Chapter 3
Invariant Representations of Objects in Natural Scenes in the Temporal Cortex Visual Areas

Edmund T. Rolls

1. Introduction

Evidence on how information about visual stimuli is represented in the temporal cortical visual areas and on how these representations are formed is described. The neurophysiological recordings are made mainly in nonhuman primates, macaques, first because the temporal lobe, in which this processing occurs, is much more developed than in nonprimates, and second because the findings are relevant to understanding the effects of brain damage in patients, as will be shown. In this chapter, attention is paid to neural systems involved in processing information about faces, because with the large number of neurons devoted to this class of stimuli, this system has proved amenable to experimental analysis; because of the importance of face recognition and expression identification in primate, including human, social and emotional behavior; and because of the application of understanding this neural system to understanding the effects of damage to this system in humans. It is also shown that the temporal cortical visual areas have neuronal populations that provide invariant representations of objects. Although there is some segregation of face identity and object identity representations in different cytoarchitectonic regions, the proportion of face-selective neurons in any one region reaches only 20%, so that no region is devoted exclusively to faces (see Section 2).

In Section 2, I show that there are two main populations of face-selective neurons in the temporal cortical visual areas. The first population is tuned to the identity of faces and has representations that are invariant with respect to, for example, retinal position, size, and even view. These invariant representations are ideally suited to provide the inputs to brain regions such as the orbitofrontal cortex and amygdala that learn the reinforcement associations of an individual's face, for then the learning, and the appropriate social and emotional responses, generalize to other views of the same face. Moreover, these inferior temporal cortex neurons have sparse distributed representations of faces, which are shown

University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford, OX1 3UD, England

Y2

to be well suited as inputs to the stimulus–reinforcer association learning mechanisms in the orbitofrontal cortex and amygdala that allow different emotional and social responses to be made to the faces of different individuals, depending on the reinforcers received. The properties of these neurons tuned to face identity or object identity are described in Sections 3–11. Section 12 describes a second main population of neurons that are in the cortex in the superior temporal sulcus, which encode other aspects of faces such as face expression, eye gaze, face view, and whether the head is moving. This second population of neurons thus provides important additional inputs to parts of the brain such as the orbitofrontal cortex and amygdala that are involved in social communication and emotional behavior. This second population of neurons may in some cases encode reinforcement value (e.g., face expression neurons), or provide social information that is very relevant to whether reinforcers will be received, such as neurons that signal eye gaze, or whether the head is turning toward or away from the receiver. Sections 13 and 14 show how the brain may learn these invariant representations of objects and faces. Section 15 shows how attention operates computationally in natural visual scenes, and Section 16 describes the biased competition approach to how attention can modulate representations in the brain. In Sections 17 and 18, I describe the representations of faces in two areas, the amygdala and orbitofrontal cortex, to which the temporal cortical areas have direct projections. I also review evidence (Section 18) that damage to the human orbitofrontal cortex can impair face (and voice) expression identification.

The orbitofrontal cortex is also shown to be involved in the rapid reversal of behavior to stimuli (which could be the face of an individual) when the reinforcement contingencies change, and therefore to have an important role in social and emotional behavior. Moreover, the human orbitofrontal cortex is shown to be activated in a simple model of human social interaction when a face expression change indicates that the face of a particular individual is no longer reinforcing. The representations in the orbitofrontal cortex are thus of the reward or affective value of the visual stimuli that are useful in emotional behavior, in contrast to the representations in the temporal cortical visual areas, where the representations that are built are primarily of the identity of the visual stimulus.

2. Neuronal Responses Found in Different Temporal Lobe Cortex Visual Areas

Visual pathways project by a number of cortico-cortical stages from the primary visual cortex until they reach the temporal lobe visual cortical areas (Baizer et al. 1991; Maunsell and Newsome 1987; Seltzer and Pandya 1978), in which some neurons that respond selectively to faces are found (Bruce et al. 1981; Desimone 1991; Desimone and Gross 1979; Desimone et al. 1984; Gross et al. 1985; Perrett et al. 1982; Rolls 1981, 1984, 1991, 1992a, 2000a, 2005, 2006; Rolls and Deco 2002). The inferior temporal visual cortex, area TE, is divided on the basis of cytoarchitecture, myeloarchitecture, and afferent input into areas TEa, TEm, TE3, TE2, and TE1. In addition, there is a set of different areas in the

Y2

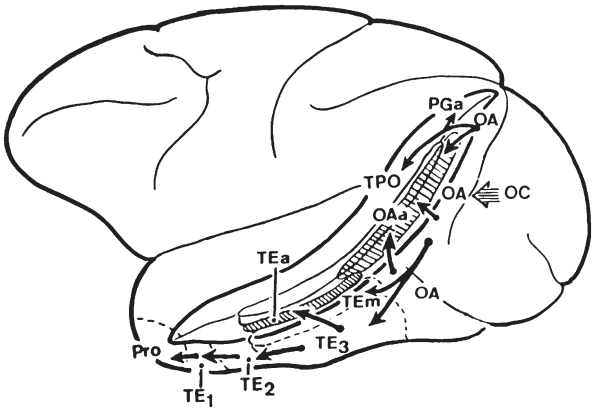


FIG. 1. Lateral view of the macaque brain (*left*) and coronal section (*right*) showing the different architectonic areas (e.g., *TEm*, *TPO*) in and bordering the anterior part of the superior temporal sulcus (*STS*) of the macaque (see text). (After Seltzer and Pandya 1978) (seltzer.eps) 4

cortex in the superior temporal sulcus (Baylis et al. 1987; Seltzer and Pandya 1978) (Fig. 1). Of these latter areas, TPO receives inputs from temporal, parietal, and occipital cortex; PGa and IPa from parietal and temporal cortex; and TS and TAa primarily from auditory areas (Seltzer and Pandya 1978).

Considerable specialization of function was found in recordings made from more than 2600 neurons in these architectonically defined areas (Baylis et al. 1987). Areas TPO, PGa, and IPa are multimodal, with neurons that respond to visual, auditory, and/or somatosensory inputs; the inferior temporal gyrus and adjacent areas (TE3, TE2, TE1, TEa, and TEm) are primarily unimodal visual areas; areas in the cortex in the anterior and dorsal part of the superior temporal sulcus (e.g., TPO, IPa, and IPg) have neurons specialized for the analysis of moving visual stimuli; and neurons responsive primarily to faces are found more frequently in areas TPO, TEa, and TEm, where they comprise approximately 20% of the visual neurons responsive to stationary stimuli, in contrast to the other temporal cortical areas, in which they comprise 4% to 10%. The stimuli that activate other cells in these TE regions include simple visual patterns such as gratings and combinations of simple stimulus features (Gross et al. 1985; Tanaka et al. 1990). Because face-selective neurons have a wide distribution, it might be expected that only large lesions, or lesions that interrupt outputs of these visual areas, would produce readily apparent face-processing deficits. Moreover, neurons with responses related to facial expression, movement, and gesture are more likely to be found in the cortex in the superior temporal sulcus, whereas neurons with activity related to facial identity are more likely to be found in the TE areas (Hasselmo et al. 1989a).

In human functional magnetic resonance imaging (fMRI) studies, evidence for specialization of function is described (Grill-Spector and Malach 2004; Haxby et al. 2002; Spiridon and Kanwisher 2002) related to face processing (in Y2

the fusiform face area, which may correspond to parts of the macaque inferior temporal visual cortex in which face neurons are common); to face expression and gesture (i.e., moving faces) (in the cortex in the superior temporal sulcus, which corresponds to the macaque cortex in the superior temporal sulcus); to objects (in an area that may correspond to the macaque inferior temporal cortex in which object but not face representations are common, as already described); and to spatial scenes (in a parahippocampal area, which probably corresponds to the macaque parahippocampal gyrus areas in which neurons are tuned to spatial view and to combinations of objects and the places in which they are located (Georges-François et al. 1999; Robertson et al. 1998; Rolls 1999c; Rolls and Kesner 2006; Rolls and Xiang 2005, 2006; Rolls et al. 1997b, 1998, 2005). However, there is much debate arising from these human fMRI studies about how specific each region is for a different type of function, in that such studies do not provide clear evidence on whether individual neurons can be very selective for face identity versus face expression versus objects and thereby convey specific information about these different classes of object; whether each area contains a mixture of different populations of neurons each tuned to different specific classes of visual stimuli, or neurons with relatively broad tuning that respond at least partly to different classes of stimuli; and about the fine-grain topology within a cortical area. The single-neuron studies in macaques described above and below do provide clear answers to these questions. The neuronal recording studies show that individual neurons can be highly tuned in that they convey information about face identity, or about face expression, or about objects, or about spatial view. The recording studies show that within these different classes, individual neurons by responding differently to different members of the class convey information about whose face it is, what the face expression is, etc., using a sparse distributed code with an approximately exponential firing rate probability distribution. The neuronal recording studies also show that each cytoarchitectonically defined area contains different proportions of face identity versus object neurons, but that the proportion of face-selective neurons in any one area is not higher than 20% of the visually responsive neurons in an area, so that considerable intermixing of specifically tuned neurons is the rule (Baylis et al. 1987). The neuronal recording studies also show that at the fine spatial scale, clusters of neurons extending for approximately 0.5–1 mm with tuning to one aspect of stimuli are common (e.g., face identity, or the visual texture of stimuli, or a particular class of head motion), and this can be understood as resulting from self-organizing mapping based on local cortical connectivity when a high dimensional space of objects, faces, etc., must be represented on a two-dimensional cortical sheet (Rolls and Deco 2002).

3. The Selectivity of One Population of Neurons for Faces

The neurons described in our studies as having responses selective for faces are selective in that they respond 2 to 20 times more (and statistically significantly more) to faces than to a wide range of gratings, simple geometric stimuli, or

complex three-dimensional (3-D) objects (Baylis et al. 1985, 1987; Rolls and Deco 2002; Rolls 1984, 1992a, 1997, 2000a, 2006). The recordings are made while the monkeys perform a visual fixation task in which, after the fixation spot has disappeared, a stimulus subtending typically 8° is presented on a video monitor (or, in some earlier studies, while monkeys perform a visual discrimination task). The responses to faces are excitatory, with firing rates often reaching 100 spikes/s, sustained, and have typical latencies of 80–100 ms. The neurons are typically unresponsive to auditory or tactile stimuli and to the sight of arousing or aversive stimuli. These findings indicate that explanations in terms of arousal, emotional or motor reactions, and simple visual feature sensitivity are insufficient to account for the selective responses to faces and face features observed in this population of neurons (Baylis et al. 1985; Perrett et al. 1982; Rolls and Baylis 1986). Observations consistent with these findings have been published by Desimone et al. (1984), who described a similar population of neurons located primarily in the cortex in the superior temporal sulcus that responded to faces but not to simpler stimuli such as edges and bars or to complex non-face stimuli (see also Gross et al. 1985).

These neurons are specialized to provide information about faces in that they provide much more information (on average, 0.4 bits) about which (of 20) face stimuli is being seen than about which (of 20) non-face stimuli is being seen (on average, 0.07 bits) (Rolls and Tovee 1995a; Rolls et al. 1997a). These information theoretical procedures provide an objective and quantitative way to show what is "represented" by a particular population of neurons, and indicate that different categories of visual stimulus are represented by different populations of inferior temporal cortex neurons (see also Hasselmo et al. 1989a).

4. The Selectivity of These Neurons for Individual Face Features or for Combinations of Face Features

Masking out or presenting parts of the face (e.g., eyes, mouth, or hair) in isolation reveal that different cells respond to different features or subsets of features. For some cells, responses to the normal organization of cut-out or line-drawn facial features are significantly larger than to images in which the same facial features are jumbled (Perrett et al. 1982; Rolls et al. 1994). These findings are consistent with the hypotheses developed below that by competitive self-organization some neurons in these regions respond to parts of faces by responding to combinations of simpler visual properties received from earlier stages of visual processing, and that other neurons respond to combinations of parts of faces and thus respond only to whole faces. Moreover, the finding that for some of these latter neurons the parts must be in the correct spatial configuration shows that the combinations formed can reflect not just the features present, but also their spatial arrangement; this provides a way in which binding can be implemented in neural networks (Elliffe et al. 2002; Rolls and Deco 2002). Further evidence that neurons in these regions respond to combinations of features in

the correct spatial configuration was found by Tanaka et al. (1990) using combinations of features that are used by comparable neurons to define objects.

5. Distributed Encoding of Face and Object Identity

An important question for understanding brain function is whether a particular object (or face) is represented in the brain by the firing of one or a few gnostic (or "grandmother") cells (Barlow 1972), or whether instead the firing of a population or ensemble of cells each with different profiles of responsiveness to the stimuli provides the representation. It has been shown that the representation of which particular object (face) is present is rather distributed. Baylis, Rolls, and Leonard (1985) showed this with the responses of temporal cortical neurons that typically responded to several members of a set of 5 faces, with each neuron having a different profile of responses to each face. In a further study using 23 faces and 45 non-face natural images, a distributed representation was again found (Rolls and Tovee 1995a), with the average sparseness being 0.65. The sparseness of the representation provided by a neuron can be defined as

$$a = \left(\sum_{s=1,S} r_s/S\right)^2 \Big/ \sum_{s=1,S} \left(r_s^2/S\right)$$

where r_s is the mean firing rate of the neuron to stimulus s in the set of S stimuli [see Rolls and Treves (1998) and Franco et al. (2006)]. If the neurons are binary (either firing or not to a given stimulus), then a would be 0.5 if the neuron responded to 50% of the stimuli and 0.1 if a neuron responded to 10% of the stimuli. If the spontaneous firing rate was subtracted from the firing rate of the neuron to each stimulus, so that the changes of firing rate, that is, the active responses of the neurons, were used in the sparseness calculation, then the "response sparseness" had a lower value, with a mean of 0.33 for the population of neurons.

The distributed nature of the representation can be further understood by the finding that the firing rate distribution of single neurons when a wide range of natural visual stimuli are being viewed is approximately exponentially distributed, with rather few stimuli producing high firing rates, and increasingly large numbers of stimuli producing lower and lower firing rates (Baddeley et al. 1997; Franco et al. 2006; Rolls and Tovee 1995a; Treves et al. 1999) (Fig. 2). The sparseness of such an exponential distribution of firing rates is 0.5. It has been shown that the distribution may arise from the threshold nonlinearity of neurons combined with short-term variability in the responses of neurons (Treves et al. 1999).

Complementary evidence comes from applying information theory to analyze how information is represented by a population of these neurons. The information required to identify which of S equiprobable events occurred (or stimuli were shown) is $\log_2 S$ bits. (Thus, 1 bit is required to specify which of 2 stimuli was shown, 2 bits to specify which of 4 stimuli was shown, 3 bits to specify which of 8 stimuli was shown, etc.) The important point for the present purposes is that

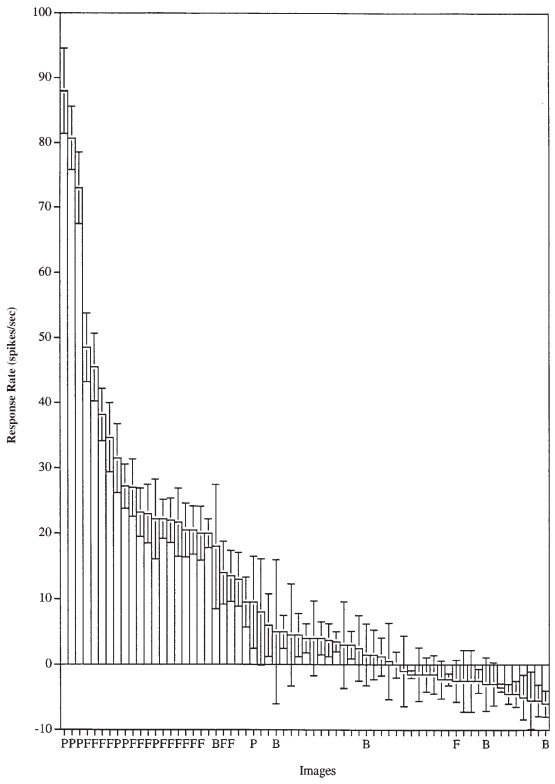


FIG. 2. Firing rate distribution of a single neuron in the temporal visual cortex to a set of 23 face (*F*) and 45 non-face images of natural scenes. The firing rate to each of the 68 stimuli is shown. *P*, a face profile stimulus; *B*, a body part stimulus such as a hand. (After Rolls and Tovee 1995a) (fratedist.eps)

if the encoding was local (or grandmother cell-like), then the number of stimuli encoded by a population of neurons would be expected to rise approximately linearly with the number of neurons in the population. In contrast, with distributed encoding, provided that the neuronal responses are sufficiently independent, and are sufficiently reliable (not too noisy), the number of stimuli encodable by the population of neurons might be expected to rise exponentially as the number of neurons in the sample of the population was increased. The information available about which of 20 equiprobable faces had been shown that was available from the responses of different numbers of these neurons is shown in Fig. 3. First, it is clear that some information is available from the responses of just one neuron, on average, approximately 0.34 bits. Thus, knowing the activity of just one neuron in the population does provide some evidence about which stimulus was present. This evidence that information is available in the responses of individual neurons in this way, without having to know the state of all the

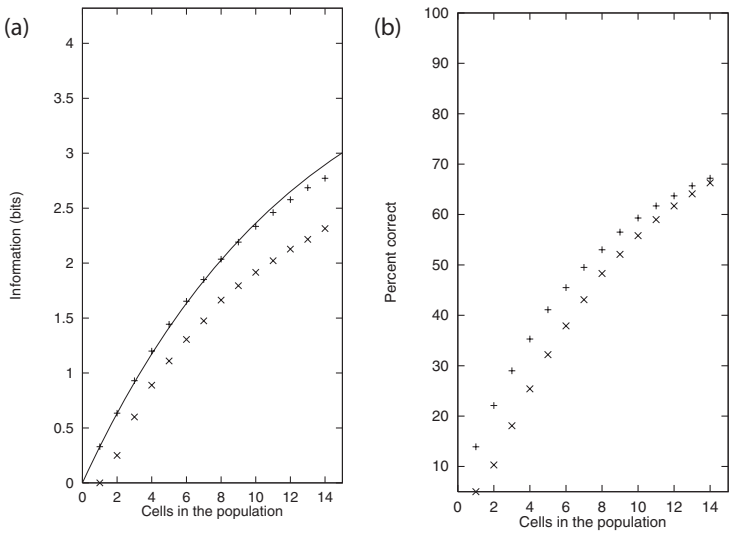


FIG. 3. **a** The values for the average information available in the responses of different numbers of these neurons on each trial, about which of a set of 20 face stimuli has been shown. The decoding method was dot product (DP, *x*) or probability estimation (*PE*, +), and the effects obtained with cross validation procedures utilizing 50% of the trials as test trials are shown. The remainder of the trials in the cross-validation procedure were used as training trials. The *full line* indicates the amount of information expected from populations of increasing size, when assuming random correlations within the constraint given by the ceiling (the information in the stimulus set, I = 4.32 bits). **b** The percent correct for the corresponding data to those shown in **a**. (After Rolls, Treves, and Tovee 1997) (multicellinfo20f.eps)

other neurons in the population, indicates that information is made explicit in the firing of individual neurons in a way that will allow neurally plausible decoding, involving computing a sum of input activities each weighted by synaptic strength, to work (see following). Second, it is clear (see Fig. 3) that the information rises approximately linearly, and the number of stimuli encoded thus rises approximately exponentially, as the number of cells in the sample increases (Abbott et al. 1996; Rolls and Treves 1998; Rolls et al. 1997a).

This direct neurophysiological evidence thus demonstrates that the encoding is distributed, and the responses are sufficiently independent and reliable, that the representational capacity increases exponentially with the number of neurons in the ensemble (Fig. 4). The consequence of this is that large numbers of stimuli, and fine discriminations between them, can be represented without having to measure the activity of an enormous number of neurons. [It has been shown that the main reason why the information tends to asymptote, as shown in Fig. 3, as the number of neurons in the sample increases is just that the ceiling is being approached of how much information is required to discriminate between the

Y2

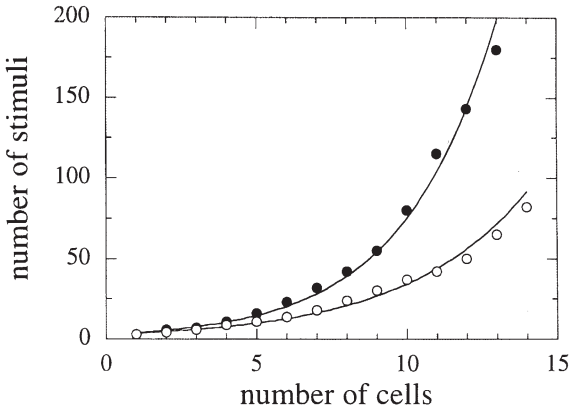


FIG. 4. The number of stimuli (in this case from a set of 20 faces) that are encoded in the responses of different numbers of neurons in the temporal lobe visual cortex, based on the results shown in Fig. 3. The decoding method was dot product (DP, *open circles*) or probability estimation (PE, *filled circles*). (After Rolls, Treves, and Tovee 1997; Abbott, Rolls, and Tovee 1996) (multicellexprot.eps)

set of stimuli, which with 20 stimuli is $\log_2 20 = 4.32$ bits (Abbott et al. 1996; Rolls et al. 1997a)].

It has in addition been shown that there are neurons in the inferior temporal visual cortex that encode view invariant representations of objects, and for these neurons the same type of representation is found, namely distributed encoding with independent information conveyed by different neurons (Booth and Rolls 1998).

The analyses just described were obtained with neurons that were not simultaneously recorded, but we have more recently shown that with simultaneously recorded neurons similar results are obtained, that is, the information about which stimulus was shown increases approximately linearly with the number of neurons, showing that the neurons convey information that is nearly independent (Panzeri et al. 1999b; Rolls et al. 2004). [Consistently, Gawne and Richmond (1993) showed that even adjacent pairs of neurons recorded simultaneously from the same electrode carried information that was approximately 80% independent.] In the research described by Panzeri et al. (1999b), Rolls et al. (2003b), and Franco et al. (2004), we developed methods for measuring the information in the relative time of firing of simultaneously recorded neurons, which might be significant if the neurons became synchronized to some but not other stimuli in a set, as postulated by Singer (1999). We found that for the set of cells currently available, almost all the information was available in the firing rates of the cells, and very little (not more than approximately 5% of the total information) was available about these static images in the relative time of firing of different simultaneously recorded neurons (Franco et al. 2004; Panzeri et al. 1999b; Rolls et al. 2003b, 2004). Thus, the evidence is that for representations of faces and objects Y2

in the inferior temporal visual cortex (and of space in the primate hippocampus and of odors in the orbitofrontal cortex; see Rolls et al. 1996, 1998), most of the information is available in the firing rates of the neurons.

To obtain direct evidence on whether stimulus-dependent synchrony is important in encoding information in natural and normal visual processing, we (Aggelopoulos et al. 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 theoretical 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 that 5% in any stimulus-dependent synchrony that was present, as illustrated in Fig. 5 (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 theoretical analyses under natural scene conditions; this has been found for the inferior temporal cortex, a brain region where features are put together to form representations of objects (Rolls and Deco 2002), and where attention has strong effects, at least in scenes with blank backgrounds (Rolls et al. 2003a). The finding as assessed by information theoretical 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. 2003b, 2004). It would of course also be of interest to test the same hypothesis in earlier visual areas, such as V4, with quantitative, information theoretical, techniques. In connection with rate codes, it should be noted that a rate code implies using the number of spikes that arrive in a given time, and that this time can be very short, as little as 20 to 50 ms, for very useful amounts of information to be made available from a population of neurons (Rolls 2003; Rolls and Tovee 1994; Rolls et al. 1994, 1999, 2006b; Tovee and Rolls 1995; Tovee et al. 1993).

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

Y2

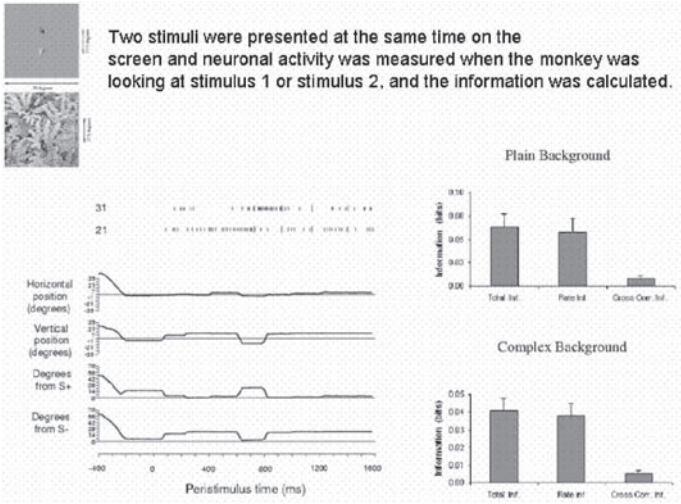


FIG. 5. *Right.* The information available from the firing rates (*Rate Inf*) or from stimulus-dependent synchrony (*Cross-Corr Inf*) from populations of simultaneously recorded inferior temporal cortex neurons about which stimulus had been presented in a complex natural scene. The total information (*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) (infoscene.eps)

associativity between the source of the biasing signals, in for example the prefrontal cortex, and the inferior temporal cortex neurons (Rolls and 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 (2005c), in which nonlinear 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 feature combination neurons which respond to features in the correct relative spatial locations (Elliffe et al. 2002; Rolls and Deco 2002), and not by temporal synchrony and attention (Singer 1999; Singer and Gray 1995; von der Malsburg 1990).

With respect to the synchrony model, von der Malsburg (1990) suggested that features that should be bound together would be linked by temporal binding. There has been considerable neurophysiological investigation of this possibility (Singer 1999; Singer and Gray 1995). 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

Y2

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 312, for example. 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.

It is unlikely that there are further processing areas beyond those described where ensemble coding changes into grandmother cell (local) encoding. Anatomically, there does not appear to be a whole further set of visual processing areas present in the brain; and outputs from the temporal lobe visual areas such as those described, are taken to limbic and related regions such as the amygdala and orbitofrontal cortex, and via the entorhinal cortex to the hippocampus, where associations between the visual stimuli and other sensory representations are formed (Rolls and Deco 2002; Rolls 2005). Indeed, tracing this pathway onward, we have found a population of neurons with face-selective responses in the amygdala (Leonard et al. 1985; Rolls 2000b) and orbitofrontal cortex (Rolls et al. 2006a), and in the majority of these neurons, different responses occur to different faces, with ensemble (not local) coding still being present. The amygdala in turn projects to another structure that may be important in other behavioral responses to faces, the ventral striatum, and comparable neurons have also been found in the ventral striatum (Williams et al. 1993).

6. Advantages of the Distributed Representation of Objects and Faces for Brain Processing

The advantages of the distributed encoding found are now considered, and apply to both fully distributed and to sparse distributed (but not to local) encoding schemes, as explained elsewhere (Rolls 2005; Rolls and Deco 2002; Rolls and Treves 1998).

6.1. Exponentially High Coding Capacity

This property arises from a combination of the encoding being sufficiently close to independent by the different neurons (i.e., factorial), and sufficiently distributed. Part of the biological significance of the exponential encoding capacity found is that a receiving neuron or neurons can obtain information about which one of a very large number of stimuli is present by receiving the activity of relatively small numbers of inputs (of the order of hundreds) from each of the neuronal populations from which it receives. In particular, the characteristics of the actual visual cells described here indicate that the activity of 15 would be able to encode 192 face stimuli (at 50% accuracy); of 20 neurons, 768 stimuli; of 25 neurons; 3 072 stimuli; of 30 neurons, 12 288 stimuli; and of 35 neurons, 49 152 stimuli (the values are for the optimal decoding case) (Abbott et al. 1996). Given

that most neurons receive a limited number of synaptic contacts, of the order of several thousand, this type of encoding is ideal. (It should be noted that the capacity of the distributed representations was calculated from ensembles of neurons each already shown to provide information about faces. If inferior temporal cortex neurons were chosen at random, 20 times as many neurons would be needed in the sample if face-selective neurons comprised 5% of the population. This brings the number of inputs required from an ensemble up to reasonable numbers given brain connectivity, a number of the order of the thousands of synapses being received by each neuron.) This type of encoding would enable, for example, neurons in the amygdala and orbitofrontal cortex to form pattern associations of visual stimuli with reinforcers such as the taste of food when each neuron received a reasonable number, perhaps in the order of hundreds, of inputs from the visually responsive neurons in the temporal cortical visual areas which specify which visual stimulus or object is being seen (Rolls 1990, 1992a,b; Rolls and Deco 2002; Rolls and Treves 1998). It is useful to realize that although the sensory representation may have exponential encoding capacity, this does not mean that the associative networks that receive the information can store such large numbers of different patterns. Indeed, there are strict limitations on the number of memories that associative networks can store (Rolls and Treves 1990, 1998; Treves and Rolls 1991). The particular value of the exponential encoding capacity of sensory representations is that very fine discriminations can be made, as there is much information in the representation, and that the representation can be decoded if the activity of even a limited number of neurons in the representation is known.

One of the underlying themes here is the neural representation of faces and objects. How would one know that one had found a neuronal representation of faces or objects in the brain? The criterion suggested (Rolls and Treves 1998) is that when one can identify the face or object that is present (from a large set of stimuli, which might be thousands or more) with a realistic number of neurons, say of the order of 100, and with some invariance, then one has a useful representation of the object.

The properties of the representation of faces, of objects (Booth and Rolls 1998), and of olfactory and taste stimuli, have been evident when the readout of the information was by measuring the firing rate of the neurons, typically over a 20-, 50-, or 500-ms period. Thus, at least where objects are represented in the visual, olfactory, and taste systems (e.g., individual faces, odors, and tastes), information can be read out without taking into account any aspects of the possible temporal synchronization between neurons, or temporal encoding within a spike train (Aggelopoulos et al. 2005; Franco et al. 2004; Panzeri et al. 1999b; Rolls et al. 1997a, 2003b, 2004; Tovee et al. 1993).

6.2. Ease with Which the Code Can Be Read by Receiving Neurons

For brain plausibility, it is also a requirement that neurons should be able to read the code. This is why when we have estimated the information from populations

of neurons, we have used in addition to a probability estimating measure (PE, optimal, in the Bayesian sense), also a dot product measure, which is a way of specifying that all that is required of decoding neurons would be the property of adding up postsynaptic potentials produced through each synapse as a result of the activity of each incoming axon (Abbott et al. 1996; Rolls et al. 1997a). It was found that with such a neurally plausible algorithm (the dot product, DP, algorithm), which calculates which average response vector the neuronal response vector on a single test trial was closest to by performing a normalized dot product (equivalent to measuring the angle between the test and the average vector), the same generic results were obtained, with only a 40% reduction of information compared to the more efficient (PE) algorithm. This is an indication that the brain could utilize the exponentially increasing capacity for encoding stimuli as the number of neurons in the population increases. For example, by using the representation provided by the neurons described here as the input to an associative or autoassociative memory, which computes effectively the dot product on each neuron between the input vector and the synaptic weight vector, most of the information available would in fact be extracted (Franco et al. 2004; Rolls and Deco 2002; Rolls and Treves 1990, 1998; Treves and Rolls 1991).

6.3. Higher Resistance to Noise

This, like the next few properties, is an advantage of distributed over local representations, which applies to artificial systems as well, but is presumably of particular value in biological systems in which some of the elements have an intrinsic variability in their operation. Because the decoding of a distributed representation involves assessing the activity of a whole population of neurons, and computing a dot product or correlation, a distributed representation provides more resistance to variation in individual components than does a local encoding scheme (Panzeri et al. 1996; Rolls and Deco 2002).

6.4. Generalization

Generalization to similar stimuli is again a property that arises in neuronal networks if distributed but not if local encoding is used. The generalization arises as a result of the fact that a neuron can be thought of as computing the inner or dot product of the stimulus representation with its weight vector. If the weight vector leads to the neuron having a response to one visual stimulus, then the neuron will have a similar response to a similar visual stimulus. This computation of correlations between stimuli operates only with distributed representations. If an output is based on a single input or output pair, then if either is lost, the correlation drops to zero (Rolls and Treves 1998; Rolls and Deco 2002).

6.5. Completion

Completion occurs in associative memory networks by a similar process. Completion is the property of recall of the whole of a pattern in response to any part

of the pattern. Completion arises because any part of the stimulus representation, or pattern, is effectively correlated with the whole pattern during memory storage. Completion is thus a property of distributed representations, and not of local representations. It arises, for example, in autoassociation (attractor) neuronal networks, which are characterized by recurrent connectivity. It is thought that such networks are important in the cerebral cortex, where the association fibers between nearby pyramidal cells may help the cells to retrieve a representation that depends on many neurons in the network (Rolls and Deco 2002; Rolls and Treves 1998).

6.6. Graceful Degradation or Fault Tolerance

This also arises only if the input patterns have distributed representations, and not if they are local. Local encoding suffers sudden deterioration once the few neurons or synapses carrying the information about a particular stimulus are destroyed.

6.7. Speed of Readout of the Information

The information available in a distributed representation can be decoded by an analyzer more quickly than can the information from a local representation, given comparable firing rates. Within a fraction of an interspike interval, with a distributed representation, much information can be extracted (Panzeri et al. 1999a; Rolls et al. 1997a; Rolls et al. 2006b; Treves 1993; Treves et al. 1996, 1997). In effect, spikes from many different neurons can contribute to calculating the angle between a neuronal population and a synaptic weight vector within an interspike interval (Franco et al. 2004; Rolls and Deco 2002). With local encoding, the speed of information readout depends on the exact model considered, but if the rate of firing needs to be taken into account, this will necessarily take time, because of the time needed for several spikes to accumulate in order to estimate the firing rate.

7. Invariance in the Neuronal Representation of Stimuli

One of the major problems that must be solved by a visual system is the building of a representation of visual information that allows recognition to occur relatively independently of size, contrast, spatial frequency, position on the retina, angle of view, etc. This is required so that if the receiving associative networks (in, e.g., the amygdala, orbitofrontal cortex, and hippocampus) learn about one view, position, etc., of the object, the animal generalizes correctly to other positions or views of the object. It has been shown that the majority of face-selective inferior temporal cortex neurons have responses that are relatively invariant with respect to the size of the stimulus (Rolls and Baylis 1986). The median size change tolerated with a response of greater than half the maximal response was Y2

12 times. Also, the neurons typically responded to a face when the information in it had been reduced from 3-D to a 2-D representation in gray on a monitor, with a response which was on average 0.5 of that to a real face. Another transform over which recognition is relatively invariant is spatial frequency. For example, a face can be identified when it is blurred (when it contains only low spatial frequencies), and when it is high-pass spatial frequency filtered (when it looks like a line drawing). It has been shown that if the face images to which these neurons respond are low-pass filtered in the spatial frequency domain (so that they are blurred), then many of the neurons still respond when the images contain frequencies only up to 8 cycles per face. Similarly, the neurons still respond to high-pass filtered images (with only high spatial frequency edge information) when frequencies down to only 8 cycles per face are included (Rolls et al. 1985). Face recognition shows similar invariance with respect to spatial frequency (Rolls et al. 1985). Further analysis of these neurons with narrow (octave) bandpass spatial frequency filtered face stimuli shows that the responses of these neurons to an unfiltered face can not be predicted from a linear combination of their responses to the narrow band stimuli (Rolls et al. 1987). This lack of linearity of these neurons, and their responsiveness to a wide range of spatial frequencies, indicate that in at least this part of the primate visual system recognition does not occur using Fourier analysis of the spatial frequency components of images.

Inferior temporal visual cortex neurons also often show considerable translation (shift) invariance, not only under anesthesia (see Gross et al. 1985), but also in the awake behaving primate (Tovee et al. 1994). It was found that in most cases the responses of the neurons were little affected by which part of the face was fixated, and that the neurons responded (with a greater than half-maximal response) even when the monkey fixated 2° to 5° beyond the edge of a face which subtended 8° to 17° at the retina. Moreover, the stimulus selectivity between faces was maintained this far eccentric within the receptive field.

Until recently, research on translation invariance considered the case in which there is only one object in the visual field. What happens in a cluttered, natural, environment? Do all objects that can activate an inferior temporal neuron do so whenever they are anywhere within the large receptive fields of inferior temporal cortex neurons (Sato 1989)? If so, the output of the visual system might be confusing for structures which receive inputs from the temporal cortical visual areas. In an investigation of this, it was found that the mean firing rate across all cells to a fixated effective face with a noneffective face in the parafovea (centered 8.5° from the fovea) was 34 spikes/s. On the other hand, the average response to a fixated non-effective face with an effective face in the periphery was 22 spikes/s (Rolls and Tovee 1995b). Thus these cells gave a reliable output about which stimulus is actually present at the fovea, in that their response was larger to a fixated effective face than to a fixated noneffective face, even when there are other parafoveal stimuli effective for the neuron.

It has now been shown that the receptive fields of inferior temporal cortex neurons, while large (typically 70° in diameter) when a test stimulus is presented

against a blank background, become much smaller, as little as several degrees in diameter, when objects are seen against a complex natural background (Rolls et al. 2003a). Object representation and selection in complex natural scenes is considered in Section 9.

8. A View-Independent Representation of Faces and Objects

It has also been shown that some temporal cortical neurons reliably responded differently to the faces of two different individuals independently of viewing angle (Hasselmo et al. 1989b), although in most cases (16/18 neurons) the response was not perfectly view independent. Mixed together in the same cortical regions are neurons with view-dependent responses (Hasselmo et al. 1989b). Such neurons might respond for example to a view of a profile of a monkey but not to a full-face view of the same monkey (Perrett et al. 1985a). These findings, of view-dependent, partially view-independent, and view-independent representations in the same cortical regions are consistent with the hypothesis discussed below that view-independent representations are being built in these regions by associating together neurons that respond to different views of the same individual.

Further evidence that some neurons in the temporal cortical visual areas have object-based rather than view-based responses comes from a study of a population of neurons that responds to moving faces (Hasselmo et al. 1989b). For example, four neurons responded vigorously to a head undergoing ventral flexion, irrespective of whether the view of the head was full face, of either profile, or even of the back of the head. These different views could only be specified as equivalent in object-based coordinates. Further, for all of the ten neurons that were tested in this way, the movement specificity was maintained across inversion, responding, for example, to ventral flexion of the head irrespective of whether the head was upright or inverted. In this procedure, retinally encoded or viewer-centered movement vectors are reversed, but the object-based description remains the same. It is an important property of these neurons that they can encode a description of an object that is based on relative motions of different parts of the object and which is not based on flow relative to the observer. The implication of this type of encoding is that the upper eyelids closing could be encoded as the same social signal that eye contact is being broken independently of the particular in-plane rotation (tilt, as far as being fully inverted) of the face being observed (or of the observer's head).

Also consistent with object-based encoding is the finding of a small number of neurons that respond to images of faces of a given absolute size, irrespective of the retinal image size or distance (Rolls and Baylis 1986).

Neurons with view-invariant responses of objects seen naturally by macaques have also been found (Booth and Rolls 1998). The stimuli were presented for 0.5 s on a color video monitor while the monkey performed a visual fixation task.

The stimuli were images of ten real plastic objects that had been in the monkey's cage for several weeks to enable him to build view-invariant representations of the objects. Control stimuli were views of objects that had never been seen as real objects. The neurons analyzed were in the TE cortex in and close to the ventral lip of the anterior part of the superior temporal sulcus. Many neurons were found that responded to some views of some objects. However, for a smaller number of neurons, the responses occurred only to a subset of the objects (using ensemble encoding), irrespective of the viewing angle. Further evidence consistent with these findings is that some studies have shown that the responses of some visual neurons in the inferior temporal cortex do not depend on the presence or absence of critical features for maximal activation (Perrett et al. 1982; Tanaka 1993, 1996). For example, Mikami et al (1994) have shown that some TE cells respond to partial views of the same laboratory instrument(s), even when these partial views contain different features. In a different approach, Logothetis et al. (1994) have reported that in monkeys extensively trained (over thousands of trials) to treat different views of computer-generated wire-frame "objects" as the same, a small population of neurons in the inferior temporal cortex did respond to different views of the same wire-frame object (Logothetis and Sheinberg 1996). The difference in the approach taken by Booth and Rolls (1998) was that no explicit training was given in invariant object recognition, as Rolls' hypothesis (1992a) is that view-invariant representations can be learned by associating together the different views of objects as they are moved and inspected naturally in a period that may be in the order of a few seconds.

9. The Representation of Objects in Complex Natural Scenes

9.1. Object-Based Attention and Object Selection in Complex Natural Scenes

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 modeling of attention has been with a small number, typically two, of objects in an otherwise blank scene. In this section, I 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.

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 et al. (2003a) analyzed the responses of inferior temporal cortex neurons to stimuli presented in complex natural backgrounds. The monkey had to search for two

Y2

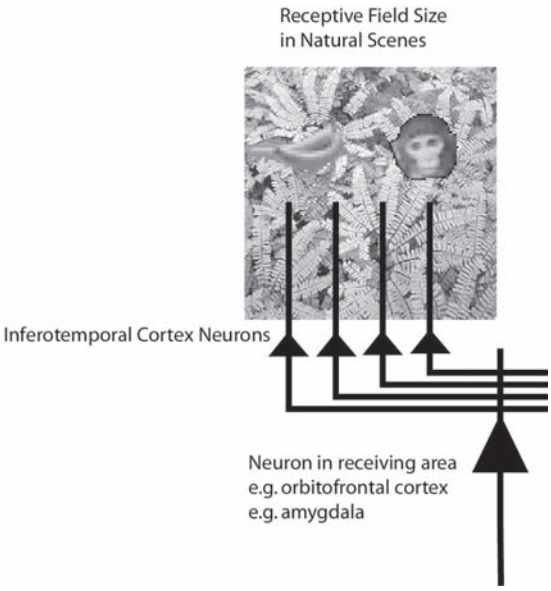


FIG. 6. Objects shown in a natural scene, in which the task was to search for and touch one of the stimuli. The objects in the task as run were smaller. The diagram shows that if the receptive fields of inferior temporal cortex neurons are large in natural scenes with multiple objects, then any receiving neuron in structures such as the orbitofrontal 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. (background.eps)

objects on a screen, and a touch of one object was rewarded with juice, and that of another object was punished with saline (Fig. 6). 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 (Fig. 7). It is proposed that this reduced translation invariance in natural scenes helps an unambiguous representation of an object that 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. 8 (Rolls et al. 2003a). The results summarized in Fig. 8 for 5° stimuli show that the receptive fields were large (77.6°) with a single stimulus Y2

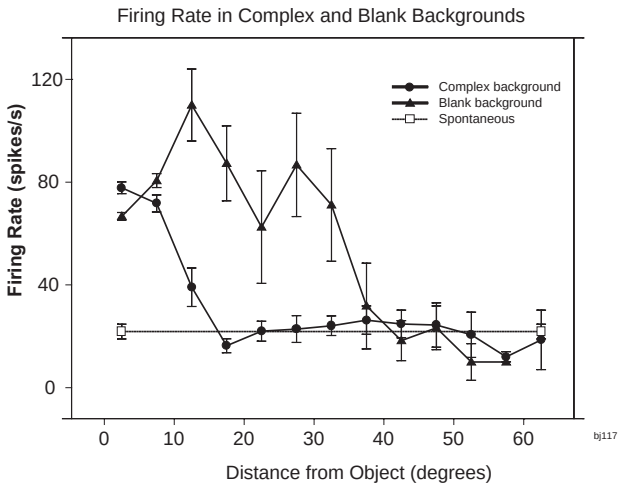


FIG. 7. Firing of a temporal cortex cell to an effective stimulus presented either in a blank background or in a natural scene, as a function of the angle in degrees at which the monkey was fixating away from the effective stimulus. The task was to search for and touch the stimulus. (After Rolls et al. 2003) (rateinbackground.eps)

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, 19.2°), and when it was not (middle right, 15.6°) (Rolls et al. 2003a). (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. 8.) The computational basis for these relatively minor effects of object-based attention when objects are viewed in natural scenes is considered in Section 15.

These findings on how objects are represented in natural scenes make the interface to memory and to action systems simpler, in that what is at the fovea can be interpreted (e.g., by an associative memory in the orbitofrontal cortex or amygdala) partly independently of the surroundings, and choices and actions can be directed if appropriate to what is at the fovea (Ballard 1993; Rolls and Deco 2002).

9.2. The Representation of Information About the Relative Positions of Multiple Objects in a Scene

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

Y2

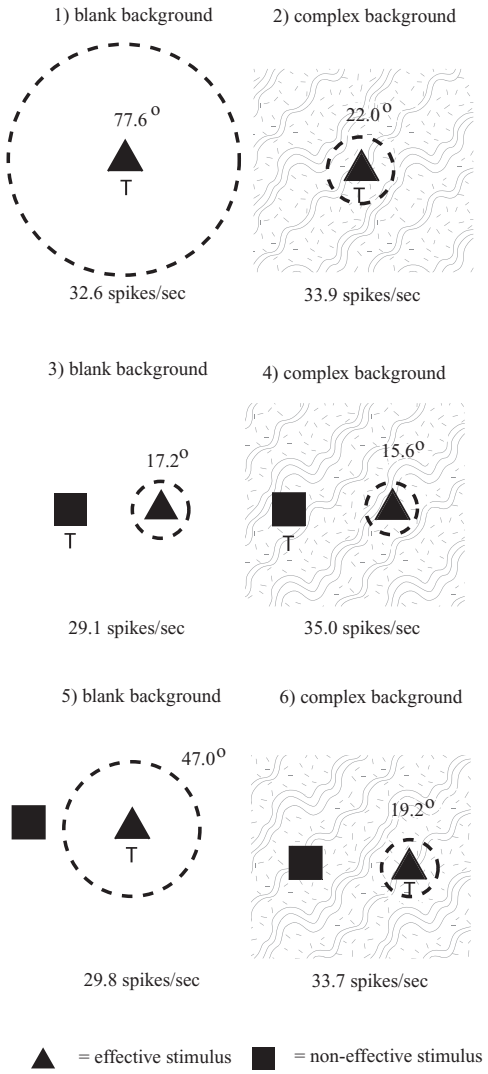


FIG. 8. Summary of the receptive field sizes of inferior temporal cortex neurons to a 5° effective stimulus presented in either a blank background (blank screen) or in a natural scene (complex background). The stimulus that was a target for action in the different experimental conditions is marked by *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° × 3.5° at the retina, and occurred on each trial in a random position in the 70° × 55° 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 et al. 2003) (rec_field7.eps)

Y2

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. 9. The asymmetry is much more evident in a scene with five images present (Fig. 9A) than when only one image is shown on an otherwise blank screen (Fig. 9B). 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 asymmetrical 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.

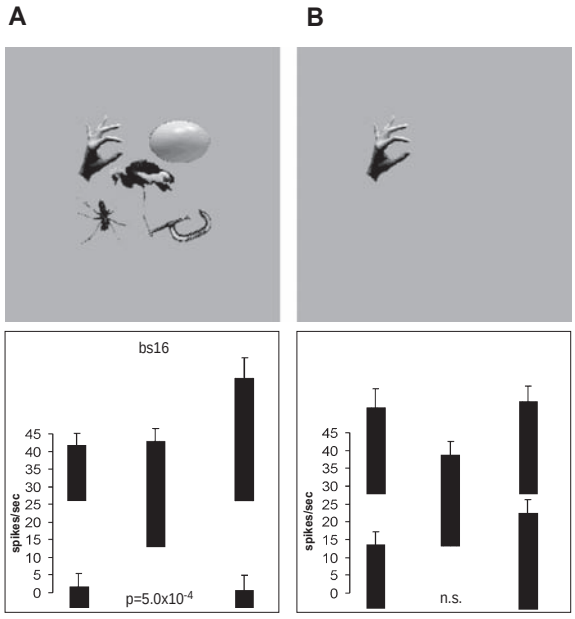


FIG. 9. **A** The responses (firing rate with the spontaneous rate subtracted, means ± SEM) of one neuron when tested with five stimuli simultaneously present in the close (10°) configuration with the parafoveal stimuli located 10° from the fovea. **B** The responses of the same neuron when only the 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) (5stim_fig1a.eps)

Y2

This information is provided by neurons that have firing rates, which 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 relationships 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 asymmetrical neurons firing.

This is, thus, the solution that these experiments indicate is used to the representation of multiple objects in a scene (Aggelopoulos and Rolls 2005), an issue which has previously been difficult to account for in neural systems with distributed representations (Mozer 1991) and for which "attention" has been a proposed solution.

10. Learning of New Representations in the Temporal Cortical Visual Areas

To investigate the hypothesis that visual experience might guide the formation of the responsiveness of neurons so that they provide an economical and ensemble-encoded representation of items actually present in the environment, the responses of inferior temporal cortex face-selective neurons have been analysed while a set of new faces were shown. It was found that some of the neurons studied in this way altered the relative degree to which they responded to the different members of the set of novel faces over the first few (1–2) presentations of the set (Rolls et al. 1989b). If in a different experiment a single novel face was introduced when the responses of a neuron to a set of familiar faces was being recorded, it was found that the responses to the set of familiar faces were not disrupted, while the responses to the novel face became stable within a few presentations. It is suggested that alteration of the tuning of individual neurons in this way results in a good discrimination over the population as a whole of the faces known to the monkey. This evidence is consistent with the categorization being performed by self-organizing competitive neuronal networks, as described below and elsewhere (Rolls 1989a; Rolls and Deco 2002; Rolls and Treves 1998; Rolls et al. 1989a).

Further evidence that these neurons can learn new representations very rapidly comes from an experiment in which binarized black-and-white images of faces that blended with the background were used. These images did not activate face-selective neurons. Full gray-scale images of the same photographs were then shown for ten 0.5-s presentations. It was found that in a number of cases, if the neuron happened to be responsive to that face, when the binarized version of the same face was shown next, the neurons responded to it (Tovee et al. 1996). This is a direct parallel to the same phenomenon that is observed psychophysically and provides dramatic evidence that these neurons are influenced by only a very few seconds (in this case 5 s) of experience with a visual stimulus. We have

shown a neural correlate of this effect using similar stimuli and a similar paradigm in a positron emission tomography (PET) neuroimaging study in humans, with a region showing an effect of the learning found for faces in the right temporal lobe and for objects in the left temporal lobe (Dolan et al. 1997).

Such rapid learning of representations of new objects appears to be a major type of learning in which the temporal cortical areas are involved. Ways in which this learning could occur are considered below. It is also the case that there is a much shorter term form of memory in which some of these neurons are involved, for whether a particular familiar visual stimulus (such as a face) has been seen recently, for some of these neurons respond differently to recently seen stimuli in short-term visual memory tasks (Baylis and Rolls 1987; Miller and Desimone 1994; Xiang and Brown 1998), and neurons in a more ventral cortical area respond during the delay in a short-term memory task (Miyashita 1993; Renart et al. 2000).

11. The Speed of Processing in the Temporal Cortical Visual Areas

Given that there is a whole sequence of visual cortical processing stages including V1, V2, V4, and the posterior inferior temporal cortex to reach the anterior temporal cortical areas, and that the response latencies of neurons in V1 are about 40 to 50 ms, and in the anterior inferior temporal cortical areas approximately 80 to 100 ms, each stage may need to perform processing for only 15 to 30 ms before it has performed sufficient processing to start influencing the next stage. Consistent with this, response latencies between V1 and the inferior temporal cortex increase from stage to stage (Thorpe and Imbert 1989). In a first approach to this issue, we measured the information available in short temporal epochs of the responses of temporal cortical face-selective neurons about which face had been seen. We found that if a period of the firing rate of 50 ms was taken, then this contained 84.4% of the information available in a much longer period of 400 ms about which of 4 faces had been seen. If the epoch was as little as 20 ms, the information was 65% of that available from the firing rate in the 400-ms period (Tovee et al. 1993). These high information yields were obtained with the short epochs taken near the start of the neuronal response, for example, in the poststimulus period of 100 to 120 ms. Moreover, we were able to show that the firing rate in short periods taken near the start of the neuronal response was highly correlated with the firing rate taken over the whole response period, so that the information available was stable over the whole response period of the neurons (Tovee et al. 1993). We were able to extend this finding to the case when a much larger stimulus set, of 20 faces, was used. Again, we found that the information available in short (e.g., 50-ms) epochs was a considerable proportion (e.g., 65%) of that available in a 400-ms-long firing rate analysis period (Tovee and Rolls 1995). These investigations thus showed that there was considerable information about which stimulus had been seen in short time epochs near the start

of the response of temporal cortex neurons. Moreover, we have shown that the information is available in the number of action potentials from each neuron (which might be 1, 2, or 3) in these short time periods (a rate code), and not in the order in which the spikes arrive from different neurons (Rolls et al. 2006b).

The next approach has been to use a visual backward masking paradigm. In this paradigm, there is a brief presentation of a test stimulus that is rapidly followed (within 1–100 ms) by the presentation of a second stimulus (the mask), which impairs or masks the perception of the test stimulus. It has been shown (Rolls and Tovee 1994) that when there is no mask inferior temporal cortex neurons respond to a 16-ms presentation of the test stimulus for 200 to 300 ms, far longer than the presentation time. It is suggested that this reflects the operation of a short-term memory system implemented in cortical circuitry, the importance of which in learning invariant representations is considered below in Section 13. If the pattern mask followed the onset of the test face stimulus by 20 ms (a stimulus onset asynchrony of 20 ms), face-selective neurons in the inferior temporal cortex of macaques responded for a period of 20 to 30 ms before their firing was interrupted by the mask (Rolls and Tovee 1994; Rolls et al. 1999). We went on to show that under these conditions (a test-mask stimulus onset asynchrony of 20 ms), human observers looking at the same displays could just identify which of six faces was shown (Rolls et al. 1994).

These results provide evidence that a cortical area can perform the computation necessary for the recognition of a visual stimulus in 20 to 30 ms (although it is true that for conscious perception, the firing needs to occur for 40–50 ms; see Rolls 2003). This condition provides a fundamental constraint that must be accounted for in any theory of cortical computation. The results emphasize just how rapidly cortical circuitry can operate. Although this speed of operation does seem fast for a network with recurrent connections (mediated by, e.g., recurrent collateral or inhibitory interneurons), analyses of networks with analogue membranes that integrate inputs, and with spontaneously active neurons, do show that such networks can settle very rapidly (Rolls and Treves 1998; Treves 1993; Treves et al. 1996). This approach has been extended to multilayer networks such as those found in the visual system, and again very rapid propagation (in 40–50 ms) of information through such a four-layer network with recurrent collaterals operating at each stage has been found (Panzeri et al. 2001). The computational approaches thus show that there is sufficient time for feedback processing using recurrent collaterals within each cortical stage during the fast cortical processing of visual inputs.

12. Different Neural Systems Are Specialized for Face Expression Decoding and for Face Recognition

It has been shown that some neurons respond to face identity and others to face expression (Hasselmo et al. 1989a). The neurons responsive to expression were found primarily in the cortex in the superior temporal sulcus, whereas the neurons

responsive to identity (described in the preceding sections) were found in the inferior temporal gyrus including areas TEa and TEm. Information about facial expression is of potential use in social interactions (Rolls 1984, 1986a,b, 1990, 1999b, 2005). Damage to this population may contribute to the deficits in social and emotional behavior that are part of the Kluver–Bucy syndrome produced by temporal lobe damage in monkeys (Leonard et al. 1985; Rolls 1981, 1984, 1986a,b, 1990, 1999b, 2005).

A further way in which some of these neurons in the cortex in the superior temporal sulcus may be involved in social interactions is that some of them respond to gestures, for example, to a face undergoing ventral flexion, as described above and by Perrett et al. (1985b). The interpretation of these neurons as being useful for social interactions is that in some cases these neurons respond not only to ventral head flexion, but also to the eyes lowering and the eyelids closing (Hasselmo et al. 1989a). These two movements (head lowering and eyelid lowering) often occur together when a monkey is breaking social contact with another. It is also important when decoding facial expression to retain some information about the head direction of the face stimulus being seen relative to the observer, for this is very important in determining whether a threat is being made in your direction. The presence of view-dependent, head and body gesture (Hasselmo et al. 1989b), and eye gaze (Perrett et al. 1985b), representations in some of these cortical regions where face expression is represented is consistent with this requirement. In contrast, the TE areas (more ventral, mainly in the macaque inferior temporal gyrus), in which neurons tuned to face identity (Hasselmo et al. 1989a) and with view-independent responses (Hasselmo et al. 1989b) are more likely to be found, may be more related to a view-invariant representation of identity. Of course, for appropriate social and emotional responses, both types of subsystem would be important, for it is necessary to know both the direction of a social gesture, and the identity of the individual, to make the correct social or emotional response.

13. Possible Computational Mechanisms in the Visual Cortex for Face and Object Recognition

The neurophysiological findings described above, and wider considerations on the possible computational properties of the cerebral cortex (Rolls 1989a,b, 1992a; Rolls and Treves 1998), lead to the following outline working hypotheses on object (including face) recognition by visual cortical mechanisms (Rolls and Deco 2002).

Cortical visual processing for object recognition is considered to be organized as a set of hierarchically connected cortical regions consisting at least of V1, V2, V4, posterior inferior temporal cortex (TEO), inferior temporal cortex (e.g., TE3, TEa and TEm), and anterior temporal cortical areas (e.g., TE2 and TE1), as shown in Fig. 10. There is convergence from each small part of a region to the succeeding region (or layer in the hierarchy) in such a way that the receptive

Y2

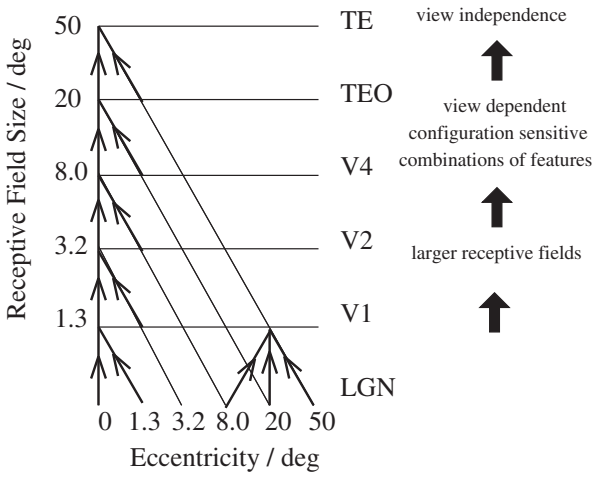


FIG. 10. Schematic diagram showing convergence achieved by the forward projections in the visual system and the types of representation that may be built by competitive networks operating at each stage of the system from the primary visual cortex (V1) to the inferior temporal visual cortex (area *TE*) (see text). *LGN*, lateral geniculate nucleus. Area *TEO* forms the posterior inferior temporal cortex. The receptive fields in the inferior temporal visual cortex (e.g., in the TE areas) cross the vertical midline (not shown) (4_7.eps)

field sizes of neurons (e.g., 1° near the fovea in V1) become larger by a factor of approximately 2.5 with each succeeding stage (and the typical parafoveal receptive field sizes found would not be inconsistent with the calculated approximations of, e.g., 8° in V4, 20° in TEO, and 50° in inferior temporal cortex) (Boussaoud et al. 1991) (see Fig. 10). Such zones of convergence would overlap continuously with each other. This connectivity would be part of the architecture by which translation invariant representations are computed. Each layer is considered to act partly as a set of local self-organizing competitive neuronal networks with overlapping inputs. The region within which competition would be implemented would depend on the spatial properties of inhibitory interneurons and might operate over distances of 1 to 2 mm in the cortex. These competitive nets operate by a single set of forward inputs leading to (typically nonlinear, e.g., sigmoid) activation of output neurons; of competition between the output neurons mediated by a set of feedback inhibitory interneurons that receive from many of the principal (in the cortex, pyramidal) cells in the net and project back (via inhibitory interneurons) to many of the principal cells, which serves to decrease the firing rates of the less active neurons relative to the rates of the more active neurons; and then of synaptic modification by a modified Hebb rule, such that synapses to strongly activated output neurons from active input axons strengthen and those from inactive input axons weaken (Rolls and Deco 2002; Rolls and Treves 1998).

Y2

Translation, size, and view invariance could be computed in such a system by utilizing competitive learning operating across short time scales to detect regularities in inputs when real objects are transforming in the physical world (Rolls 1992a, 2000a; Rolls and Deco 2002; Wallis and Rolls 1997). The hypothesis is that because objects have continuous properties in space and time in the world, an object at one place on the retina might activate feature analyzers at the next stage of cortical processing, and when the object was translated to a nearby position, because this would occur in a short period (e.g., 0.5 s), the membrane of the postsynaptic neuron would still be in its "Hebb-modifiable" state (caused, for example, by calcium entry as a result of the voltage-dependent activation of NMDA receptors, or by continuing firing of the neuron implemented by recurrent collateral connections forming a short term memory), and the presynaptic afferents activated with the object in its new position would thus become strengthened on the still-activated postsynaptic neuron. It is suggested that the short temporal window (e.g., 0.5 s) of Hebb modifiability helps neurons to learn the statistics of objects moving in the physical world, and at the same time to form different representations of different feature combinations or objects, as these are physically discontinuous and present less regular correlations to the visual system. Földiák (1991) has proposed computing an average activation of the postsynaptic neuron to assist with translation invariance. I also suggest that other invariances, for example, size, spatial frequency, rotation, and view invariance, could be learned by similar mechanisms to those just described (Rolls 1992a). It is suggested that the process takes place at each stage of the multiple layer cortical processing hierarchy, so that invariances are learned first over small regions of space, and then over successively larger regions; this limits the size of the connection space within which correlations must be sought.

Increasing complexity of representations could also be built in such a multiple layer hierarchy by similar competitive learning mechanisms. To avoid the combinatorial explosion, it is proposed that low-order combinations of inputs would be what is learned by each neuron. Evidence consistent with this suggestion that neurons are responding to combinations of a few variables represented at the preceding stage of cortical processing is that some neurons in V2 and V4 respond to end-stopped lines, to tongues flanked by inhibitory subregions, or to combinations of colors (see references cited by Rolls 1991); in posterior inferior temporal cortex to stimuli that may require two or more simple features to be present (Tanaka et al. 1990); and in the temporal cortical face processing areas to images which require the presence of several features in a face (such as eyes, hair, and mouth) to respond (Perrett et al. 1982; Yamane et al. 1988). It is an important part of this suggestion that some local spatial information would be inherent in the features which were being combined (Elliffe et al. 2002). For example, cells might not respond to the combination of an edge and a small circle unless they were in the correct spatial relationship to each other. [This is in fact consistent with the data of Tanaka et al. (1990) and with our data on face neurons (Rolls et al. 1994), in that some faces neurons require the face features to be in the correct spatial configuration, and not jumbled.] The local spatial information in

Y2

the features being combined would ensure that the representation at the next level would contain some information about the (local) arrangement of features. Further low-order combinations of such neurons at the next stage would include sufficient local spatial information so that an arbitrary spatial arrangement of the same features would not activate the same neuron, and this is the proposed, and limited, solution that this mechanism would provide for the feature-binding problem (Elliffe et al. 2002).

It is suggested that view-independent representations could be formed by the same type of computation, operating to combine a limited set of views of objects. The plausibility of providing view-independent recognition of objects by combining a set of different views of objects has been proposed by a number of investigators (Koenderink and Van Doorn 1979; Logothetis et al. 1994; Poggio and Edelman 1990; Ullman 1996). Consistent with the suggestion that the view-independent representations are formed by combining view-dependent representations in the primate visual system is the fact that in the temporal cortical areas, neurons with view-independent representations of faces are present in the same cortical areas as neurons with view-dependent representations (from which the view-independent neurons could receive inputs) (Booth and Rolls 1998; Hasselmo et al. 1989b; Perrett et al. 1987). This solution to "object-based" representations is very different from that traditionally proposed for artificial vision systems, in which the coordinates in 3-D space of objects are stored in a database, and general-purpose algorithms operate on these to perform transforms such as translation, rotation, and scale change in 3-D space (Ullman 1996), or a linked list of feature parts is used (Marr 1982). In the present, much more limited but more biologically plausible scheme, the representation would be suitable for recognition of an object, and for linking associative memories to objects, but would be less good for making actions in 3-D space to particular parts of, or inside, objects, as the 3-D coordinates of each part of the object would not be explicitly available. It is therefore proposed that visual fixation is used to locate in foveal vision part of an object to which movements must be made, and that local disparity and other measurements of depth then provide sufficient information for the motor system to make actions relative to the small part of space in which a local, view-dependent, representation of depth would be provided (Ballard 1990; Rolls and Deco 2002).

14. A Computational Model of Invariant Visual Object and Face Recognition

To test and clarify the hypotheses just described about how the visual system may operate to learn invariant object recognition, we have performed simulations that implement many of the ideas just described and which are consistent with and based on much of the neurophysiology summarized here. The network simulated (VisNet) can perform object, including face, recognition in a biologically plausible way, and after training shows, for example, translation and view

invariance (Rolls and Deco 2002; Rolls and Milward 2000; Wallis and Rolls 1997; Wallis et al. 1993).

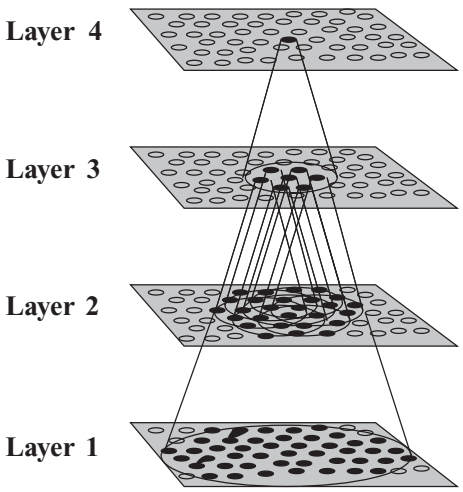
In the four-layer network, the successive layers correspond approximately to V2, V4, the posterior temporal cortex, and the anterior temporal cortex. The forward connections to a cell in one layer are derived from a topologically corresponding region of the preceding layer, using a Gaussian distribution of connection probabilities to determine the exact neurons in the preceding layer to which connections are made. This schema is constrained to preclude the repeated connection of any cells. Each cell receives 100 connections from the 32×32 cells of the preceding layer, with a 67% probability that a connection comes from within 4 cells of the distribution center. Figure 11 shows the general convergent network architecture used, and may be compared with Fig. 10. Within each layer, lateral inhibition between neurons has a radius of effect just greater than the radius of feed-forward convergence just defined. The lateral inhibition is simulated via a linear local contrast-enhancing filter active on each neuron. (Note that this differs from the global "winner-take-all" paradigm implemented by Földiák 1991.) The cell activation is then passed through a nonlinear cell activation function, which also produces contrast enhancement of the firing rates.

So that the results of the simulation might be made particularly relevant to understanding processing in higher cortical visual areas, the inputs to layer 1 come from a separate input layer that provides an approximation to the encoding found in visual area 1 (V1) of the primate visual system.

The synaptic learning rule used can be summarized as follows:

$$\delta w_{ij} = k\, m_i\, r'_j \text{ and}$$

$$m_i^t = (1 - \eta) r_i^{(t)} + \eta m_i^{(t-1)}$$



Y2 FIG. 11. Hierarchical network structure of VisNet (visnetarchi.eps)

where r'_j is the j^{th} input to the neuron, r_i is the output of the i^{th} neuron, w_{ij} is the j^{th} weight on the i^{th} neuron, η governs the relative influence of the trace and the new input (typically 0.4–0.6), and $m_i^{(t)}$ represents the value of the i^{th} cell's memory trace at time t. In the simulation, the neuronal learning was bounded by normalization of each cell's dendritic weight vector, as in standard competitive learning (Rolls and Treves 1998; Rolls and Deco 2002).

To train the network to produce a translation-invariant representation, one stimulus was placed successively in a sequence of nine positions across the input, then the next stimulus was placed successively in the same sequence of nine positions across the input, and so on through the set of stimuli. The idea was to enable the network to learn whatever was common at each stage of the network about a stimulus shown in different positions. To train on view invariance, different views of the same object were shown in succession, then different views of the next object were shown in succession, and so on. It has been shown that the network can learn to form neurons in the last layer of the network that respond to one of a set of simple shapes (such as "T, L, and +") with translation invariance, or to a set of five to eight faces with translation, view, or size invariance, provided that the trace learning rule (and not a simple Hebb rule) is used (Figs. 12, 13) (Rolls and Deco 2002; Wallis and Rolls 1997).

There have been a number of investigations to explore this type of learning further. Rolls and Milward (2000) explored the operation of the trace learning rule used in the VisNet architecture, and showed that the rule operated especially well if the trace incorporated activity from previous presentations of the same object, but no contribution from the current neuronal activity being produced by the current exemplar of the object. The explanation for this is that this temporally asymmetrical rule (the presynaptic term from the current exemplar, and the trace

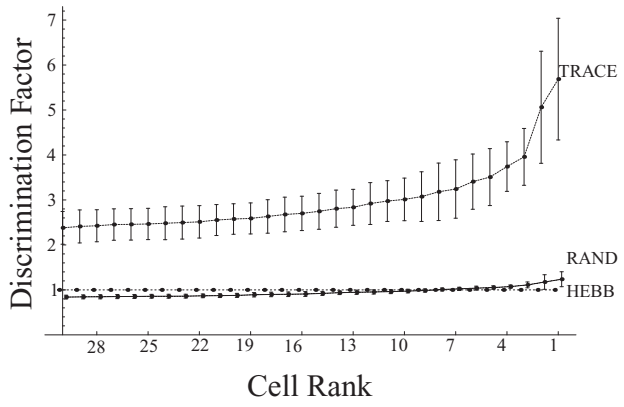


FIG. 12. Comparison of VisNet network discrimination when trained with the trace learning rule, with a *HEBB* rule (no trace), and when not trained (random, *RAND*) on three stimuli, +, T, and L, at nine different locations. (After Wallis and Rolls 1997) (tlcrank2lrn.eps)

Y2

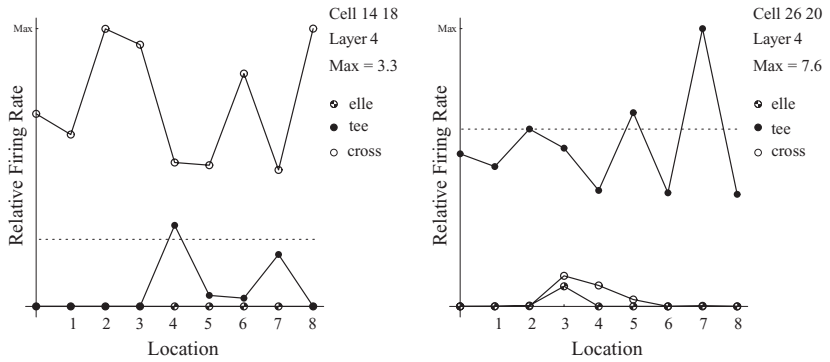


FIG. 13. Response profiles for two fourth- layer neurons in VisNet, discrimination factors 4.07 and 3.62, in the L, T, and + invariance learning experiment. (After Wallis and Rolls 1997) (l4tlc2a.eps)

from the preceding exemplars) encourages neurons to respond to the current exemplar in the same way as they did to previous exemplars. It is of interest to consider whether intracellular processes related to LTP might implement an approximation of this rule, given that it is somewhat more powerful than the standard trace learning rule described above. Rolls and Stringer (2001) went on to show that part of the power of this type of trace rule can be related to gradient descent and temporal difference learning (Sutton and Barto 1998). Elliffe et al. (2002) examined the issue of spatial binding in this general class of hierarchical architecture studied originally by Fukushima (1980, 1989, 1991), and showed how by forming high spatial precision feature combination neurons early in processing, it is possible for later layers to maintain high precision for the relative spatial position of features within an object, yet achieve invariance for the spatial position of the whole object.

These results show that the proposed learning mechanism and neural architecture can produce cells with responses selective for stimulus identity with considerable position or view invariance (Rolls and Deco 2002). This ability to form invariant representations is an important property of the temporal cortical visual areas, for if a reinforcement association leading to an emotional or social response is learned to one view of a face, that learning will automatically generalize to other views of the face. This is a fundamental aspect of the way in which the brain is organized to allow this type of capability for emotional and social behavior (Rolls 1999b, 2005). Further developments include operation of the system in a cluttered environment (Stringer and Rolls 2000), generalization from trained to untrained views of objects (Stringer and Rolls 2002), a new training algorithm named continuous transformation learning (Stringer et al. 2006), and a unifying theory of how invariant representations of optic flow produced by rotating or looming objects could be produced in the brain (Rolls and Stringer 2006).

15. Object Representation and Attention in Natural Scenes: A Computational Account

The results described in Section 9 and summarized in Fig. 8 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°) 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 et al. 2003a).

Trappenberg et al. (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 (Fig. 14). 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 because of the cortical higher magnification factor of the fovea incorporated into the model. [The cortical magnification factor can be expressed as the number of millimeters of cortex representing 1° of visual field. The cortical magnification factor decreases rapidly with increasing eccentricity from the fovea (Cowey and Rolls 1975; Rolls and Cowey 1970).] This difference results in smaller receptive fields of IT neurons in complex scenes because the object tends to need to be close to the fovea to trigger the attractor into the state representing that object. (In other words, if the object is far from the fovea in a cluttered scene, then the object will not trigger neurons in IT that 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 the competition for which attractor state is retrieved is weighted toward 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, because 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 that 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

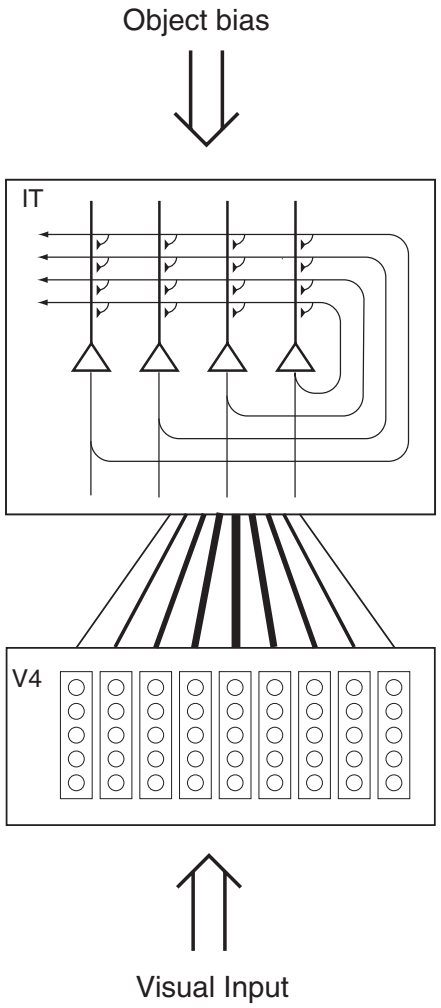


FIG. 14. The architecture of the inferior temporal cortex (IT) model of Trappenberg et al. (2002) operating as an attractor network with inputs from the fovea given preferential weighting by the greater magnification factor of the fovea. The model also has a top-down object-selective bias input. The model was used to analyze how object vision and recognition operate in complex natural scenes (itcattractor.eps)

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 (Rolls and Deco 2002; Rolls et al. 2003a).

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 2000a; Rolls et al. 2003a). At least part of the answer to this may be that inferior temporal cortex neurons must have the capability to have large receptive fields if they are to handle large objects (Rolls and 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 effects of bottom-up and top-down effects described elsewhere in this chapter.

The implementation of the simulations is described by Trappenberg et al. (2002), and some of the results obtained with the architecture (Fig. 14) 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. 15A by the line labeled "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 finding demonstrates that the attractor dynamics can support translation invariant object recognition even though the translation invariant weight vectors between V4 and IT are explicitly mapped by a modulation factor 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. 15A with the line labeled "natural background." Many objects in the visual scene are now

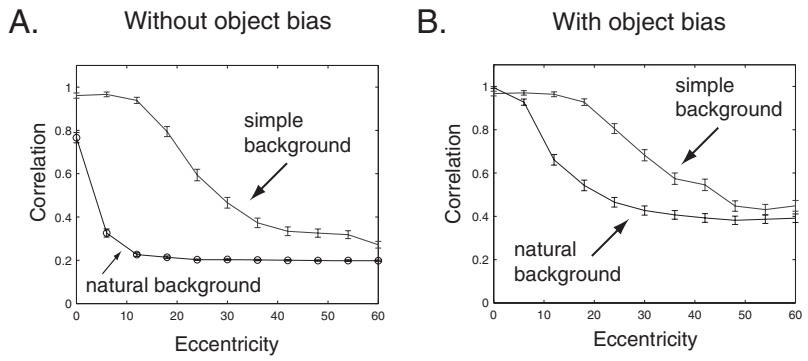


FIG. 15. Correlations as measured by the normalized dot product between the object vector used to train IT and the state of the IT network after settling into a stable state with a single object in the visual scene (blank background) or with other trained objects at all possible locations in the visual scene (natural background). There is no object bias included in the results shown in **A**, whereas an object bias is included in the results shown in **B**, with $k^{ITBIAS} = 0.7$ in the experiments with a natural background and $k^{ITBIAS} = 0.1$ in the experiments with a blank background (vis6_fig2r.eps)

competing for recognition by the attractor network, and 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. 15B where an object bias was used. The object bias biases the IT network toward 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 (compare Fig. 15B to Fig. 15A). This models the effect found neurophysiologically (Rolls et al. 2003a).

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.

Y2

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 fact that objects close to the fovea are weighted by the cortical magnification factor. 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 et al. 2003a). The model accounts for this by an effect of top-down bias, which simply biases the neurons toward 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 et al. 2000, 2001; Rolls and Deco 2002). 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 foregoing simulation by the competition inherent in an attractor network.

We have also been able to obtain similar results in a hierarchical feed-forward network where each layer operates as a competitive network (Deco and Rolls 2004). This network thus captures many of the properties of our hierarchical model of invariant object recognition (Elliffe et al. 2002; Rolls 1992a; Rolls and Deco 2002; Rolls and Milward 2000; Rolls and Stringer 2001, 2006; Stringer and Rolls 2000, 2002; Stringer et al. 2006; Wallis and 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 (Fig. 16).

Deco and Rolls (2004) trained the network described shown in Fig. 16 with two objects, and used the trace learning rule (Rolls and Milward 2000; Wallis and Rolls 1997) 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.

Figure 17 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 (solid line for blank background, dashed line for cluttered background) 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 (solid line in Fig. 17), larger receptive fields were observed. The decrease in neuronal response as a function of distance from the fovea is mainly the effect of the magnification factor implemented in V1. On the other hand, when the object was in a complex cluttered background, the 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

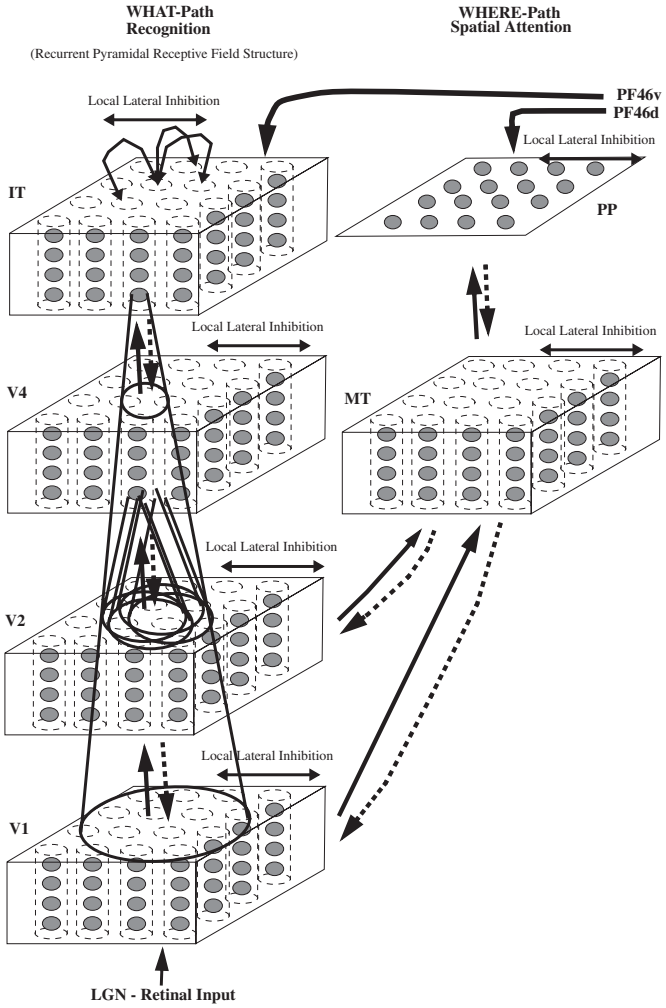


FIG. 16. Cortical architecture for hierarchical and attention-based visual perception. The system is essentially composed of five modules structured such that they resemble the two known main visual paths of the mammalian visual cortex. Information from the retino-geniculo-striate pathway enters the 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 occipitoparietal 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* show the top-down back-projections. Short-term memory systems in the prefrontal cortex (PF46) apply top-down attentional bias to the object or spatial processing streams. (After Deco and Rolls 2004) (visnet3archi.eps)

Y2

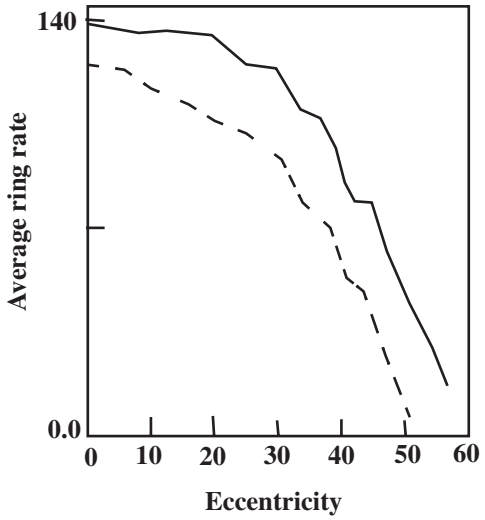


FIG. 17. Average firing activity of an inferior temporal cortex neuron as a function of eccentricity from the fovea, in the simulation of Deco and Rolls (2004). When the object was in a blank background (*solid line*), large receptive fields are observed because of the translation invariance of inferior temporal neurons. The decay is mainly the result of 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. (After Deco and Rolls 2004) (rf1_5a.eps)

character of the competition expressed in the inferior temporal cortex module (caused by 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 fields 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. 18). 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), shrinkage of the receptive field size is observed in the complex background (dashed line). When attentional object bias is introduced (b), the shrinkage of the receptive field because of the complex background is slightly reduced (dashed line). Rolls et al. (2003a) also found that in natural scenes that the effect of object-based attention on the response properties of inferior temporal cortex neurons was relatively Y2

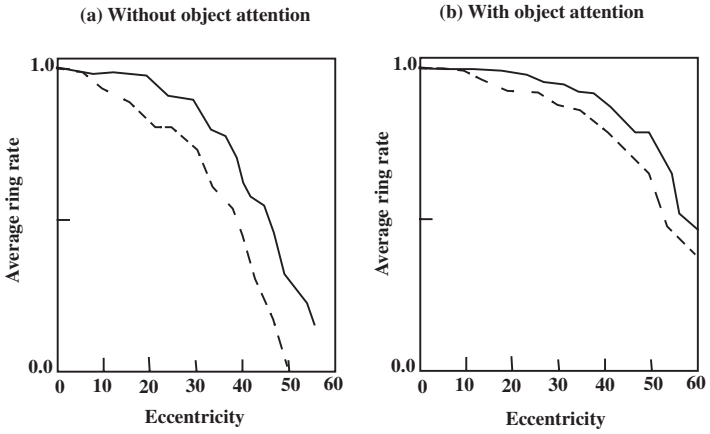


FIG. 18. Influence of object-based attentional top-down bias from prefrontal area 46v on the effective size of an inferior temporal cortex neuron for the case of an object in a blank (*solid line*) or a cluttered (*dashed line*) background. The average firing activity was normalized to the maximum value 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 because of the complex background was slightly reduced. (After Deco and Rolls 2004) (rf1_6a.eps)

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 versus when it was not. In the framework of the model (Deco and Rolls 2004), the reduction of the shrinkage of the receptive field is caused by the biasing of the competition in the inferior temporal cortex layer in favor 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 mechanism for facilitation of the search for specific objects in complex natural scenes.

16. A Biased Competition Model of Object and Spatial Attentional Effects on the Representations in the Visual System

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 work of Chelazzi et al. (1993) has led to a promising account of attention termed the biased competition hypothesis (Desimone and Duncan 1995; 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 favor 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 (Deco and Rolls 2005a), namely, single cells (Deco and Lee 2002; Rolls and Deco 2002), fMRI (Corchs and Deco 2002), psychophysics (Deco and Rolls 2005a; Rolls and Deco 2002), and neuropsychology (Deco and Rolls 2002). In the context of working memory, further developments (Deco and Rolls 2003) 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 (2005c) extended previous concepts of the role of biased competition in attention by providing the first analysis at the integrate-and-fire neuronal level, which allows the neuronal nonlinearities in the system to be explicitly modeled, 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; this 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 et al. (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 (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 toward 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 and Rolls 2003, 2005b, 2006).

In relation to representation in the brain, the impact of these findings is that they show details of the mechanisms by which representations can be modulated by attention, and moreover can account for many phenomena in attention using models in which the firing rate of neurons is represented and in which stimulus-dependent synchrony is not involved.

17. A Representation of Faces in the Amygdala

Outputs from the temporal cortical visual areas reach the amygdala and the orbitofrontal cortex, and evidence is accumulating that these brain areas are involved in social and emotional responses to faces (Rolls 1990, 1999b, 2000b, 2005; Rolls and Deco 2002). For example, lesions of the amygdala in monkeys disrupt social and emotional responses to faces, and we have identified a population of neurons with face-selective responses in the primate amygdala (Leonard et al. 1985), some of which may respond to facial and body gestures (Brothers et al. 1990). In humans, bilateral dysfunction of the amygdala can impair face expression identification, although primarily of fear (Adolphs et al. 1995; Adolphs et al. 2002), so that the impairment seems much less severe than that produced by orbitofrontal cortex damage.

18. A Representation of Faces in the Orbitofrontal Cortex

Rolls et al. (2006a) have found a number of face-responsive neurons in the orbitofrontal cortex, and they are also present in adjacent prefrontal cortical areas (Wilson et al. 1993). The orbitofrontal cortex face-responsive neurons, first observed by Thorpe et al. (1983), then by Rolls et al. (2006a), tend to respond with longer latencies than temporal lobe neurons (140–200 ms typically, compared with 80–100 ms); they also convey information about which face is being seen, by having different responses to different faces (Fig. 19), and are typically rather harder to activate strongly than temporal cortical face-selective neurons, in that many of them respond much better to real faces than to 2-D images of faces on a video monitor (Rolls and Baylis 1986). Some of the orbitofrontal cortex face-selective neurons are responsive to face gesture or movement. The findings are consistent with the likelihood that these neurons are activated via the inputs from the temporal cortical visual areas in which face-selective neurons are found. The significance of the neurons is likely to be related to the fact that faces convey information that is important in social reinforcement, both by conveying face expression (Hasselmo et al. 1989a), which can indicate reinforcement, and by encoding information about which individual is present, also important in evaluating and utilizing reinforcing inputs in social situations (Rolls et al. 2006a).

Y2

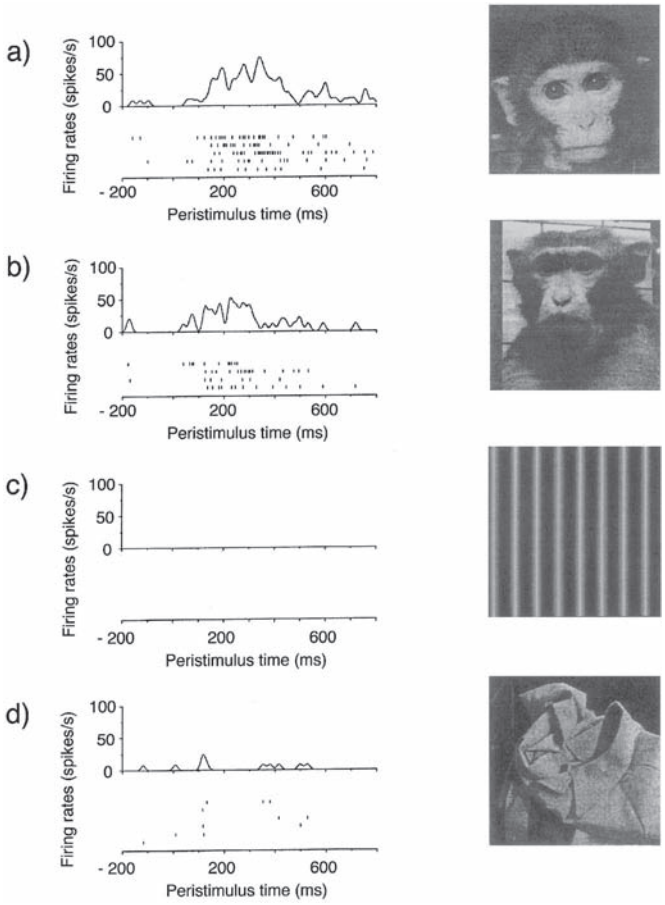


FIG. 19. Orbitofrontal cortex face-selective neuron as found in macaques. Peristimulus rastergrams and time histograms are shown. Each trial is a *row in the rastergram*. Several trials for each stimulus are shown. The ordinate is in spikes/s. The neuron responded best to face (**a**), also it responded, although less, to face (**b**), had different responses to other faces (not shown), and did not respond to non-face stimuli (e.g., **c** and **d**). The stimulus appeared at time 0 on a video monitor. (After Rolls 1999a; Rolls et al. 2005) (4.21a.eps)

We have also been able to obtain evidence that nonreward used as a signal to reverse behavioral choice is represented in the human orbitofrontal cortex (for background, see Rolls 2005). Kringelbach and Rolls (2003) used the faces of two different people, and if one face was selected then that face smiled, and if the other was selected, the face showed an angry expression. After good performance was acquired, there were repeated reversals of the visual discrimination task. Kringelbach and Rolls (2003) found that activation of a lateral part of the orbitofrontal cortex in the fMRI study was produced on the error trials, that is,

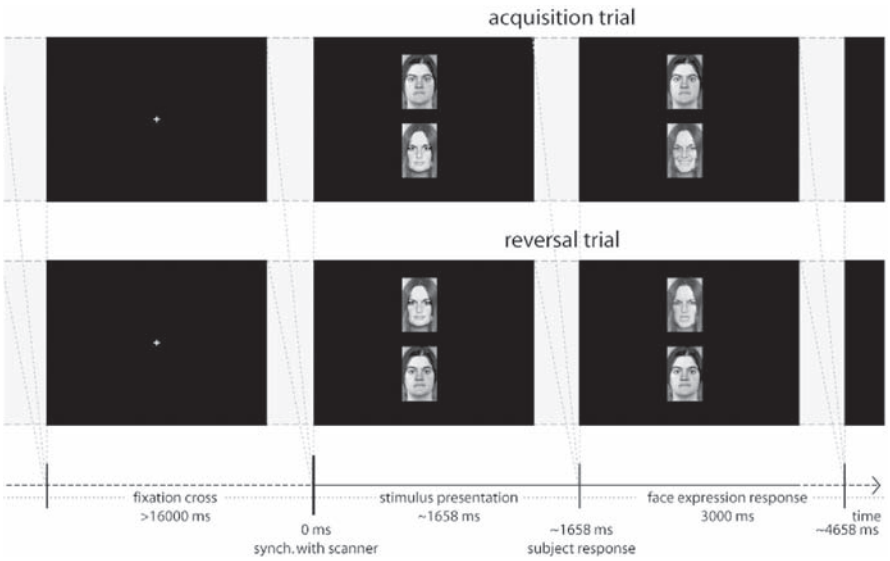


FIG. 20. Social reversal task: The trial starts synchronized with the scanner, and two people with neutral face expressions are presented to the subject. The subject has to select one of the people by pressing the corresponding button, and the person will then either smile or show an angry face expression for 3000 ms, depending on the current mood of the person. The task for the subject is to keep track of the mood of each person and choose the "happy" person as much as possible (*upper row*). Over time (after between 4 and 8 correct trials), this will change so that the "happy" person becomes "angry" and vice versa, and the subject has to learn to adapt her choices accordingly (*bottom row*). Randomly intermixed trials with either two men, or two women, were used to control for possible gender and identification effects, and a fixation cross was presented between trials for at least 16000 ms. (After Kringelbach and Rolls 2003) (OFCfacereversaltask.eps)

when the human chose a face and did not obtain the expected reward (Figs. 20, 21). Control tasks showed that the response was related to the error, and the mismatch between what was expected and what was obtained, in that just showing an angry face expression did not selectively activate this part of the lateral orbitofrontal cortex. An interesting aspect of this study that makes it relevant to human social behavior is that the conditioned stimuli were faces of particular individuals and the unconditioned stimuli were face expressions. Moreover, the study reveals that the human orbitofrontal cortex is very sensitive to social feedback when it must be used to change behavior (Kringelbach and Rolls 2003, 2004; Rolls 2005).

To investigate the possible significance of face-related inputs to the orbitofrontal cortex visual neurons described above, we also tested the responses to faces of patients with orbitofrontal cortex damage. We included tests of face (and also voice) expression decoding, because these are ways in which the reinforcing quality of individuals is often indicated. Impairments in the identification of facial

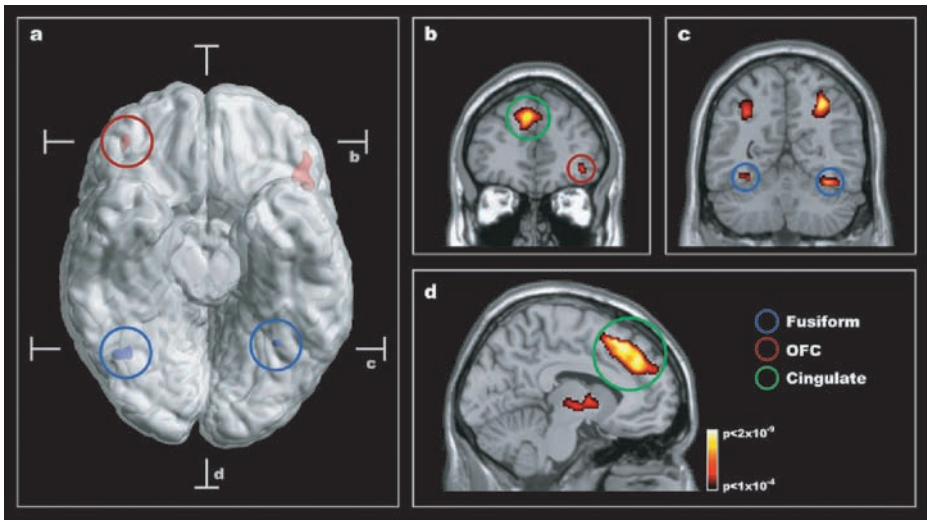


FIG. 21. Social reversal: composite figure showing that changing behavior based on face expression is correlated with increased brain activity in the human orbitofrontal cortex. **a** The figure is based on two different group statistical contrasts from the neuroimaging data, which are superimposed on a ventral view of the human brain with the cerebellum removed, and with indication of the location of the two coronal slices (**b**, **c**) and the transverse slice (**d**). The *red* activations in the orbitofrontal cortex (denoted *OFC*, maximal activation: $Z = 4.94$; 42, 42, –8; and $Z = 5.51$; x, y, $z = -46$, 30, –8) shown on the rendered brain arise from a comparison of reversal events with stable acquisition events, while the *blue* activations in the fusiform gyrus (denoted *Fusiform*, maximal activation: $Z > 8$; 36, –60, –20 and $Z = 7.80$; –30, –56, –16) arise from the main effects of face expression. **b** The coronal slice through the frontal part of the brain shows the cluster in the right orbitofrontal cortex across all nine subjects when comparing reversal events with stable acquisition events. Significant activity was also seen in an extended area of the anterior cingulate/paracingulate cortex (denoted *Cingulate*, maximal activation: $Z = 6.88$; –8, 22, 52; *green circle*). **c** The coronal slice through the posterior part of the brain shows the brain response to the main effects of face expression with significant activation in the fusiform gyrus and the cortex in the intraparietal sulcus (maximal activation: $Z > 8$; 32, –60, 46; and $Z > 8$; –32, –60, 44). **d** The transverse slice shows the extent of the activation in the anterior cingulate/paracingulate cortex when comparing reversal events with stable acquisition events. Group statistical results are superimposed on a ventral view of the human brain with the cerebellum removed, and on coronal and transverse slices of the same template brain (activations are thresholded at $P = 0.0001$ for purposes of illustration to show their extent). (After Kringelbach and Rolls 2003) (OFCfacereversal.eps)

and vocal emotional expression were demonstrated in a group of patients with ventral frontal lobe damage who had socially inappropriate behavior (Hornak et al. 1996; Rolls 1999a). The expression identification impairments could occur independently of perceptual impairments in facial recognition, voice discrimination, or environmental sound recognition. The face and voice expression problems did not necessarily occur together in the same patients, providing an

Y2

indication of separate processing. Poor performance on both expression tests was correlated with the degree of alteration of emotional experience reported by the patients. There was also a strong positive correlation between the degree of altered emotional experience and the severity of the behavioural problems (e.g., disinhibition) found in these patients. A comparison group of patients with brain damage outside the ventral frontal lobe region, without these behavioral problems, was unimpaired on the face expression identification test, was significantly less impaired at vocal expression identification, and reported little subjective emotional change (Hornak et al. 1996; Rolls 1999a).

To obtain clear evidence that the changes in face and voice expression identification, emotional behavior, and subjective emotional state were related to orbitofrontal cortex damage itself, and not to damage to surrounding areas, which is present in many closed head injury patients, we performed further assessments in patients with circumscribed lesions made surgically in the course of treatment (Hornak et al. 2003). This study also enabled us to determine whether there was functional specialization within the orbitofrontal cortex, and whether damage to nearby and connected areas (such as the anterior cingulate cortex) in which some of the patients had lesions could produce similar effects. We found that some patients with bilateral lesions of the orbitofrontal cortex had deficits in voice and face expression identification, and the group had impairments in social behavior and significant changes in their subjective emotional state (Hornak et al. 2003). The same group of patients had deficits on a probabilistic monetary reward reversal task, indicating that they have difficulty not only in representing reinforcers such as face expression, but also in using reinforcers (such as monetary reward) to influence behavior (Hornak et al. 2004). Some patients with unilateral damage restricted to the orbitofrontal cortex also had deficits in voice expression identification, and the group did not have significant changes in social behavior, or in their subjective emotional state. Patients with unilateral lesions of the anteroventral part of the anterior cingulate cortex and/or medial prefrontal cortex area BA9 were in some cases impaired on voice and face expression identification, had some change in social behavior, and had significant changes in their subjective emotional state. Patients with dorsolateral prefrontal cortex lesions or with medial lesions outside the anterior cingulate cortex and medial prefrontal BA9 areas were unimpaired on any of these measures of emotion. In all cases in which voice expression identification was impaired, there were no deficits in control tests of the discrimination of unfamiliar voices and the recognition of environmental sounds.

These results (Hornak et al. 2003) thus confirm that damage restricted to the orbitofrontal cortex can produce impairments in face and voice expression identification, which may be primary reinforcers. The system is sensitive, in that even patients with unilateral orbitofrontal cortex lesions may be impaired. The impairment is not a generic impairment of the ability to recognize any emotions in others, in that frequently voice but not face expression identification was impaired, and vice versa. This implies some functional specialization for visual versus auditory emotion-related processing in the human orbitofrontal cortex. The results

also show that the changes in social behavior can be produced by damage restricted to the orbitofrontal cortex. The patients were particularly likely to be impaired on emotion recognition (they were less likely to notice when others were sad, or happy, or disgusted); on emotional empathy (they were less likely to comfort those who are sad, or afraid, or to feel happy for others who are happy); on interpersonal relationships (not caring what others think, and not being close to his/her family); and were less likely to cooperate with others; were impatient and impulsive; and had difficulty in making and keeping close relationships. The results also show that changes in subjective emotional state (including frequently sadness, anger, and happiness) can be produced by damage restricted to the orbitofrontal cortex (Hornak et al. 2003). In addition, the patients with bilateral orbitofrontal cortex lesions were impaired on the probabilistic reversal learning task (Hornak et al. 2004). The findings overall thus make clear the types of deficit found in humans with orbitofrontal cortex damage, and can be directly related to underlying fundamental processes in which the orbitofrontal cortex is involved (see Rolls 2005), including decoding and representing primary reinforcers (including face expression), being sensitive to changes in reinforcers, and rapidly readjusting behaviour to stimuli when the reinforcers available change.

The results (Hornak et al. 2003) also extend these investigations to the anterior cingulate cortex (including some of medial prefrontal cortex area BA9) by showing that lesions in these regions can produce voice and/or face expression identification deficits and marked changes in subjective emotional state.

It is of interest that the range of face expressions for which identification is impaired by orbitofrontal cortex damage (Hornak et al. 1996; Hornak et al. 2003; Rolls 1999a) is more extensive than the impairment in identifying primarily fear face expressions produced by amygdala damage in humans (Adolphs et al. 2002; Calder et al. 1996) (for review, see Rolls 2005). In addition, the deficits in emotional and social behavior described above that are produced by orbitofrontal cortex damage in humans seem to be more pronounced than changes in emotional behavior produced by amygdala damage in humans, although deficits in autonomic conditioning can be demonstrated (Phelps 2004). This result suggests that in humans and other primates the orbitofrontal cortex may become more important than the amygdala in emotion, and possible reasons for this, including the more powerful architecture for rapid learning and reversal that may be facilitated by the functional architecture of the neocortex with its highly developed recurrent collateral connections, which may help to support short-term memory attractor states, are considered by Rolls (2005).

19. Conclusions

Neurophysiological investigations of the inferior temporal cortex are revealing at least part of the way in which neuronal firing encodes information about faces and objects and are showing that one representation implements several types of invariance. The representation found has clear utility for the receiving Y2

networks. These neurophysiological findings are stimulating the development of computational neuronal network models, which suggest that part of the process involves the operation of a modified Hebb learning rule with a short-term memory trace to help the system learn invariances from the statistical properties of the inputs it receives. Neurons in the inferior temporal cortex, which encode the identity of faces and have considerable invariance and a sparse distributed representation, are ideal as an input to stimulus–reinforcer association learning mechanisms in the orbitofrontal cortex and amygdala that enable appropriate emotional and social responses to be made to different individuals. The neurons in the cortex in the superior temporal sulcus, which respond to face expression, or for other neurons to eye gaze, or for others to head movement, encode reinforcement-related information that is important in making the correct emotional and social responses to a face. Neurons of both these main types are also found in the orbitofrontal cortex (Rolls et al. 2006a) and are important in human social and emotional behavior, which is changed after damage to the orbitofrontal cortex. A more comprehensive description of the reinforcement-related signals and processing in brain regions such as the orbitofrontal cortex that are important in emotional and social behavior, and how these depend on inputs from the temporal cortex visual areas, is provided in *Emotion Explained* (Rolls 2005).

Acknowledgments. The author has worked on some of the investigations described here with N. Aggelopoulos, P. Azzopardi, G.C. Baylis, H. Critchley, G. Deco, P. Földiák, L. Franco, M. Hasselmo, J.Hornak, M. Kringelbach, C.M. Leonard, T.J. Milward, D.I. Perrett, S.M. Stringer, M.J. Tovee, T. Trappenberg, A. Treves, and G. Wallis, and their collaboration is sincerely acknowledged. Different parts of the research described were supported by the Medical Research Council, PG8513790; by a Human Frontier Science Program grant; by an EC Human Capital and Mobility grant; by the MRC Oxford Interdisciplinary Research Centre in Cognitive Neuroscience; and by the Oxford McDonnell-Pew Centre in Cognitive Neuroscience.

5 *References*

Abbott LF, Rolls ET, Tovee MJ (1996) Representational capacity of face coding in monkeys. Cereb Cortex 6:498–505

Adolphs R, Baron-Cohen S, Tranel D (2002) Impaired recognition of social emotions following amygdala damage. J Cognit Neurosci 14:1264–1274

Adolphs R, Tranel D, Damasio H, Damasio AR (1995) Fear and the human amygdala. J Neurosci 15:5879–5891

Aggelopoulos NC, Rolls ET (2005) Natural scene perception: inferior temporal cortex neurons encode the positions of different objects in the scene. Eur J Neurosci 22:2903–2916

Aggelopoulos NC, Franco L, Rolls ET (2005) Object perception in natural scenes: encoding by inferior temporal cortex simultaneously recorded neurons. J Neurophysiol 93: 1342–1357

Y2

Baddeley RJ, Abbott LF, Booth MJA, Sengpiel F, Freeman T, Wakeman EA, Rolls ET (1997) Responses of neurons in primary and inferior temporal visual cortices to natural scenes. Proc R Soc Lond B 264:1775–1783
Baizer JS, Ungerleider LG, Desimone R (1991) Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. J Neurosci 11:168–190
Ballard DH (1990) Animate vision uses object-centred reference frames. In: Eckmiller R (ed) Advanced neural computers. North-Holland, Amsterdam, pp 229–236
Ballard DH (1993) Subsymbolic modelling of hand-eye coordination. In: Broadbent DE (ed) The simulation of human intelligence. Blackwell, Oxford, pp 71–102
Barlow HB (1972) Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1:371–394
Baylis GC, Rolls ET (1987) Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp Brain Res 65:614–622
Baylis GC, Rolls ET, Leonard CM (1985) Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. Brain Res 342:91–102
Baylis GC, Rolls ET, Leonard CM (1987) Functional subdivisions of the temporal lobe neocortex. J Neurosci 7:330–342
Biederman I (1972) Perceiving real-world scenes. Science 177: 77–80
Booth MCA, Rolls ET (1998) View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cereb Cortex 8:510–523
Boussaoud D, Desimone R, Ungerleider LG (1991) Visual topography of area TEO in the macaque. J Comp Neurol 306:554–575
Brothers L, Ring B, Kling A (1990) Response of neurons in the macaque amygdala to complex social stimuli. Behav Brain Res 41:199–213
Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. J Neurophysiol 46:369–384
Calder AJ, Young AW, Rowland D, Perrett DI, Hodges JR, Etcoff NL (1996) Facial emotion recognition after bilateral amygdala damage: differentially severe impairment of fear. Cognit Neuropsychol 13:699–745
Chelazzi L, Miller E, Duncan J, Desimone R (1993) A neural basis for visual search in inferior temporal cortex. Nature (Lond) 363:345–347
Corchs S, Deco G (2002) Large-scale neural model for visual attention: integration of experimental single-cell and fMRI data. Cereb Cortex 12:339–348
Cowey A, Rolls ET (1975) Human cortical magnification factor and its relation to visual acuity. Exp Brain Res 21:447–454
Deco G, Lee TS (2002) A unified model of spatial and object attention based on intercortical biased competition. Neurocomputing 44–46:769–774
Deco G, Rolls ET (2002) Object-based visual neglect: a computational hypothesis. Eur J Neurosci 16:1994–2000
Deco G, Rolls ET (2003) Attention and working memory: a dynamical model of neuronal activity in the prefrontal cortex. Eur J Neurosci 18: 2374–2390
Deco G, Rolls ET (2004) A neurodynamical cortical model of visual attention and invariant object recognition. Vision Res 44:621–644
Deco G, Rolls ET (2005a) Attention, short-term memory, and action selection: a unifying theory. Prog Neurobiol 76:236–256
Deco G, Rolls ET (2005b) Synaptic and spiking dynamics underlying reward reversal in orbitofrontal cortex. Cereb Cortex 15:15–30
Deco G, Rolls ET (2005c) Neurodynamics of biased competition and co-operation for attention: a model with spiking neurons. J Neurophysiol 94:295–313

Y2

Deco G, Rolls ET (2006) Decision-making and Weber's law: a neurophysiological model. Eur J Neurosci (in press)
Deco G, Zihl J (2001) Top-down selective visual attention: a neurodynamical approach. Visual Cognit 8:119–140
Desimone R (1991) Face-selective cells in the temporal cortex of monkeys. J Cognit Neurosci 3:1–8
Desimone R, Gross CG (1979) Visual areas in the temporal cortex of the macaque. Brain Res 178:363–380
Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222
Desimone R, Albright TD, Gross CG, Bruce CJ (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:2051–2062
Dolan RJ, Fink GR, Rolls ET, Booth M, Holmes A, Frackowiak RSJ, Friston KJ (1997) How the brain learns to see objects and faces in an impoverished context. Nature (Lond) 389:596–599
Elliffe MCM, Rolls ET, Stringer SM (2002) Invariant recognition of feature combinations in the visual system. Biol Cybern 86:59–71
Földiák P (1991) Learning invariance from transformation sequences. Neural Comput 3:194–200
Franco L, Rolls ET, Aggelopoulos NC, Treves A (2004) The use of decoding to analyze the contribution to the information of the correlations between the firing of simultaneously recorded neurons. Exp Brain Res 155:370–384
Franco L, Rolls ET, Aggelopoulos NC, Jerez JM (2006) Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex.
Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 36:193–202
Fukushima K (1989) Anlaysis of the process of visual pattern recognition by the neocognitron. Neural Networks 2:413–420
Fukushima K (1991) Neural networks for visual pattern recognition. IEEE Trans 74:179–190
Gawne TJ, Richmond BJ (1993) How independent are the messages carried by adjacent inferior temporal cortical neurons? J Neurosci 13:2758–2771
Georges-François P, Rolls ET, Robertson RG (1999) Spatial view cells in the primate hippocampus: allocentric view not head direction or eye position or place. Cereb Cortex 9:197–212
Grill-Spector K, Malach R (2004) The human visual cortex. Annu Rev Neurosci 27:649–677
Gross CG, Desimone R, Albright TD, Schwartz EL (1985) Inferior temporal cortex and pattern recognition. Exp Brain Res 11:179–201
Hasselmo ME, Rolls ET, Baylis GC (1989a) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. Behav Brain Res 32:203–218
Hasselmo ME, Rolls ET, Baylis GC, Nalwa V (1989b) Object-centred encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. Exp Brain Res 75:417–429
Haxby JV, Hoffman EA, Gobbini MI (2002) Human neural systems for face recognition and social communication. Biol Psychiatry 51:59–67
Hornak J, Rolls ET, Wade D (1996) Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. Neuropsychologia 34:247–261

Y2

Hornak J, Bramham J, Rolls ET, Morris RG, O'Doherty J, Bullock PR, Polkey CE (2003) Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. Brain 126:1691–1712

Hornak J, O'Doherty J, Bramham J, Rolls ET, Morris RG, Bullock PR, Polkey CE (2004) Reward-related reversal learning after surgical excisions in orbitofrontal and dorsolateral prefrontal cortex in humans. J Cognit Neurosci 16:463–478

Koenderink JJ, Van Doorn AJ (1979) The internal representation of solid shape with respect to vision. Biol Cybern 32:211–217

Kringelbach ML, Rolls ET (2003) Neural correlates of rapid reversal learning in a simple model of human social interaction. Neuroimage 20:1371–1383

Kringelbach ML, Rolls ET (2004) The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Prog Neurobiol 72:341–372

Leonard CM, Rolls ET, Wilson FAW, Baylis GC (1985) Neurons in the amygdala of the monkey with responses selective for faces. Behav Brain Res 15:159–176

Logothetis NK, Sheinberg DL (1996) Visual object recognition. Annu Rev Neurosci 19:577–621

Logothetis NK, Pauls J, Bülthoff HH, Poggio T (1994) View-dependent object recognition by monkeys. Curr Biol 4:401–414

Marr D (1982) Vision. Freeman, San Francisco

Martinez-Trujillo J, Treue S (2002) Attentional modulation strength in cortical area MT depends on stimulus contrast. Neuron 35:365–370

Maunsell JH, Newsome WT (1987) Visual processing in monkey extrastriate cortex. Annu Rev Neurosci 10:363–401

Mikami A, Nakamura K, Kubota K (1994) Neuronal responses to photographs in the superior temporal sulcus of the rhesus monkey. Behav Brain Res 60:1–13

Miller EK, Desimone R (1994) Parallel neuronal mechanisms for short-term memory. Science 263:520–522

Miyashita Y (1993) Inferior temporal cortex: where visual perception meets memory. Annu Rev Neurosci 16:245–263

Mozer M (1991) The perception of multiple objects: a connectionist approach. MIT Press, Cambridge

Panzeri S, Biella G, Rolls ET, Skaggs WE, Treves A (1996) Speed, noise, information and the graded nature of neuronal responses. Network 7:365–370

Panzeri S, Treves A, Schultz S, Rolls ET (1999a) On decoding the responses of a population of neurons from short time epochs. Neural Comput 11:1553–1577

Panzeri S, Schultz SR, Treves A, Rolls ET (1999b) Correlations and the encoding of information in the nervous system. Proc R Soc Lond B 266:1001–1012

Panzeri S, Rolls ET, Battaglia F, Lavis R (2001) Speed of feed-forward and recurrent processing in multilayer networks of integrate-and-fire neurons. Network Comput Neural Syst 12:423–440

Perrett DI, Rolls ET, Caan W (1982) Visual neurons responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329–342

Perrett DI, Smith PA, Potter DD, Mistlin AJ, Head AS, Milner AD, Jeeves MA (1985a) Visual cells in the temporal cortex sensitive to face view and gaze direction. Proc R Soc Lond B 223:293–317

Perrett DI, Smith PAJ, Mistlin AJ, Chitty AJ, Head AS, Potter DD, Broennimann R, Milner AD, Jeeves MA (1985b) Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: a preliminary report. Behav Brain Res 16:153–170

Y2

Perrett D, Mistlin A, Chitty A (1987) Visual neurons responsive to faces. Trends Neurosci 10:358–364
Phelps EA (2004) Human emotion and memory: interactions of the amygdala and hippocampal complex. Curr Opin Neurobiol 14:198–202
Poggio T, Edelman S (1990) A network that learns to recognize three-dimensional objects. Nature (Lond) 343:263–266
Renart A, Parga N, Rolls ET (2000) A recurrent model of the interaction between the prefrontal cortex and inferior temporal cortex in delay memory tasks. In: Solla SA, Leen TK, Mueller KR (eds) Advances in neural information processing systems, vol 12. MIT Press, Cambridge, pp 171–177
Renart A, Moreno R, de la Rocha J, Parga N, Rolls ET (2001) A model of the IT-PF network in object working memory which includes balanced persistent activity and tuned inhibition. Neurocomputing 38–40:1525–1531
Reynolds J, Desimone R (1999) The role of neural mechanisms of attention in solving the binding problem. Neuron 24:19–29
Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. J Neurosci 19:1736–1753
Robertson RG, Rolls ET, Georges-François P (1998) Spatial view cells in the primate hippocampus: effects of removal of view details. J Neurophysiol 79:1145–1156
Rolls ET (1981) Responses of amygdaloid neurons in the primate. In: Ben-Ari Y (ed) The amygdaloid complex. Elsevier, Amsterdam, pp 383–393
Rolls ET (1984) Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. Human Neurobiol 3:209–222
Rolls ET (1986a) A theory of emotion, and its application to understanding the neural basis of emotion. In: Oomura Y (ed) Emotions. Neural and chemical control. Karger, Basel, pp 325–344
Rolls ET (1986b) Neural systems involved in emotion in primates. In: Plutchik R, Kellerman H (eds) Emotion: theory, research, and experience, vol 3. Biological foundations of emotion. Academic Press, New York, pp 125–143
Rolls ET (1989a) Functions of neuronal networks in the hippocampus and neocortex in memory. In: Byrne JH, Berry WO (eds) Neural models of plasticity: experimental and theoretical approaches. Academic Press, San Diego, pp 240–265
Rolls ET (1989b) The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus. In: Durbin R, Miall C, Mitchison G (eds) The computing neuron. Addison-Wesley, Wokingham, England, pp 125–159
Rolls ET (1990) A theory of emotion, and its application to understanding the neural basis of emotion. Cognit Emotion 4:161–190
Rolls ET (1991) Neural organisation of higher visual functions. Curr Opin Neurobiol 1:274–278
Rolls ET (1992a) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Philos Trans R Soc Lond B 335:11–21
Rolls ET (1992b) Neurophysiology and functions of the primate amygdala. In: Aggleton JP (ed) The amygdala. Wiley-Liss, New York, pp 143–165
Rolls ET (1997) A neurophysiological and computational approach to the functions of the temporal lobe cortical visual areas in invariant object recognition. In: Jenkin M, Harris L (eds) Computational and psychophysical mechanisms of visual coding. Cambridge University Press, Cambridge, pp 184–220
Rolls ET (1999a) The functions of the orbitofrontal cortex. Neurocase 5:301–312
Rolls ET (1999b) The brain and emotion. Oxford University Press, Oxford

Y2

Rolls ET (1999c) Spatial view cells and the representation of place in the primate hippocampus. Hippocampus 9:467–480
Rolls ET (2000a) Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron 27:205–218
Rolls ET (2000b) Neurophysiology and functions of the primate amygdala, and the neural basis of emotion. In: Aggleton JP (ed) The amygdala: a functional analysis, 2nd edn. Oxford University Press, Oxford, pp 447–478
Rolls ET (2003) Consciousness absent and present: a neurophysiological exploration. Prog Brain Res 144:95–106
Rolls ET (2005) Emotion explained. Oxford University Press, Oxford
Rolls ET (2006) The representation of information about faces in the temporal and frontal lobes. Neuropsychologia (in press)
Rolls ET, Baylis GC (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. Exp Brain Res 65:38–48
Rolls ET, Cowey A (1970) Topography of the retina and striate cortex and its relationship to visual acuity in rhesus monkeys and squirrel monkeys. Exp Brain Res 10:298–310
Rolls ET, Deco G (2002) Computational neuroscience of vision. Oxford University Press, Oxford
Rolls ET, Kesner RP (2006) A computational theory of hippocampal function, and empirical tests of the theory. Prog Neurobiol (in press)
Rolls ET, 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 Comput 12:2547–2572
Rolls ET, Stringer SM (2001) Invariant object recognition in the visual system with error correction and temporal difference learning. Network Comput Neural Syst 12:111–129
Rolls ET, Stringer SM (2006) Invariant global motion recognition in the dorsal visual system: a unifying theory. Neural Comput (in press)
Rolls ET, Tovee MJ (1994) Processing speed in the cerebral cortex and the neurophysiology of visual masking. Proc R Soc Lond B 257:9–15
Rolls ET, Tovee MJ (1995a) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol 73:713–726
Rolls ET, Tovee MJ (1995b) The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the visual field. Exp Brain Res 103:409–420
Rolls ET, Treves A (1990) The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. Network 1:407–421
Rolls ET, Treves A (1998) Neural networks and brain function. Oxford University Press, Oxford
Rolls ET, Xiang J-Z (2005) Reward-spatial view representations and learning in the hippocampus. J Neurosci 25:6167–6174
Rolls ET, Xiang J-Z (2006) Spatial view cells in the primate hippocampus, and memory recall. Rev Neurosci 17:175–200
Rolls ET, Baylis GC, Leonard CM (1985) Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus in the monkey. Vision Res 25:1021–1035
Rolls ET, Baylis GC, Hasselmo ME (1987) The responses of neurons in the cortex in the superior temporal sulcus of the monkey to band-pass spatial frequency filtered faces. Vision Res 27:311–326

Y2

Rolls ET, Baylis GC, Hasselmo M, Nalwa V (1989a) The representation of information in the temporal lobe visual cortical areas of macaque monkeys. In: Kulikowski JJ, Dickinson CM, Murray IJ (eds) Seeing contour and colour. Pergamon, Oxford

Rolls ET, Baylis GC, Hasselmo ME, Nalwa V (1989b) The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. Exp Brain Res 76:153–164

Rolls ET, Tovee MJ, Purcell DG, Stewart AL, Azzopardi P (1994) The responses of neurons in the temporal cortex of primates, and face identification and detection. Exp Brain Res 101:473–484

Rolls ET, Critchley HD, Treves A (1996) The representation of olfactory information in the primate orbitofrontal cortex. J Neurophysiol 75:1982–1996

Rolls ET, Treves A, Tovee MJ (1997a) The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. Exp Brain Res 114:177–185

Rolls ET, Robertson RG, Georges-François P (1997b) Spatial view cells in the primate hippocampus. Eur J Neurosci 9:1789–1794

Rolls ET, Treves A, Robertson RG, Georges-François P, Panzeri S (1998) Information about spatial view in an ensemble of primate hippocampal cells. J Neurophysiol 79:1797–1813

Rolls ET, Tovee MJ, Panzeri S (1999) The neurophysiology of backward visual masking: information analysis. J Cognit Neurosci 11:335–346

Rolls ET, Aggelopoulos NC, Zheng F (2003a) The receptive fields of inferior temporal cortex neurons in natural scenes. J Neurosci 23:339–348

Rolls ET, Franco L, Aggelopoulos NC, Reece S (2003b) An information theoretic approach to the contributions of the firing rates and correlations between the firing of neurons. J Neurophysiol 89:2810–2822

Rolls ET, Aggelopoulos NC, Franco L, Treves A (2004) Information encoding in the inferior temporal cortex: contributions of the firing rates and correlations between the firing of neurons. Biol Cybern 90:19–32

Rolls ET, Xiang J-Z, Franco L (2005) Object, space and object-space representations in the primate hippocampus. J Neurophysiol 94:833–844

Rolls ET, Critchley HD, Browning AS, Inoue K (2006a) Face-selective and auditory neurons in the primate orbitofrontal cortex. Exp Brain Res 170:74–87

Rolls ET, Franco L, Aggelopoulos NC, Perez JM (2006b) Information in the first spike, the order of spikes, and the number of spikes provided by neurons in the inferior temporal visual cortex. Vision Res (in press)

Sato T (1989) Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. Exp Brain Res 77:23–30

Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. Brain Res 149:1–24

Singer W (1999) Neuronal synchrony: a versatile code for the definition of relations? Neuron 24:49–65

Singer W, Gray CM (1995) Visual feature integration and the temporal correlation hypothesis. Annu Rev Neurosci 18:555–586

Spiridon M, Kanwisher N (2002) How distributed is visual category information in human occipito-temporal cortex? An fMRI study. Neuron 35:1157–1165

Stringer SM, Rolls ET (2000) Position invariant recognition in the visual system with cluttered environments. Neural Networks 13:305–315

Y2

Stringer SM, Rolls ET (2002) Invariant object recognition in the visual system with novel views of 3D objects. Neural Comput 14:2585–2596
Stringer SM, Perry G, Rolls ET, Proske JH (2006) Learning invariant object recognition in the visual system with continuous transformations. Biol Cybern 94:128–142
Sutton RS, Barto AG (1998) Reinforcement learning. MIT Press, Cambridge
Tanaka K (1993) Neuronal mechanisms of object recognition. Science 262:685–688
Tanaka K (1996) Inferotemporal cortex and object vision. Annu Rev Neurosci 19: 109–139
Tanaka K, Saito C, Fukada Y, Moriya M (1990) Integration of form, texture, and color information in the inferotemporal cortex of the macaque. In: Iwai E, Mishkin M (eds) Vision, memory and the temporal lobe. Elsevier, New York, pp 101–109.
Thorpe SJ, Imbert M (1989) Biological constraints on connectionist models. In: Pfeifer R, Schreter Z, Fogelman-Soulie F (eds) Connectionism in perspective. Elsevier, Amsterdam, pp 63–92
Thorpe SJ, Rolls ET, Maddison S (1983) Neuronal activity in the orbitofrontal cortex of the behaving monkey. Exp Brain Res 49:93–115
Tovee MJ, Rolls ET (1995) Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. Visual Cognit 2:35–58
Tovee MJ, Rolls ET, Treves A, Bellis RP (1993) Information encoding and the responses of single neurons in the primate temporal visual cortex. J Neurophysiol 70:640–654
Tovee MJ, Rolls ET, Azzopardi P (1994) Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. J Neurophysiol 72:1049–1060
Tovee MJ, Rolls ET, Ramachandran VS (1996) Rapid visual learning in neurones of the primate temporal visual cortex. Neuroreport 7:2757–2760
Trappenberg TP, Rolls ET, Stringer SM (2002) Effective size of receptive fields of inferior temporal cortex neurons in natural scenes. In: Dietterich TG, Becker S, Ghahramani Z (eds) Advances in neural information processing systems, 14, vol 1. MIT Press, Cambridge, pp 293–300
Treves A (1993) Mean-field analysis of neuronal spike dynamics. Network 4:259–284
Treves A, Rolls ET (1991) What determines the capacity of autoassociative memories in the brain? Network 2:371–397
Treves A, Rolls ET, Tovee MJ (1996) On the time required for recurrent processing in the brain. In: Torre V, Conti F (eds) Neurobiology: ionic channels, neurons, and the brain. Plenum, New York, pp 325–353
Treves A, Rolls ET, Simmen M (1997) Time for retrieval in recurrent associative memories. Physica D 107:392–400
Treves A, Panzeri S, Rolls ET, Booth M, Wakeman EA (1999) Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. Neural Computat 11:611–641
Ullman S (1996) High-level vision: object recognition and visual cognition. Bradford/MIT Press, Cambridge
Usher M, Niebur E (1996) Modelling the temporal dynamics of IT neurons in visual search: a mechanism for top-down selective attention. J Cognit Neurosci 8:311–327
von der Malsburg C (1990) A neural architecture for the representation of scenes. In: McGaugh JL, Weinberger NM, Lynch G (eds) Brain organisation and memory: cells, systems and circuits. Oxford University Press, New York, pp 356–372
Wallis G, Rolls ET (1997) Invariant face and object recognition in the visual system. Prog Neurobiol 51:167–194

Y2

Wallis G, Rolls ET, Földiák P (1993) Learning invariant responses to the natural transformations of objects. In: International Joint Conference on Neural Networks, vol 2, pp 1087–1090

Williams GV, Rolls ET, Leonard CM, Stern C (1993) Neuronal responses in the ventral striatum of the behaving macaque. Behav Brain Res 55:243–252

Wilson FAW, O'Scalaidhe SPO, Goldman-Rakic PS (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex. Science 260:1955–1958

Xiang J-Z, Brown MW (1998) Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. Neuropharmacology 37:657–676

Yamane S, Kaji S, Kawano K (1988) What facial features activate face neurons in the inferotemporal cortex of the monkey? Exp Brain Res 73:209–214

Y2