the maximum value to compare the neuronal activity as a function of the eccentricity. When no attentional object bias is introduced (a), a shrinkage of the receptive field is observed. When attentional object bias is introduced (b), the shrinkage of the receptive field due to the complex background is slightly reduced. Rolls, Aggelopoulos et al. (2003) also found that in natural scenes, the effect of object-based attention on the response properties of inferior temporal cortex neurons was relatively small. They found only a small difference in the receptive field size or firing rate in the complex background when the effective stimulus was selected for action, vs when it was not. In the framework of the model (Deco & Rolls, 2004), the reduction of the shrinkage of the receptive field is due to the biasing of the competition in the inferior temporal cortex layer in favour of the specific IT neuron tested, so that it shows more translation invariance (i.e. a slightly larger receptive field). The increase of the receptive field of an IT neuron, although small, produced by the external top-down attentional bias offers



A. Without object attention.



B. With object attention.

Fig. 9. Influence of object-based attentional top-down bias from prefrontal area 46v on the effective size of an inferior temporal cortex neuron for the case of an object in a blank (solid line) or a cluttered (dashed line) background. The average firing activity was normalized to the maximum value in order to compare the neuronal activity as a function of the eccentricity. When no attentional object bias was introduced (a), a reduction of the receptive field was observed. When attentional object bias was introduced (b), the reduction of the receptive field size due to the complex background was slightly reduced. (After Deco and Rolls (2004).)

a mechanism for facilitation of the search for specific objects in complex natural scenes.

2. Information representation, attention, and feature binding in the inferior temporal visual cortex without stimulus-dependent temporal synchrony

In the hierarchical model of object recognition we have developed, feature binding is implemented by neurons at each level of the hierarchy that respond to combinations of features in the correct relative spatial position (Elliffe et al., 2002; Rolls, 1992; Rolls & Deco, 2002; Rolls & Milward, 2000; Wallis & Rolls, 1997). In the models of attention on which we have been working, the attention operates in biased competition networks to bias the firing rates of neurons (Deco & Rolls, 2005a, 2005b; Rolls & Deco, 2002). This biasing of firing rates is completely consistent with the information theoretic analyses we have performed which indicate that information is encoded mainly by the firing rates of inferior temporal cortex neurons (Franco, Rolls, Aggelopoulos, & Treves, 2004; Rolls, Aggelopoulos,

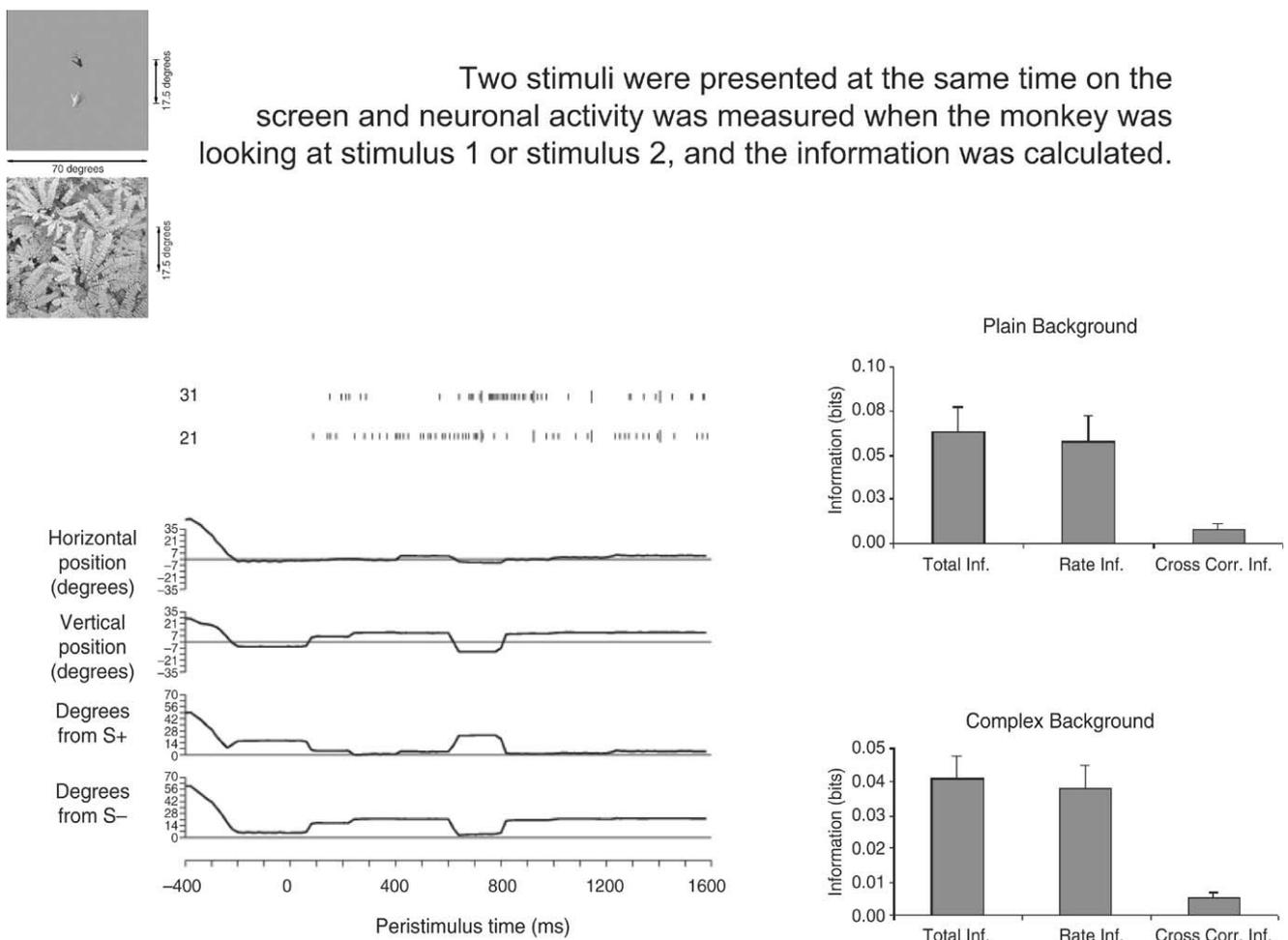


Fig. 10. (Right) The information available from the firing rates (Rate Inf) or from stimulus-dependent synchrony (Cross-Corr Inf) from populations of simultaneously recorded inferior temporal cortex neurons about which stimulus had been presented in a complex natural scene. The total information (Total Inf) is that available from both the rate and the stimulus-dependent synchrony, which do not necessarily contribute independently. (Left) Eye position recordings and spiking activity from two neurons on a single trial of the task. (Neuron 31 tended to fire more when the macaque looked at one of the stimuli, S-, and neuron 21 tended to fire more when the macaque looked at the other stimulus, S+. Both stimuli were within the receptive field of the neuron.) (After Aggelopoulos et al. (2005).)

Franco, & Treves, 2004; Rolls, Franco, Aggelopoulos, & Reece, 2003; Rolls, Treves, & Tovee, 1997; Rolls, Treves, Tovee, & Panzeri, 1997; Tovee & Rolls, 1995; Tovee, Rolls, Treves, & Bellis, 1993). In contrast, a different hypothesis suggests that feature binding and attention is implemented by stimulus-dependent synchronization of neurons (Malsburg, 1990, Chap. 18; Malsburg & Schneider, 1986; Singer, 1999, 2000, Chap. 23).

To obtain direct evidence on whether stimulus-dependent synchrony is important in encoding information in a system in which it should apply, Aggelopoulos, Franco, and Rolls (2005) analyzed the activity of simultaneously recorded neurons using an object-based attention task in which macaques searched for a target object to touch in a complex natural scene. In the task, object-based attention was required as the macaque knew which of the two objects he was searching for. Feature binding was required in that two objects (each requiring correct binding of the features of that object but not the other object) were present, and segmentation was required to segment the objects from their background. This is a real-world task with natural visual scenes in which if temporal

synchrony was important in neuronal encoding, it should be present. Information theoretic techniques were used to assess how much information was provided by the firing rates of the neurons about the stimuli, and how much by the stimulus-dependent cross-correlations between the firing of different neurons that were sometimes present. The use of information theoretic procedures was important, for it allowed the relative contributions of rates and stimulus-dependent synchrony to be quantified (Franco et al., 2004). It was found that between 99% and 94% of the information was present in the firing rates of inferior temporal cortex neurons, and less than 5% in any stimulus-dependent synchrony that was present, as illustrated in Fig. 10 (Aggelopoulos et al., 2005). The implication of these results is that any stimulus-dependent synchrony that is present is not quantitatively important as measured by information theoretic analyses under natural scene conditions. This has been found for inferior temporal cortex, a brain region where features are put together to form representations of objects (Rolls & Deco, 2002), and where attention has strong effects, at least in scenes with blank backgrounds (Rolls, Aggelopoulos et al. (2003). The finding as assessed

by information theoretic methods of the importance of firing rates and not stimulus-dependent synchrony is consistent with previous information theoretic approaches (Franco et al., 2004; Rolls et al., 2004; Rolls, Franco, et al., 2003). It would of course also be of interest to test the same hypothesis in earlier visual areas, such as V4, with quantitative, information theoretic, techniques. In connection with rate codes, it should be noted that this implies the number of spikes that arrive in a given time is important, and that this time can be very short, as little as 20–50 ms, for very useful amounts of information to be made available from a population of neurons (Rolls, 2003; Rolls & Deco, 2002; Rolls, Franco, Aggelopoulos, & Jerez, in press; Rolls & Tovee, 1994; Rolls, Tovee, & Panzeri, 1999; Rolls, Tovee, Purcell, Stewart, & Azzopardi, 1994; Tovee & Rolls, 1995).

The implications of these findings for the computational bases of attention are important. First the findings indicate that top-down attentional biasing inputs could, by providing biasing inputs to the appropriate object-selective neurons, facilitate bottom-up information about objects without any need to alter the time relations between the firing of different neurons. The neurons to which the top-down biases should be applied could in principle be learned by simple Hebbian associativity between the source of the biasing signals, in for example the prefrontal cortex, and the inferior temporal cortex neurons (Rolls & Deco, 2002). Thus rate encoding would be sufficient for the whole system to implement attention, a conclusion supported by the spiking network model of attention of Deco and Rolls (2005b), in which non-linear interactions between top-down and bottom-up signals without specific temporal encoding can implement the details of the interactions found neurophysiologically in V4 and V2. Second, the findings are consistent with the hypothesis that feature binding is implemented by neurons that respond to features in the correct relative spatial locations (Elliffe et al., 2002; Rolls & Deco, 2002), and not by temporal synchrony and attention (Abeles, 1991; Hummel & Biederman, 1992; Malsburg, 1990, Chap. 18; Singer, 1999; Singer et al., 1990; Singer & Gray, 1995).

With respect to the synchrony model, Malsburg (1990, Chap. 18) suggested that features that should be bound together would be linked by temporal binding. There has been considerable neurophysiological investigation of this possibility (Abeles, 1991; Hummel & Biederman, 1992; Singer, 1999; Singer et al., 1990; Singer & Gray, 1995). We note that a problem with this approach is that temporal binding might enable features 1, 2 and 3, which might define one stimulus to be bound together and kept separate from for example another stimulus consisting of features 2, 3 and 4, but would require a further temporal binding (leading in the end potentially to a combinatorial explosion) to indicate the relative spatial positions of the 1, 2 and 3 in the 123 stimulus, so that it can be discriminated from e.g. 312. Thus temporal synchrony could it seems at best be useful for grouping features (e.g. features 1, 2 and 3 are part of object 1, and features 4, and 6 are part of object 2), but would not without a great deal more in the way of implementation be useful to encode the relative spatial positions of features within an object, or of objects in a scene.

3. The representation of information about the relative positions of multiple objects in a scene using a rate code

These experiments have been extended to address the issue of how several objects are represented in a complex scene. The issue arises because the relative spatial locations of objects in a scene must be encoded (and is possible even in short presentation times without eye movements (Biederman, 1972)) (and this has been held to involve some spotlight of attention); and because as shown above what is represented in complex natural scenes is primarily about what is at the fovea, yet we can locate more than one object in a scene even without eye movements. Aggelopoulos and Rolls (2005) showed that with five objects simultaneously present in the receptive field of inferior temporal cortex neurons, although all the neurons responded to their effective stimulus when it was at the fovea, some could also respond to their effective stimulus when it was in a parafoveal position 10° from the fovea. An example of such a neuron is shown in Fig. 11. The asymmetry is much more evident in a scene with 5 images present (Fig. 11A) than when only one image is shown on an otherwise blank screen (Fig. 11B). Competition between different stimuli in the receptive field thus reveals the asymmetry in the receptive field of inferior temporal visual cortex neurons.

The asymmetry provides a way of encoding the position of multiple objects in a scene. Depending on which asymmetric neurons are firing, the population of neurons provides information to the next processing stage not only about which image is present at or close to the fovea, but where it is with respect to the fovea. This information is provided by neurons that have firing rates that reflect the relevant information, and stimulus-dependent synchrony is not necessary. Top-down attentional biasing input could thus, by biasing the appropriate neurons, facilitate bottom-up information about objects without any need to alter the time relations between the firing of different neurons. The exact position of the object with respect to the fovea, and effectively thus its spatial position relative to other objects in the scene, would then be made evident by the subset of asymmetric neurons firing.

This is thus the solution that these experiments indicate is used to the representation of multiple objects in a scene, an issue that has previously been difficult to account for in neural systems with distributed representations Mozer (1991) and for which ‘attention’ has been a proposed solution.

4. A biased competition model of object and spatial attention

Visual attention exerts top-down influences on the processing of sensory information in the visual cortex, and therefore is intrinsically associated with intercortical neural interactions. Thus, elucidating the neural basis of visual attention is an excellent paradigm for understanding the basic mechanisms of intercortical neurodynamics. Recent cognitive neuroscience developments allow a more direct study of the neural mechanisms underlying attention in humans and primates. In particular, the seminal work of Chelazzi et al. (1993) has led to a promising

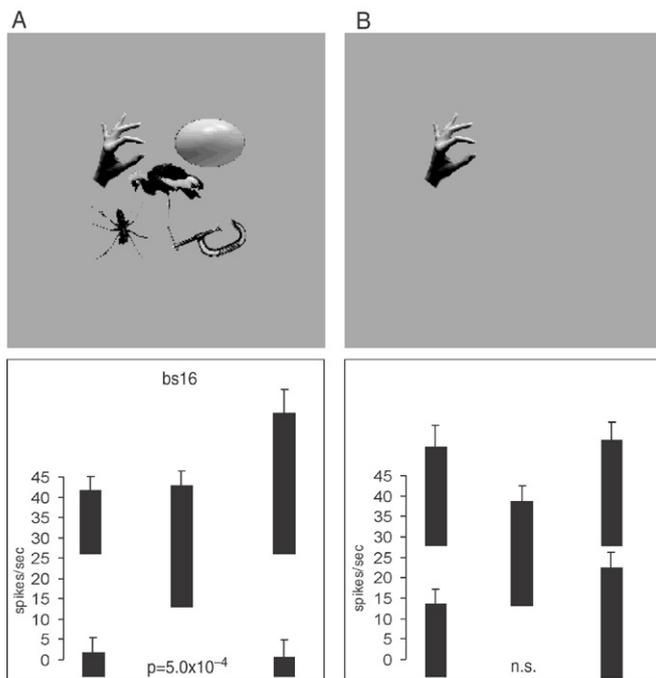


Fig. 11. A. The responses (firing rate with the spontaneous rate subtracted, means sem) of one neuron when tested with 5 stimuli simultaneously present in close (10°) configuration with the parafoveal stimuli located 10° from the fovea. B. The responses of the same neuron when only the effective stimulus was presented in each position. The firing rate for each position is that when the effective stimulus for the neuron was in that position. The p value is that from the ANOVA calculated over the four parafoveal positions. (After Aggelopoulos and Rolls (2005).)

account of attention termed the ‘*biased competition hypothesis*’ (see also Duncan (1996), Moran and Desimone (1985) and Reynolds and Desimone (1999)). According to this hypothesis, attentional selection operates in parallel by biasing an underlying competitive interaction between multiple stimuli in the visual field toward one stimulus or another, so that behaviorally relevant stimuli are processed in the cortex while irrelevant stimuli are filtered out. Thus, attending to a stimulus at a particular location or with a particular feature biases the underlying neural competition in a certain brain area in favour of neurons that respond to the location, or the features, of the attended stimulus.

Neurodynamical models for biased competition have been proposed and successfully applied in the context of attention and working memory. In the context of attention, Usher and Niebur (1996) introduced an early model of biased competition. Deco and Zihl (2001) extended Usher and Niebur’s model to simulate the psychophysics of visual attention by visual search experiments in humans. Their neurodynamical formulation is a large-scale hierarchical model of the visual cortex whose global dynamics is based on biased competition mechanisms at the neural level. Attention then appears as an emergent effect related to the dynamical evolution of the whole network. This large-scale formulation has been able to simulate and explain in a unifying framework visual attention in a variety of tasks and at different cognitive neuroscience experimental measurement levels, namely: single-

cells (Deco & Lee, 2002; Rolls & Deco, 2002), fMRI (Corchs & Deco, 2002, 2004), psychophysics (Deco, Pollatos, & Zihl, 2002; Deco & Rolls, 2005a), and neuropsychology (Deco & Rolls, 2002). In the context of working memory, further developments (Deco, Rolls, & Horwitz, 2004; Szabo, Almeida, Deco, & Stetter, 2004) managed to model in a unifying form attentional and memory effects in the prefrontal cortex, integrating single-cell and fMRI data, and different paradigms in the framework of biased competition.

In particular Deco and Rolls (2005b) extended previous concepts of the role of biased competition in attention (Desimone & Duncan, 1995; Duncan, 1996; Usher & Niebur, 1996) by providing the first analysis at the integrate-and-fire neuronal level, which allows the neuronal non-linearities in the system to be explicitly modelled, in order to investigate realistically the processes that underlie the apparent gain modulation effect of top-down attentional control. In the integrate-and-fire model, the competition is implemented realistically by the effects of the excitatory neurons on the inhibitory neurons, and their return inhibitory synaptic connections. That was also the first integrate-and-fire analysis of top-down attentional influences in vision that explicitly models the interaction of several different brain areas. Part of the originality of the model is that in the form in which it can account for attentional effects in V2 and V4 in the paradigms of Reynolds, Chelazzi, and Desimone (1999) in the context of biased competition, the model with the same parameters effectively makes predictions which show that the ‘contrast gain’ effects in MT of Martinez-Trujillo and Treue (2002) can be accounted for by the same model.

These detailed and quantitative analyses of neuronal dynamical systems are an important step towards understanding the operation of complex processes such as top-down attention, which necessarily involve the interaction of several brain areas. They are being extended to provide neurally plausible models of decision-making (Deco & Rolls, 2003, 2005c, 2006; Rolls, 2005).

Acknowledgements

The first author’s research was supported by the Medical Research Council, the Wellcome Trust, and the Oxford McDonnell Centre for Cognitive Neuroscience.

The second author was supported by ICREA.

References

- Abeles, A. (1991). *Corticonics*. New York: Cambridge University Press.
- Aggelopoulos, N. C., Franco, L., & Rolls, E. T. (2005). Object perception in natural scenes: Encoding by inferior temporal cortex simultaneously recorded neurons. *Journal of Neurophysiology*, *93*, 1342–1357.
- Aggelopoulos, N. C., & Rolls, E. T. (2005). Natural scene perception: Inferior temporal cortex neurons encode the positions of different objects in the scene. *European Journal of Neuroscience*, *22*, 2903–2916.
- Biederman, I. (1972). Perceiving real-world scenes. *Science*, *177*, 77–80.
- Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex*, *8*, 510–523.

- 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.
- Corchs, S., & Deco, G. (2004). Feature-based attention in human visual cortex: Simulation of fMRI data. *Neuroimage*, *21*, 36–45.
- Cowey, A., & Rolls, E. T. (1975). Human cortical magnification factor and its relation to visual acuity. *Experimental Brain Research*, *21*, 447–454.
- Deco, G., & Lee, T. S. (2002). A unified model of spatial and object attention based on inter-cortical biased competition. *Neurocomputing*, *44–46*, 775–781.
- Deco, G., Pollatos, O., & Zihl, J. (2002). The time course of selective visual attention: Theory and experiments. *Vision Research*, *42*, 2925–2945.
- Deco, G., & Rolls, E. T. (2002). Object-based visual neglect: A computational hypothesis. *European Journal of Neuroscience*, *16*, 1994–2000.
- Deco, G., & Rolls, E. T. (2003). Attention and working memory: A dynamical model of neuronal activity in the prefrontal cortex. *European Journal of Neuroscience*, *18*, 2374–2390.
- Deco, G., & Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*, *44*, 621–644.
- Deco, G., & Rolls, E. T. (2005a). Attention, short term memory, and action selection: A unifying theory. *Progress in Neurobiology*, *76*, 236–256.
- Deco, G., & Rolls, E. T. (2005b). Neurodynamics of biased competition and cooperation for attention: A model with spiking neurons. *Journal of Neurophysiology*, *94*, 295–313.
- Deco, G., & Rolls, E. T. (2005c). Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. *Cerebral Cortex*, *15*, 15–30.
- Deco, G., & Rolls, E. T. (2006). A neurophysiological model of decision-making and Weber's law. *European Journal of Neuroscience*, *24*, 901–916.
- Deco, G., Rolls, E. T., & Horwitz, B. (2004). 'What' and 'where' in visual working memory: A computational neurodynamical perspective for integrating fMRI and single-neuron data. *Journal of Cognitive Neuroscience*, *16*, 683–701.
- Deco, G., & Zihl, J. (2001). Top-down selective visual attention: A neurodynamical approach. *Visual Cognition*, *8*, 119–140.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Duncan, J. (1996). Cooperating brain systems in selective perception and action. In T. Inui, & J. L. McClelland (Eds.), *Attention and performance XVI* (pp. 549–578). Cambridge, MA: MIT Press.
- Elliffré, M. C. M., Rolls, E. T., & Stringer, S. M. (2002). Invariant recognition of feature combinations in the visual system. *Biological Cybernetics*, *86*, 59–71.
- Franco, L., Rolls, E. T., Aggelopoulos, N. C., & Treves, A. (2004). The use of decoding to analyze the contribution to the information of the correlations between the firing of simultaneously recorded neurons. *Experimental Brain Research*, *155*, 370–384.
- Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, *32*, 203–218.
- Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, *99*, 480–517.
- Malsburg, C. v. d. (1990). A neural architecture for the representation of scenes. In J. L. McGaugh, N. M. Weinberger, & G. Lynch (Eds.), *Brain organization and memory: Cells, systems and circuits* (pp. 356–372). New York: Oxford University Press.
- Malsburg, C. v. d., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, *54*, 29–40.
- Martinez-Trujillo, J., & Treue, S. (2002). Attentional modulation strength in cortical area MT depends on stimulus contrast. *Neuron*, *35*, 365–370.
- 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.
- 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.
- Mozer, M. C. (1991). *The perception of multiple objects: A connectionist approach*. Cambridge, MA: MIT Press.
- Renart, A., Parga, N., & Rolls, E. T. (2000). A recurrent model of the interaction between the prefrontal cortex and inferior temporal cortex in delay memory tasks. In S. Solla, T. Leen, & K. -R. Mueller (Eds.), *Advances in neural information processing systems: Vol. 12* (pp. 171–177). Cambridge, MA: MIT Press.
- Reynolds, J., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*, 1736–1753.
- Reynolds, J., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, *24*, 19–29.
- Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society*, *335*, 11–21.
- Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron*, *27*, 205–218.
- Rolls, E. T. (2003). Consciousness absent and present: A neurophysiological exploration. *Progress in Brain Research*, *144*, 95–106.
- Rolls, E. T. (2005). *Emotion explained*. Oxford: Oxford University Press.
- Rolls, E. T. (2006). The representation of information about faces in the temporal and frontal lobes of primates including humans. *Neuropsychologia*, in press (doi:10.1016/j.neuropsychologia.2006.04.019).
- Rolls, E. T., Aggelopoulos, N. C., Franco, L., & Treves, A. (2004). Information encoding in the inferior temporal visual cortex: Contributions of the firing rates and the correlations between the firing of neurons. *Biological Cybernetics*, *90*, 19–32.
- Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *Journal of Neuroscience*, *23*, 339–348.
- Rolls, E. T., & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Experimental Brain Research*, *65*, 38–48.
- Rolls, E. T., & Cowey, A. (1970). Topography of the retina and striate cortex and its relationship to visual acuity in rhesus monkeys and squirrel monkeys. *Experimental Brain Research*, *10*, 298–310.
- Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision*. Oxford: Oxford University Press.
- Rolls, E. T., Franco, L., Aggelopoulos, N. C., & Jerez, J. M. (2006). Information in the first spike, the order of spikes, and the number of spikes provided by neurons in the inferior temporal visual cortex. *Vision Research* (in press).
- Rolls, E. T., Franco, L., Aggelopoulos, N. C., & Reece, S. (2003). An information theoretic approach to the contributions of the firing rates and the correlations between the firing of neurons. *Journal of Neurophysiology*, *89*, 2810–2822.
- Rolls, E. T., & Milward, T. (2000). A model of invariant object recognition in the visual system: Learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation*, *12*, 2547–2572.
- Rolls, E. T., & Stringer, S. M. (2001). Invariant object recognition in the visual system with error correction and temporal difference learning. *Network: Computation in Neural Systems*, *12*, 111–129.
- Rolls, E. T., & Stringer, S. M. (2006). Invariant global motion recognition in the dorsal visual system: A unifying theory. *Neural Computation* (in press-a).
- Rolls, E. T., & Stringer, S. M. (2007). Invariant visual object recognition: A model, with lighting invariance. *Journal de Physiologie, Paris* (in press-b).
- Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society, B* *257*, 9–15.
- Rolls, E. T., & Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, *73*, 713–726.
- Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, *11*, 335–346.

- Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research*, *101*, 474–484.
- Rolls, E. T., Treves, A., & Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. *Experimental Brain Research*, *114*, 149–162.
- Rolls, E. T., Treves, A., Tovee, M., & Panzeri, S. (1997). Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *Journal of Computational Neuroscience*, *4*, 309–333.
- Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*, 49–65.
- Singer, W. (2000). Response synchronisation: A universal coding strategy for the definition of relations. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 325–338). Cambridge, MA: MIT Press.
- Singer, W., Gray, C., Engel, A., Konig, P., Artola, A., & Brocher, S. (1990). Formation of cortical cell assemblies. *Cold Spring Harbor Symposium on Quantitative Biology*, *55*, 939–952.
- Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, *18*, 555–586.
- Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, *240*, 338–340.
- Stringer, S. M., Perry, G., Rolls, E. T., & Proske, J. H. (2006). Learning invariant object recognition in the visual system with continuous transformations. *Biological Cybernetics*, *94*, 128–142.
- Stringer, S. M., & Rolls, E. T. (2000). Position invariant recognition in the visual system with cluttered environments. *Neural Networks*, *13*, 305–315.
- Stringer, S. M., & Rolls, E. T. (2002). Invariant object recognition in the visual system with novel views of 3D objects. *Neural Computation*, *14*, 2585–2596.
- Szabo, M., Almeida, R., Deco, G., & Stetter, M. (2004). Cooperation and biased competition model can explain attentional filtering in the prefrontal cortex. *European Journal of Neuroscience*, *19*, 1969–1977.
- Tovee, M. J., & Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cognition*, *2*, 35–58.
- Tovee, M. J., Rolls, E. T., & Azzopardi, P. (1994). Translation invariance and the responses of neurons in the temporal visual cortical areas of primates. *Journal of Neurophysiology*, *72*, 1049–1060.
- Tovee, M. J., Rolls, E. T., Treves, A., & Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology*, *70*, 640–654.
- Trappenberg, T. P., Rolls, E. T., & Stringer, S. M. (2002). Effective size of receptive fields of inferior temporal visual cortex neurons in natural scenes. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems: Vol. 14* (pp. 293–300). Cambridge, MA: MIT Press.
- Usher, M., & Niebur, E. (1996). Modelling the temporal dynamics of IT neurons in visual search: A mechanism for top-down selective attention. *Journal of Cognitive Neuroscience*, *8*, 311–327.
- Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, *51*, 167–194.