

# The representation of information about faces in the temporal and frontal lobes

Edmund T. Rolls\*

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

Available online 23 June 2006

## Abstract

Neurophysiological evidence is described showing that some neurons in the macaque inferior temporal visual cortex have responses that are invariant with respect to the position, size and view of faces and objects, and that these neurons show rapid processing and rapid learning. Which face or object is present is encoded using a distributed representation in which each neuron conveys independent information in its firing rate, with little information evident in the relative time of firing of different neurons. This ensemble encoding has the advantages of maximising the information in the representation useful for discrimination between stimuli using a simple weighted sum of the neuronal firing by the receiving neurons, generalisation and graceful degradation. 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, generalise to other views of the same face. A theory is described of how such invariant representations may be produced in a hierarchically organised set of visual cortical areas with convergent connectivity. The theory proposes that neurons in these visual areas use a modified Hebb synaptic modification rule with a short-term memory trace to capture whatever can be captured at each stage that is invariant about objects as the objects change in retinal view, position, size and rotation. Another population of neurons in the cortex in the superior temporal sulcus encodes other aspects of faces such as face expression, eye gaze, face view and whether the head is moving. These neurons thus provide important additional inputs to parts of the brain such as the orbitofrontal cortex and amygdala that are involved in social communication and emotional behaviour.

Outputs of these systems reach the amygdala, in which face-selective neurons are found, and also the orbitofrontal cortex, in which some neurons are tuned to face identity and others to face expression. In humans, activation of the orbitofrontal cortex is found when a change of face expression acts as a social signal that behaviour should change; and damage to the orbitofrontal cortex can impair face and voice expression identification, and also the reversal of emotional behaviour that normally occurs when reinforcers are reversed.

© 2006 Published by Elsevier Ltd.

*Keywords:* Face expression; Face identity; Inferior temporal visual cortex; Orbitofrontal cortex

## 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 non-human primates, macaques, firstly because the temporal lobe, in which this processing occurs, is much more developed than in non-primates, and secondly because the findings are relevant to understanding the effects of brain damage in patients, as will be shown. In this paper, attention will be 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 behaviour; and because of the application of understanding this neural system to understanding the effects of damage to this system in humans. It will also be 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

\* Tel.: +44 1865 271348; fax: +44 1865 310447.  
*E-mail address:* Edmund.Rolls@psy.ox.ac.uk.  
*URL:* www.cns.ox.ac.uk.

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, generalise to other views of the same face. Moreover, these inferior temporal cortex neurons have sparse distributed representations of faces, which are shown to be well suited as inputs to the stimulus–reinforcer association learning mechanisms in the orbitofrontal cortex and amygdala which 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, and how the brain may learn these invariant representations of faces, are described in Sections 3–12. Section 13 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 behaviour. 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 towards or away from the receiver. In Sections 14 and 15, I describe the responses of face-selective neurons in two areas, the amygdala and orbitofrontal cortex, to which the temporal cortical areas have direct projections. I also review evidence (in Section 15) 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 behaviour 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 behaviour. 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.

## 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, Ungerleider, & Desimone, 1991; Maunsell & Newsome, 1987; Seltzer & Pandya, 1978) in which some neurons respond selectively to faces (Bruce, Desimone, & Gross, 1981; Desimone, 1991; Desimone, Albright, Gross, & Bruce, 1984; Desimone & Gross, 1979; Gross, Desimone, Albright, & Schwartz, 1985; Perrett, Rolls, & Caan, 1982; Rolls, 1981, 1984, 1991, 1992a, 2000a, 2005; Rolls & 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

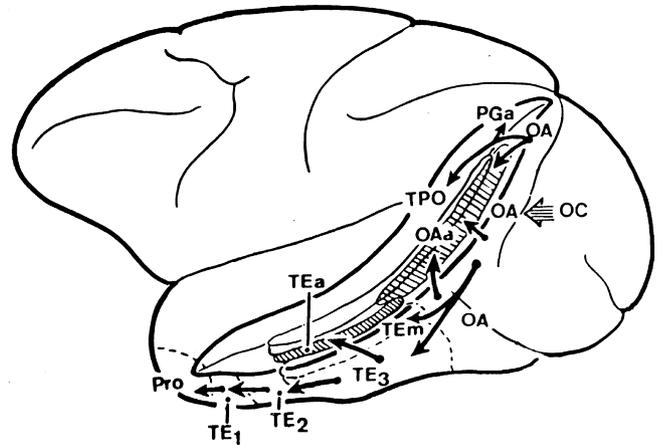


Fig. 1. Lateral view of the macaque brain (left) and coronal section (right) showing the different architectonic areas (e.g. TEm, TPO) in and bordering the anterior part of the superior temporal sulcus (STS) of the macaque (see text). (After Seltzer & Pandya, 1978.).

and TE1. In addition there is a set of different areas in the cortex in the superior temporal sulcus (Baylis, Rolls, & Leonard, 1987; Seltzer & Pandya, 1978) (see 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 & Pandya, 1978).

Considerable specialisation 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 specialised 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 4–10%. The stimuli which 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, Saito, Fukada, & Moriya, 1990). Due to the fact that 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, Rolls, & Baylis, 1989).

In human fMRI studies, evidence for specialisation of function is described (Grill-Spector & Malach, 2004; Haxby, Hoffman, & Gobbini, 2002; O'Toole, Jiang, Abdi, & Haxby, 2005; Spiridon, Fischl, & Kanwisher, 2006; Spiridon & Kanwisher, 2002) related to face processing (in 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 areas which may correspond to the macaque inferior temporal cortex in which object but not face representations are common, as described above); 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, Rolls, & Robertson, 1999; Robertson, Rolls, & Georges-François, 1998; Rolls, 1999c; Rolls, Robertson, & Georges-François, 1997; Rolls, Treves, Robertson, Georges-François, & Panzeri, 1998; Rolls & Xiang, 2005; Rolls, Xiang, & Franco, 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 cytoarchitecturally 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-organising 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 & 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 2–20 times more (and statistically significantly more) to faces than to a wide

range of gratings, simple geometrical stimuli, or complex 3D objects (Baylis, Rolls, & Leonard, 1985; Baylis et al., 1987; Rolls, 1984, 1992a, 1997, 2000a; Rolls & Deco, 2002). 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 & 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 which 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 specialised 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 & Tovee, 1995a; Rolls, Treves, & Tovee, 1997). These information theoretic procedures provide an objective and quantitative way to show what is “represented” by a particular population of neurons.

### 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 organisation 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, Tovee, Purcell, Stewart, & Azzopardi, 1994). These findings are consistent with the hypotheses developed below that by competitive self-organisation 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, Rolls, & Stringer, 2002; Rolls & 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. (e.g. 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 group 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 et al. (1985) showed this with the responses of temporal cortical neurons that typically responded to several members of a set of five 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 & Tovee, 1995a), with the average sparseness being 0.65. (The sparseness of the representation provided by a neuron can be defined as

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

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, Rolls, Aggelopoulos, and Jerez (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, i.e. 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 & Tovee, 1995a; Treves, Panzeri, Rolls, Booth, & Wakeman, 1999) (see 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 non-linearity of neurons combined with short-term variability in the responses of neurons (Treves et al., 1999).

Complementary evidence comes from applying information theory to analyse 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 two 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 if the encoding was local (or grandmother cell-like), 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

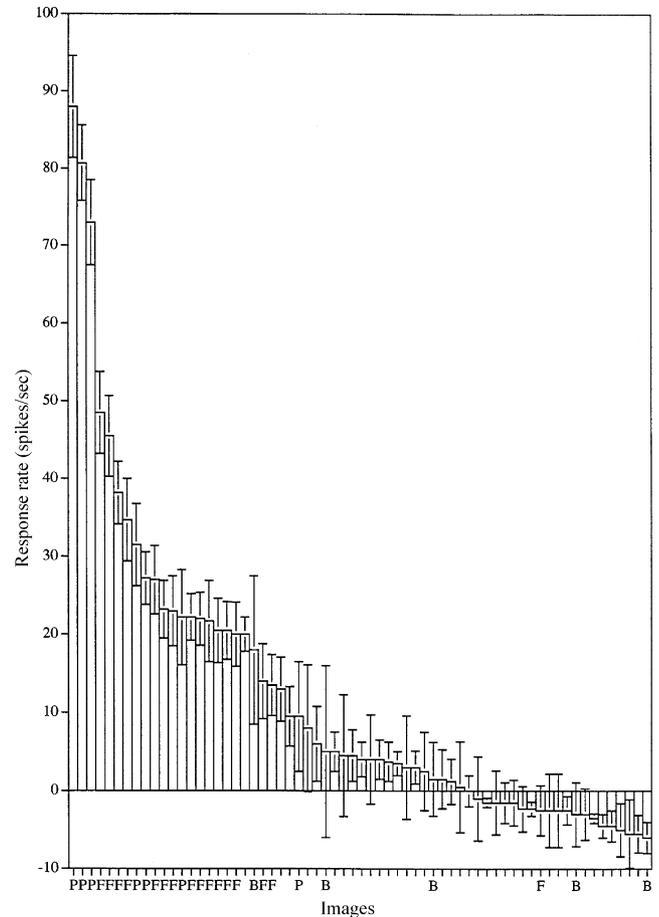


Fig. 2. Firing rate distribution of a single neuron in the temporal visual cortex to a set of 23 face (F) and 45 non-face images of natural scenes. The firing rate to each of the 68 stimuli is shown. P indicates a face profile stimulus, a B a body part stimulus such as a hand. (After Rolls & Tovee, 1995a.).

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 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 below). Second, it is clear (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, Rolls, & Tovee, 1996; Rolls & Treves, 1998; Rolls, Treves et al., 1997).

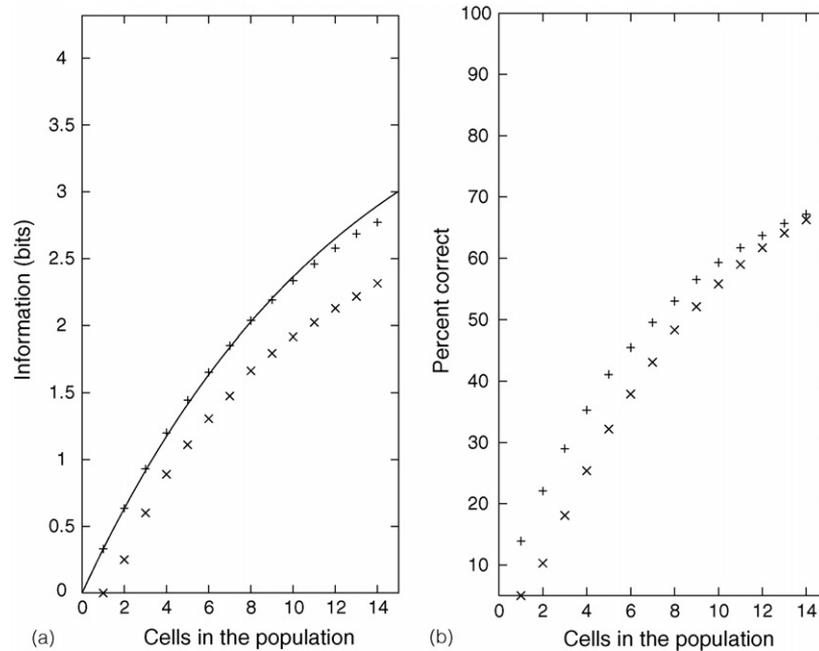


Fig. 3. (a) The values for the average information available in the responses of different numbers of these neurons on each trial, about which of a set of 20 face stimuli has been shown. The decoding method was dot product (DP,  $\times$ ) or probability estimation (PE,  $+$ ), and the effects obtained with cross-validation procedures utilising 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 et al., 1997.).

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 (see 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

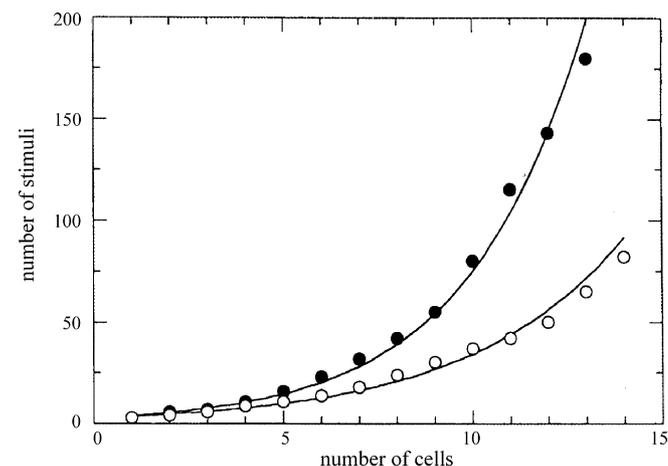


Fig. 4. The number of stimuli (in this case from a set of 20 faces) that are encoded in the responses of different numbers of neurons in the temporal lobe visual cortex, based on the results shown in Fig. 3. The decoding method was dot product (DP, open circle) or probability estimation (PE, filled circle).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

to discriminate between the set of stimuli, which with 20 stimuli is  $\log_2 20 = 4.32$  bits; Abbott et al., 1996.]

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 & 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, Schultz, Treves, & Rolls, 1999; Rolls, Aggelopoulos, Franco, & Treves, 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, Schultz et al. (1999), Rolls, Franco, Aggelopoulos, and Reece (2003) and Franco, Rolls, Aggelopoulos, and Treves (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 synchronised to some but not other stimuli in a set, as postulated by Engel, Konig, Kreiter, Schillen, and Singer (1992). 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 neu-

rons (Franco et al., 2004; Panzeri, Schultz et al., 1999; Rolls et al., 2004; Rolls, Franco et al., 2003). The same result was found in natural scenes in which two test images had to be segmented from a complex background, the features of each object had to be bound together, and the monkey had to perform a visual discrimination task between the two test stimuli (Aggelopoulos, Franco, & Rolls, 2005). Thus, the evidence is that for representations of faces and objects in the inferior temporal visual cortex (and of space in the primate hippocampus and of odours in the orbitofrontal cortex; see Rolls, Critchley, & Treves, 1996; Rolls et al., 1998), most of the information is available in the firing rates of the neurons.

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 the hippocampus, where associations between the visual stimuli and other sensory representations are formed (see Rolls, 2005; Rolls & Deco, 2002). Indeed, tracing this pathway onwards, we have found a population of neurons with face-selective responses in the amygdala (Leonard, Rolls, Wilson, & Baylis, 1985; Rolls, 2000b) and orbitofrontal cortex (Rolls, Critchley, Browning, & Inoue, 2006), 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 which may be important in other behavioural responses to faces, the ventral striatum, and comparable neurons have also been found in the ventral striatum (Williams, Rolls, Leonard, & Stern, 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 & Deco, 2002; Rolls & 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 (in 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 3072 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, in 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 in the order of 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, 1992b; Rolls & Deco, 2002; Rolls & Treves, 1998). It is useful to realise 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 & Treves, 1990, 1998; Treves & 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 & Treves, 1998) is that when one can identify the face or object that is present (from a large set of stimuli, that might be thousands or more) with a realistic number of neurons, say in 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 & 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, odours and tastes), information can be read out without taking into account any aspects of the possible temporal synchronisation between neurons (Engel et al., 1992), or temporal encoding within a spike train (Aggelopoulos et al., 2005; Franco et al., 2004; Panzeri, Schultz et al., 1999; Rolls et al., 2004; Rolls, Franco et al., 2003; Rolls, Treves et al., 1997; Tovee, Rolls, Treves, & Bellis, 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 esti-

mated 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, Treves et al., 1997). 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 normalised 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 utilise 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 & Treves, 1990, 1998; Treves & 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, Biella, Rolls, Skaggs, & Treves, 1996; Rolls & Deco, 2002).

### 6.4. Generalisation

Generalisation to similar stimuli is again a property that arises in neuronal networks if distributed but not if local encoding is used. The generalisation 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 (see further Rolls & Deco, 2002; Rolls & Treves, 1998).

### 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 characterised by recurrent connectivity. It is thought that such networks are important in the cerebral cortex, where the association fibres between nearby pyramidal cells may help the cells to retrieve a representation that depends on many neurons in the network (Rolls & Deco, 2002; Rolls & 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 analyser 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, Treves, Schultz, & Rolls, 1999; Rolls, Treves et al., 1997; Treves, 1993; Treves, Rolls, & Simmen, 1997; Treves, Rolls, & Tovee, 1996). 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 & 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 which 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 generalises correctly to other positions, views, etc., 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 & Baylis, 1986). The median size change tolerated with a response of greater than half the maximal response was 12 times. Also, the neurons typically responded to a face when the information in it had been reduced from 3D to a 2D representation in grey 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 eight 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 eight cycles per face are included (Rolls, Baylis, & Leonard, 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 cannot be predicted from a linear combination of their responses to the narrow band stimuli (Rolls, Baylis, & Hasselmo, 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, Rolls, & Azzopardi, 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 2–5° beyond the edge of a face which subtended 8–17° 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 (cf. 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 non-effective face in the parafovea (centred 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 & 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 non-effective 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, Aggelopoulos, & Zheng, 2003). Thus, the neurons provide information biased towards what is present at the fovea, and not equally about what is present anywhere in the visual field. This makes the interface to action 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 & Deco, 2002). These findings are a step towards understanding how the visual system functions in a normal environment (Gallant, Connor, & Van Essen, 1998; Rolls & Deco, 2002).

## 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, Rolls, Baylis, & Nalwa, 1989), although in most cases (16/18 neurons) the response was not perfectly view-independent. Mixed together in the same cortical regions there are neurons with view-dependent responses (Hasselmo, Rolls, Baylis, & Nalwa, 1989). 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, Smith, Potter et al., 1985). 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, Rolls, Baylis, & Nalwa, 1989). 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 10 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 that 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 & Baylis, 1986).

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

were images of 10 real plastic objects which 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 which had never been seen as real objects. The neurons analysed 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 (e.g. Perrett et al., 1982; Tanaka, 1993, 1996). For example, Mikami, Nakamura, and Kubota (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, Pauls, Bülthoff, and Poggio (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 (see Logothetis & 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. 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 (one to two) presentations of the set (Rolls, Baylis, Hasselmo, & Nalwa, 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 categorisation being performed by self-organising competitive neuronal networks, as described below and elsewhere (Rolls, 1989a; Rolls, Baylis, Hasselmo, & Nalwa, 1989a; Rolls & Deco, 2002; Rolls & Treves, 1998).

Further evidence that these neurons can learn new representations very rapidly comes from an experiment in which binarised

black and white images of faces which blended with the background were used. These did not activate face-selective neurons. Full grey-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 binarised version of the same face was shown next, the neurons responded to it (Tovee, Rolls, & Ramachandran, 1996). This is a direct parallel to the same phenomenon which 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 & Rolls, 1987; Miller & Desimone, 1994; Xiang & Brown, 1998), and neurons in a more ventral cortical area respond during the delay in a short-term memory task (Miyashita, 1993; Renart, Parga, & Rolls, 2000).

### 10. 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 40–50 ms, and in the anterior inferior temporal cortical areas approximately 80–100 ms, each stage may need to perform processing for only 15–30 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 & 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 four 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 post-stimulus period 100–120 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 & 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.

The next approach has been to use a visual backward masking paradigm. In this paradigm there is a brief presentation of a test stimulus which 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 & Tovee, 1994) that when there is no mask inferior temporal cortex neurons respond to a 16 ms presentation of the test stimulus for 200–300 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 12. 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 20–30 ms before their firing was interrupted by the mask (Rolls & Tovee, 1994; Rolls, Tovee, & Panzeri, 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 20–30 ms (although it is true that for conscious perception, the firing needs to occur for 40–50 ms; see Rolls, 2003). This provides a fundamental constraint which must be accounted for in any theory of cortical computation. The results emphasise 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 analog membranes which integrate inputs, and with spontaneously active neurons, do show that such networks can settle very rapidly (Rolls & 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, Rolls, Battaglia, & Lavis, 2001).

### 11. 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, 1989b, 1992a; Rolls & Treves, 1998), lead to the following outline working hypotheses on object (including face) recognition by visual cortical mechanisms (Rolls & Deco, 2002).

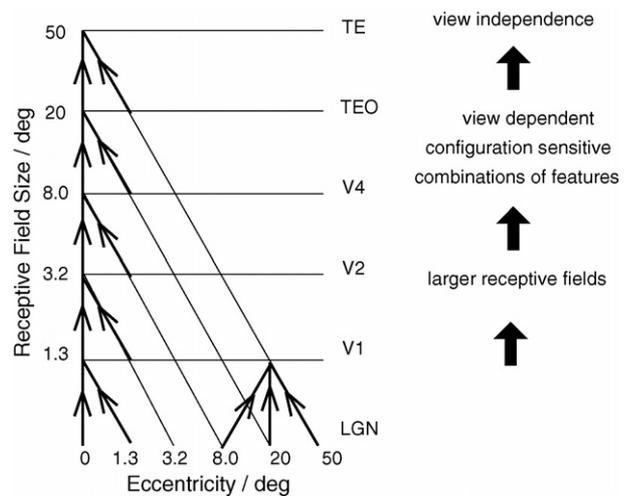


Fig. 5. Schematic diagram showing convergence achieved by the forward projections in the visual system, and the types of representation that may be built by competitive networks operating at each stage of the system from the primary visual cortex (V1) to the inferior temporal visual cortex (area TE) (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).

Cortical visual processing for object recognition is considered to be organised 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 TEb) and anterior temporal cortical areas (e.g. TE2 and TE1). 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 field sizes of neurons (e.g.  $1^\circ$  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.  $8^\circ$  in V4,  $20^\circ$  in TEO and  $50^\circ$  in inferior temporal cortex; Boussaoud, Desimone, & Ungerleider, 1991) (see Fig. 5). Such zones of convergence would overlap continuously with each other (see Fig. 5). 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-organising 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 1–2 mm in the cortex.) These competitive nets operate by a single set of forward inputs leading to (typically non-linear, e.g. sigmoid) activation of output neurons; of competition between the output neurons mediated by a set of feedback inhibitory interneurons which 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 from inactive input axons weaken (Rolls & Deco, 2002; Rolls & Treves, 1998).

Translation, size and view invariance could be computed in such a system by utilising 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 & Deco, 2002; Wallis & 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 analysers 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. 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. In order 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 colours (see references cited by Rolls, 1991); in posterior inferior temporal cortex to stimuli which may require two or more simple features to be present (Tanaka et al., 1990); and in the temporal cortical face processing areas to images that require the presence of several features in a face (such as eyes, hair and mouth) in order to respond (Perrett et al., 1982; Yamane, Kaji, & Kawano, 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 relation to each other. (This is in fact consistent with the data of Tanaka et al. (1990) and with our data on face neurons Rolls, Tovee 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 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 which 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 & Van Doorn, 1979; Logothetis et al., 1994; Poggio & 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 & Rolls, 1998; Hasselmo, Rolls, Baylis, & Nalwa, 1989; Perrett, Mistlin, & Chitty, 1987). This solution to “object-based” representations is very different from that traditionally proposed for artificial vision systems, in which the coordinates in 3D 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 3D space (e.g. 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 3D space to particular parts of, or inside, objects, as the 3D 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 (cf. Ballard, 1990; Rolls & Deco, 2002).

## 12. 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 which implement many of the ideas just described, and which are consistent with and based on much of the neurophysiology summarised above. 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 & Deco, 2002; Rolls & Milward, 2000; Wallis & Rolls, 1997; Wallis, Rolls, & Földiák, 1993).

In the four layer network, the successive layers correspond approximately to V2, V4, the posterior temporal cortex, and

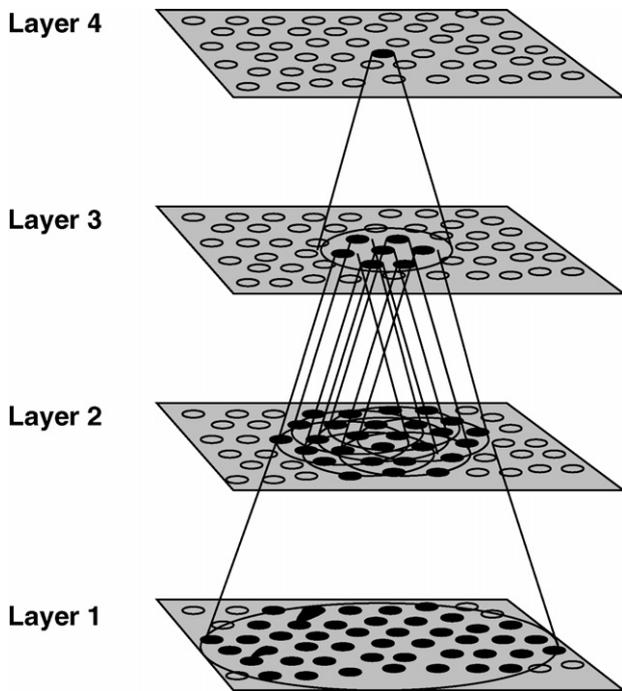


Fig. 6. Hierarchical network structure of VisNet.

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  $32 \times 32$  cells of the preceding layer, with a 67% probability that a connection comes from within 4 cells of the distribution centre. Fig. 6 shows the general convergent network architecture used, and may be compared with Fig. 5. Within each layer, lateral inhibition between neurons has a radius of effect just greater than the radius of feedforward 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 non-linear cell activation function, which also produces contrast enhancement of the firing rates.

In order 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 which provides an approximation to the encoding found in visual area 1 (V1) of the primate visual system.

The synaptic learning rule used can be summarised as follows:

$$\delta w_{ij} = k \cdot m_i \cdot r'_j$$

and

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

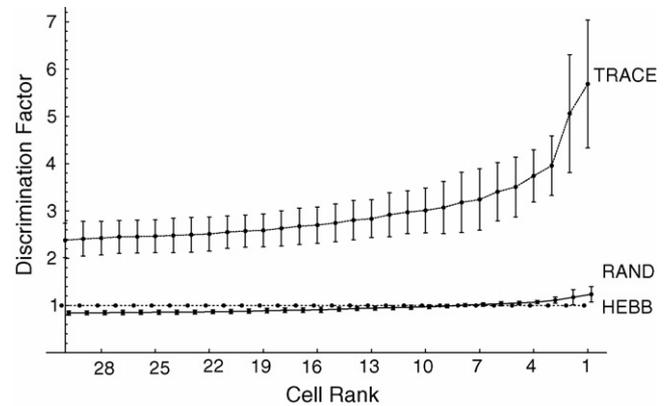


Fig. 7. Comparison of VisNet network discrimination when trained with the trace learning rule, with a Hebb rule (no trace), and when not trained (random) on three stimuli, +, T and L, at nine different locations. (After Wallis & Rolls, 1997.)

where  $r'_j$  is the  $j$ th input to the neuron,  $r_i$  the output of the  $i$ th neuron,  $w_{ij}$  the  $j$ th weight on the  $i$ th neuron,  $\eta$  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 normalisation of each cell’s dendritic weight vector, as in standard competitive learning (see Rolls & Deco, 2002; Rolls & Treves, 1998).

To train the network to produce a translation invariant representation, one stimulus was placed successively in a sequence of seven positions across the input, then the next stimulus was placed successively in the same sequence of seven 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 (see Figs. 7 and 8) (Rolls & Deco, 2002; Wallis & 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 asymmetric rule (the presynaptic term from the current exemplar, and the trace 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

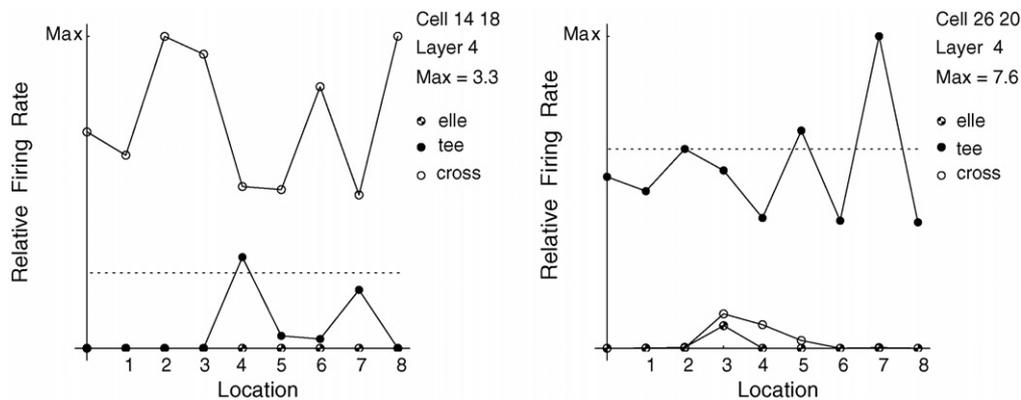


Fig. 8. Response profiles for two fourth layer neurons in VisNet – discrimination factors 4.07 and 3.62 – in the L, T and + invariance learning experiment. (Wallis & Rolls, 1997.)

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 (see Sutton & 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 & 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 generalise to other views of the face. This is a fundamental aspect of the way in which the brain is organised in order to allow this type of capability for emotional and social behaviour (Rolls, 1999b, 2005).

### 13. Different neural systems are specialised 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, Rolls, & Baylis, 1989). The neurons responsive to expression were found primarily in the cortex in the superior temporal sulcus, while 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, 1986b, 1990, 1999b, 2005). Damage to this population may contribute to the deficits in social and emotional behaviour which are part of the Kluver–Bucy syndrome produced by temporal lobe damage in monkeys (see Leonard et al., 1985; Rolls, 1981, 1984, 1986a, 1986b, 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 inter-

actions is that some of them respond to gestures, e.g. to a face undergoing ventral flexion, as described above and by Perrett, Smith, Mistlin et al. (1985). 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, Rolls, & Baylis, 1989). 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, Rolls, Baylis, & Nalwa, 1989) and eye gaze (Perrett, Smith, Mistlin et al., 1985) 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, Rolls, & Baylis, 1989) and with view-independent responses (Hasselmo, Rolls, Baylis, & Nalwa, 1989) 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, in order to make the correct social or emotional response.

### 14. 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 & 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 gesture (Brothers, Ring, & Kling, 1990). The amygdala representation of faces is discussed in Section 15.

## 15. A representation of faces in the orbitofrontal cortex

Rolls et al. (2006) have found a number of face-responsive neurons in the orbitofrontal cortex, and they are also present in adjacent prefrontal cortical areas (Wilson, O'Scalaidhe, & Goldman-Rakic, 1993). The orbitofrontal cortex face-responsive neurons, first observed by Thorpe, Rolls, and Maddison (1983), then by Rolls et al. (2006), 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 (see Fig. 9); 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 two-dimensional images of faces on a video monitor (Rolls & 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 (cf. Hasselmo, Rolls, & Baylis, 1989), which can indicate reinforcement, and by encoding information about which individual is present, also important in evaluating and utilising reinforcing inputs in social situations.

We have also been able to obtain evidence that non-reward used as a signal to reverse behavioural 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.

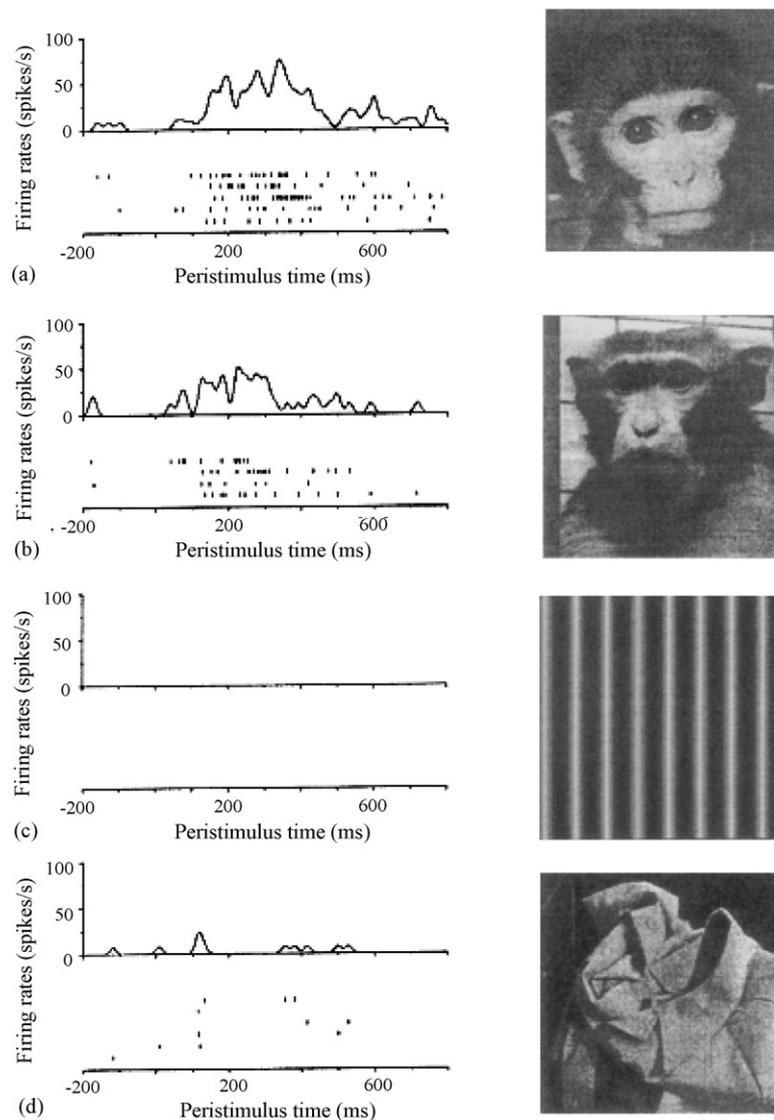


Fig. 9. Orbitofrontal cortex faceselective neuron as found in macaques. Peristimulus rastergrams and time histograms are shown. Each trial is a row in the rastergram. Several trials for each stimulus are shown. The ordinate is in spikes/s. The neuron responded best to face (a), also responded, though 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.)

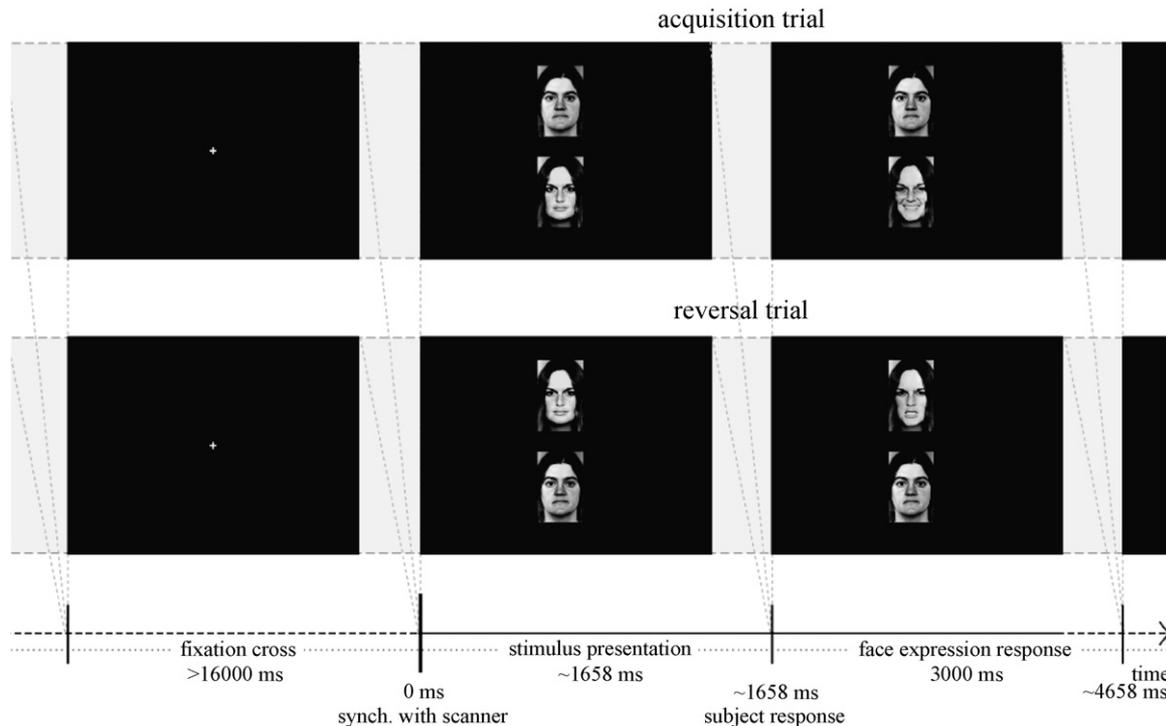


Fig. 10. Social reversal task: the trial starts synchronised with the scanner and two people with neutral face expressions are presented to the subject. The subject has to select one of the people by pressing the corresponding button, and the person will then either smile or show an angry face expression 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 four and eight 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 16,000 ms. (After Kringelbach & Rolls, 2003.)

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 when the human chose a face, and did not obtain the expected reward (see Figs. 10 and 11). 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 behaviour 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 behaviour (Kringelbach & 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 and vocal emotional expression were demonstrated in a group of patients with ventral frontal lobe damage who had socially inappropriate behaviour (Hornak, Rolls, & Wade, 1996; Rolls, 1999a). The expression identification impair-

ments 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 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 behavioural 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 behaviour, 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 specialisation 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 pro-

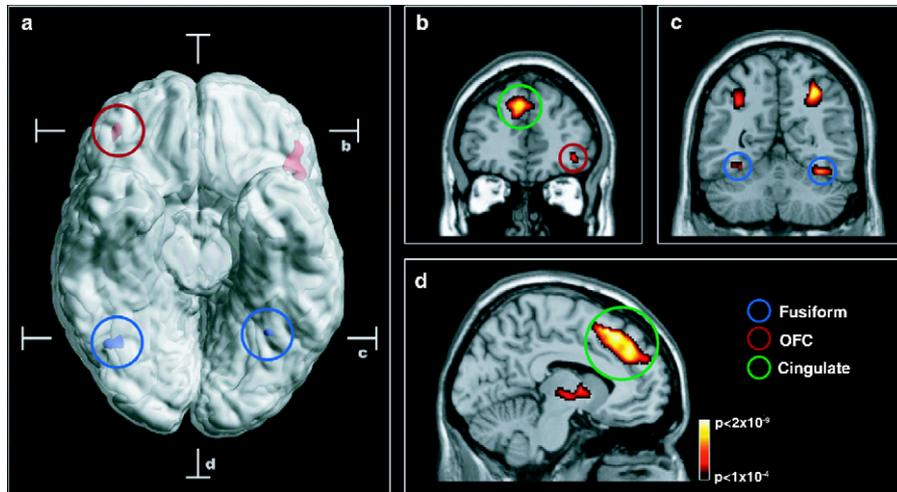


Fig. 11. Social reversal: composite figure showing that changing behaviour based on face expression is correlated with increased brain activity in the human orbitofrontal cortex. (a) The figure is based on two different group statistical contrasts from the neuroimaging data which are superimposed on a ventral view of the human brain with the cerebellum removed, and with indication of the location of the two coronal slices (b and 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 & Rolls, 2003.)

duce 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 behaviour, 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 behaviour; 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 behaviour, or in their subjective emotional state. Patients with unilateral lesions of the antero-ventral 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 behaviour, 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 pri-

mary 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 recognise any emotions in others, in that frequently voice but not face expression identification was impaired, and vice versa. This implies some functional specialisation for visual versus auditory emotion-related processing in the human orbitofrontal cortex. The results also show that the changes in social behaviour 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, 2003; Rolls, 1999a) is more extensive than the impairment in identifying primarily fear face expressions produced by amygdala damage in humans (Adolphs, Baron-Cohen, & Tranel, 2002; Calder et al., 1996) (for review, see Rolls, 2005). In addition, the deficits in emotional and social behaviour described above that are produced by orbitofrontal cortex damage in humans seem to be more pronounced than changes in emotional behaviour produced by amygdala damage in humans, although deficits in autonomic conditioning can be demonstrated (Phelps, 2004). This 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).

## 16. 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 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 which 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., 2006), and are important in human social and emotional behaviour, which are 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 behaviour, and how these depend on inputs from

the temporal cortex visual areas, is provided elsewhere (Rolls, 2005).

## Acknowledgements

The author has worked on some of the investigations described here with P. Azzopardi, G.C. Baylis, P. Földiák, M. Hasselmo, C.M. Leonard, G. Littlewort, T.J. Milward, D.I. Perrett, M.J. Tovee, 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.

## References

- Abbott, L. F., Rolls, E. T., & Tovee, M. J. (1996). Representational capacity of face coding in monkeys. *Cerebral Cortex*, *6*, 498–505.
- Adolphs, R., Baron-Cohen, S., & Tranel, D. (2002). Impaired recognition of social emotions following amygdala damage. *Journal of Cognitive Neuroscience*, *14*, 1264–1274.
- Aggelopoulos, N. C., Franco, L., & Rolls, E. T. (2005). Object perception in natural scenes: Encoding by inferior temporal cortex simultaneously recorded neurons. *Journal of Neurophysiology*, *93*, 1342–1357.
- Baddeley, R. J., Abbott, L. F., Booth, M. J. A., Sengpiel, F., Freeman, T., Wake-man, E. A., et al. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London B*, *264*, 1775–1783.
- Baizer, J. S., Ungerleider, L. G., & Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *Journal of Neuroscience*, *11*, 168–190.
- Ballard, D. H. (1990). Animate vision uses object-centred reference frames. In R. Eckmiller (Ed.), *Advanced neural computers* (pp. 229–236). Amsterdam: North-Holland.
- Ballard, D. H. (1993). Subsymbolic modelling of hand-eye coordination. In D. E. Broadbent (Ed.), *The simulation of human intelligence* (pp. 71–102). Oxford: Blackwell.
- Barlow, H. B. (1972). Single units and sensation: A neuron doctrine for perceptual psychology? *Perception*, *1*, 371–394.
- Baylis, G. C., & Rolls, E. T. (1987). Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. *Experimental Brain Research*, *65*, 614–622.
- Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Research*, *342*, 91–102.
- Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1987). Functional subdivisions of the temporal lobe neocortex. *Journal of Neuroscience*, *7*, 330–342.
- Booth, M. C. A., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cerebral Cortex*, *8*, 510–523.
- Boussaoud, D., Desimone, R., & Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque. *Journal of Comparative Neurology*, *306*, 554–575.
- Brothers, L., Ring, B., & Kling, A. (1990). Response of neurons in the macaque amygdala to complex social stimuli. *Behavioural Brain Research*, *41*, 199–213.
- Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*, 369–384.
- Calder, A. J., Young, A. W., Rowland, D., Perrett, D. I., Hodges, J. R., & Etcoff, N. L. (1996). Facial emotion recognition after bilateral amygdala damage:

- Differentially severe impairment of fear. *Cognitive Neuropsychology*, *13*, 699–745.
- Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. *Journal of Cognitive Neuroscience*, *3*, 1–8.
- Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*, 2051–2062.
- Desimone, R., & Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research*, *178*, 363–380.
- Dolan, R. J., Fink, G. R., Rolls, E. T., Booth, M., Holmes, A., Frackowiak, R. S. J., et al. (1997). How the brain learns to see objects and faces in an impoverished context. *Nature*, *389*, 596–599.
- Elliffe, M. C. M., Rolls, E. T., & Stringer, S. M. (2002). Invariant recognition of feature combinations in the visual system. *Biological Cybernetics*, *86*, 59–71.
- Engel, A. K., Konig, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual system: New vistas on integration in the nervous system. *Trends in Neurosciences*, *15*, 218–226.
- Földiák, P. (1991). Learning invariance from transformation sequences. *Neural Computation*, *3*, 194–200.
- Franco, L., Rolls, E. T., Aggelopoulos, N. C., & Jerez, J. M. (2006). Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex.
- Franco, L., Rolls, E. T., Aggelopoulos, N. C., & Treves, A. (2004). The use of decoding to analyze the contribution to the information of the correlations between the firing of simultaneously recorded neurons. *Experimental Brain Research*, *155*, 370–384.
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*, 193–202.
- Fukushima, K. (1989). Analysis 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 Transactions*, *74*, 179–190.
- Gallant, J. L., Connor, C. E., & Van Essen, D. C. (1998). Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport*, *9*, 85–89.
- Gawne, T. J., & Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? *Journal of Neuroscience*, *13*, 2758–2771.
- Georges-François, P., Rolls, E. T., & Robertson, R. G. (1999). Spatial view cells in the primate hippocampus: Allocentric view not head direction or eye position or place. *Cerebral Cortex*, *9*, 197–212.
- Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*, 649–677.
- Gross, C. G., Desimone, R., Albright, T. D., & Schwartz, E. L. (1985). Inferior temporal cortex and pattern recognition. *Experimental Brain Research*, *11*, 179–201.
- Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, *32*, 203–218.
- Hasselmo, M. E., Rolls, E. T., Baylis, G. C., & Nalwa, V. (1989). Object-centred encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research*, *75*, 417–429.
- Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2002). Human neural systems for face recognition and social communication. *Biological Psychiatry*, *51*, 59–67.
- Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R., et al. (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, E. T., Morris, R. G., Bullock, P. R., et al. (2004). Reward-related reversal learning after surgical excisions in orbitofrontal and dorsolateral prefrontal cortex in humans. *Journal of Cognitive Neuroscience*, *16*, 463–478.
- Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia*, *34*, 247–261.
- Koenderink, J. J., & Van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics*, *32*, 211–217.
- Kringelbach, M. L., & Rolls, E. T. (2003). Neural correlates of rapid reversal learning in a simple model of human social interaction. *NeuroImage*, *20*, 1371–1383.
- Kringelbach, M. L., & Rolls, E. T. (2004). The functional neuroanatomy of the human orbitofrontal cortex: Evidence from neuroimaging and neuropsychology. *Progress in Neurobiology*, *72*, 341–372.
- Leonard, C. M., Rolls, E. T., Wilson, F. A. W., & Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behavioural Brain Research*, *15*, 159–176.
- Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current Biology*, *4*, 401–414.
- Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, *19*, 577–621.
- Marr, D. (1982). *Vision*. San Francisco: WH Freeman.
- Maunsell, J. H., & Newsome, W. T. (1987). Visual processing in monkey extrastriate cortex. *Annual Reviews in Neuroscience*, *10*, 363–401.
- Mikami, A., Nakamura, K., & Kubota, K. (1994). Neuronal responses to photographs in the superior temporal sulcus of the rhesus monkey. *Behavioural Brain Research*, *60*, 1–13.
- Miller, E. K., & 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. *Annual Review of Neuroscience*, *16*, 245–263.
- O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, *17*, 580–590.
- Panzeri, S., Biella, G., Rolls, E. T., Skaggs, W. E., & Treves, A. (1996). Speed, noise, information and the graded nature of neuronal responses. *Network*, *7*, 365–370.
- Panzeri, S., Rolls, E. T., Battaglia, F., & Lavis, R. (2001). Speed of feed-forward and recurrent processing in multilayer networks of integrate-and-fire neurons. *Network: Computation in Neural Systems*, *12*, 423–440.
- Panzeri, S., Schultz, S. R., Treves, A., & Rolls, E. T. (1999). Correlations and the encoding of information in the nervous system. *Proceedings of the Royal Society of London B*, *266*, 1001–1012.
- Panzeri, S., Treves, A., Schultz, S., & Rolls, E. T. (1999a). On decoding the responses of a population of neurons from short time epochs. *Neural Computation*, *11*, 1553–1577.
- Perrett, D., Mistlin, A., & Chitty, A. (1987). Visual neurons responsive to faces. *Trends in Neurosciences*, *10*, 358–364.
- Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*, 329–342.
- Perrett, D. I., Smith, P. A. J., Mistlin, A. J., Chitty, A. J., Head, A. S., Potter, D. D., et al. (1985). Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: A preliminary report. *Behavioural Brain Research*, *16*, 153–170.
- Perrett, D. I., Smith, P. A., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London B*, *223*, 293–317.
- Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, *14*, 198–202.
- Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, *343*, 263–266.
- Renart, A., Parga, N., & Rolls, E. T. (2000). A recurrent model of the interaction between the prefrontal cortex and inferior temporal cortex in delay memory tasks. In S. A. Solla, T. K. Leen, & K.-R. Mueller (Eds.), *Advances in neural information processing systems*: Vol. 12, (Vol. 12, pp. 171–177). Cambridge, MA: MIT Press.
- Robertson, R. G., Rolls, E. T., & Georges-François, P. (1998). Spatial view cells in the primate hippocampus: Effects of removal of view details. *Journal of Neurophysiology*, *79*, 1145–1156.
- Rolls, E. T. (1981). Responses of amygdaloid neurons in the primate. In Y. Ben-Ari (Ed.), *The amygdaloid complex* (pp. 383–393). Amsterdam: Elsevier.

- Rolls, E. T. (1984). Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Human Neurobiology*, 3, 209–222.
- Rolls, E. T. (1986a). A theory of emotion, and its application to understanding the neural basis of emotion. In Y. Oomura (Ed.), *Emotions. Neural and chemical control* (pp. 325–344). Basel: Karger.
- Rolls, E. T. (1986b). Neural systems involved in emotion in primates. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience. Biological foundations of emotion: Vol. 3*, (Vol. 3, (pp. 125–143). New York: Academic Press.
- Rolls, E. T. (1989a). Functions of neuronal networks in the hippocampus and neocortex in memory. In J. H. Byrne & W. O. Berry (Eds.), *Neural models of plasticity: Experimental and theoretical approaches* (pp. 240–265). San Diego: Academic Press.
- Rolls, E. T. (1989b). The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus. In R. Durbin, C. Miall, & G. Mitchison (Eds.), *The computing neuron* (pp. 125–159). Wokingham, England: Addison-Wesley.
- Rolls, E. T. (1990). A theory of emotion, and its application to understanding the neural basis of emotion. *Cognition and Emotion*, 4, 161–190.
- Rolls, E. T. (1991). Neural organisation of higher visual functions. *Current Opinion in Neurobiology*, 1, 274–278.
- Rolls, E. T. (1992a). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society of London B*, 335, 11–21.
- Rolls, E. T. (1992b). Neurophysiology and functions of the primate amygdala. In J. P. Aggleton (Ed.), *The amygdala* (pp. 143–165). New York: Wiley-Liss.
- Rolls, E. T. (1997). A neurophysiological and computational approach to the functions of the temporal lobe cortical visual areas in invariant object recognition. In M. Jenkin & L. Harris (Eds.), *Computational and psychophysical mechanisms of visual coding* (pp. 184–220). Cambridge University Press: Cambridge.
- Rolls, E. T. (1999a). The functions of the orbitofrontal cortex. *Neurocase*, 5, 301–312.
- Rolls, E. T. (1999b). *The brain and emotion*. Oxford: Oxford University Press.
- Rolls, E. T. (1999c). Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus*, 9, 467–480.
- Rolls, E. T. (2000a). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron*, 27, 205–218.
- Rolls, E. T. (2000b). Neurophysiology and functions of the primate amygdala, and the neural basis of emotion. In J. P. Aggleton (Ed.), *The amygdala: A functional analysis* (2nd ed., pp. 447–478). Oxford: Oxford University Press.
- Rolls, E. T. (2003). Consciousness absent and present: A neurophysiological exploration. *Progress in Brain Research*, 144, 95–106.
- Rolls, E. T. (2005). *Emotion explained*. Oxford: Oxford University Press.
- Rolls, E. T., Aggelopoulos, N. C., Franco, L., & Treves, A. (2004). Information encoding in the inferior temporal cortex: Contributions of the firing rates and correlations between the firing of neurons. *Biological Cybernetics*, 90, 19–32.
- Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *Journal of Neuroscience*, 23, 339–348.
- Rolls, E. T., & Baylis, G. C. (1986). Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. *Experimental Brain Research*, 65, 38–48.
- Rolls, E. T., Baylis, G. C., & Hasselmo, M. E. (1987). The responses of neurons in the cortex in the superior temporal sulcus of the monkey to band-pass spatial frequency filtered faces. *Vision Research*, 27, 311–326.
- Rolls, E. T., Baylis, G. C., Hasselmo, M., & Nalwa, V. (1989). The representation of information in the temporal lobe visual cortical areas of macaque monkeys. In J. J. Kulikowski, C. M. Dickinson, & I. J. Murray (Eds.), *Seeing contour and colour*. Oxford: Pergamon.
- Rolls, E. T., Baylis, G. C., Hasselmo, M. E., & Nalwa, V. (1989). The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Experimental Brain Research*, 76, 153–164.
- Rolls, E. T., Baylis, G. C., & Leonard, C. M. (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 Research*, 25, 1021–1035.
- Rolls, E. T., Critchley, H. D., Browning, A. S., & Inoue, K. (2006). Face-selective and auditory neurons in the primate orbitofrontal cortex. *Experimental Brain Research*, 170, 74–87.
- Rolls, E. T., Critchley, H. D., & Treves, A. (1996). The representation of olfactory information in the primate orbitofrontal cortex. *Journal of Neurophysiology*, 75, 1982–1996.
- Rolls, E. T., & Deco, G. (2002). *Computational neuroscience of vision*. Oxford: Oxford University Press.
- Rolls, E. T., Franco, L., Aggelopoulos, N. C., & Reece, S. (2003). An information theoretic approach to the contributions of the firing rates and correlations between the firing of neurons. *Journal of Neurophysiology*, 89, 2810–2822.
- Rolls, E. T., & Milward, T. (2000). A model of invariant object recognition in the visual system: Learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation*, 12, 2547–2572.
- Rolls, E. T., Robertson, R. G., & Georges-François, P. (1997). Spatial view cells in the primate hippocampus. *European Journal of Neuroscience*, 9, 1789–1794.
- Rolls, E. T., & Stringer, S. M. (2001). Invariant object recognition in the visual system with error correction and temporal difference learning. *Network: Computation in Neural Systems*, 12, 111–129.
- Rolls, E. T., & Tovee, M. J. (1994). Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proceedings of the Royal Society of London B*, 257, 9–15.
- Rolls, E. T., & Tovee, M. J. (1995a). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology*, 73, 713–726.
- Rolls, E. T., & Tovee, M. J. (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. *Experimental Brain Research*, 103, 409–420.
- Rolls, E. T., Tovee, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, 11, 335–346.
- Rolls, E. T., Tovee, M. J., Purcell, D. G., Stewart, A. L., & Azzopardi, P. (1994). The responses of neurons in the temporal cortex of primates, and face identification and detection. *Experimental Brain Research*, 101, 473–484.
- Rolls, E. T., & Treves, A. (1990). The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network*, 1, 407–421.
- Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function*. Oxford: Oxford University Press.
- Rolls, E. T., Treves, A., Robertson, R. G., Georges-François, P., & Panzeri, S. (1998). Information about spatial view in an ensemble of primate hippocampal cells. *Journal of Neurophysiology*, 79, 1797–1813.
- Rolls, E. T., Treves, A., & Tovee, M. J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. *Experimental Brain Research*, 114, 177–185.
- Rolls, E. T., & Xiang, J.-Z. (2005). Reward-spatial view representations and learning in the hippocampus. *Journal of Neuroscience*, 25, 6167–6174.
- Rolls, E. T., Xiang, J.-Z., & Franco, L. (2005). Object, space and object-space representations in the primate hippocampus. *Journal of Neurophysiology*, 94, 833–844.
- Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Experimental Brain Research*, 77, 23–30.
- Seltzer, B., & Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Research*, 149, 1–24.
- Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex. *Human Brain Mapping*, 27, 77–89.

- Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron*, *35*, 1157–1165.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. Cambridge, MA: MIT Press.
- Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, *262*, 685–688.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *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 E. Iwai & M. Mishkin (Eds.), *Vision, memory and the temporal lobe* (pp. 101–109). New York: Elsevier.
- Thorpe, S. J., & Imbert, M. (1989). Biological constraints on connectionist models. In R. Pfeifer, Z. Schreter, & F. Fogelman-Soulie (Eds.), *Connectionism in perspective* (pp. 63–92). Amsterdam: Elsevier.
- Thorpe, S. J., Rolls, E. T., & Maddison, S. (1983). Neuronal activity in the orbitofrontal cortex of the behaving monkey. *Experimental Brain Research*, *49*, 93–115.
- Tovee, M. J., & Rolls, E. T. (1995). Information encoding in short firing rate epochs by single neurons in the primate temporal visual cortex. *Visual Cognition*, *2*, 35–58.
- Tovee, M. J., Rolls, E. T., & Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *Journal of Neurophysiology*, *72*, 1049–1060.
- Tovee, M. J., Rolls, E. T., & Ramachandran, V. S. (1996). Rapid visual learning in neurones of the primate temporal visual cortex. *Neuroreport*, *7*, 2757–2760.
- Tovee, M. J., Rolls, E. T., Treves, A., & Bellis, R. P. (1993). Information encoding and the responses of single neurons in the primate temporal visual cortex. *Journal of Neurophysiology*, *70*, 640–654.
- Treves, A. (1993). Mean-field analysis of neuronal spike dynamics. *Network*, *4*, 259–284.
- Treves, A., Panzeri, S., Rolls, E. T., Booth, M., & Waksman, E. A. (1999). Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. *Neural Computation*, *11*, 611–641.
- Treves, A., & Rolls, E. T. (1991). What determines the capacity of autoassociative memories in the brain? *Network*, *2*, 371–397.
- Treves, A., Rolls, E. T., & Simmen, M. (1997). Time for retrieval in recurrent associative memories. *Physica D*, *107*, 392–400.
- Treves, A., Rolls, E. T., & Tovee, M. J. (1996). On the time required for recurrent processing in the brain. In V. Torre & F. Conti (Eds.), *Neurobiology: Ionic channels, neurons, and the brain* (pp. 325–353). New York: Plenum.
- Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: Bradford/MIT Press.
- Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, *51*, 167–194.
- Wallis, G., Rolls, E. T., & Földiák, P. (1993). Learning invariant responses to the natural transformations of objects. *International Joint Conference on Neural Networks*, *2*, 1087–1090.
- Williams, G. V., Rolls, E. T., Leonard, C. M., & Stern, C. (1993). Neuronal responses in the ventral striatum of the behaving macaque. *Behavioural Brain Research*, *55*, 243–252.
- Wilson, F. A. W., O'Scalaidhe, S. P. O., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, *260*, 1955–1958.
- Xiang, J.-Z., & Brown, M. W. (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? *Experimental Brain Research*, *73*, 209–214.