

# Computational Neuroscience<sup>☆</sup>

ET Rolls, Oxford Centre for Computational Neuroscience, Oxford, England

© 2017 Elsevier Inc. All rights reserved.

---

<b>Introduction to Computational Neuroscience</b>	1
<b>Biologically Plausible Neuronal Networks</b>	2
Summary of the Architecture and Operation of Autoassociation or Attractor Networks	2
Summary of the Architecture and Operation of Pattern Association Networks	3
Summary of the Architecture and Operation of Competitive Networks	4
<b>Coding</b>	4
Definitions of Types of Representation	5
Sparse Distributed Graded Firing Rate Encoding of Face and Object Identity in the Inferior Temporal Visual Cortex; of Taste and Related Stimuli in the Insula and Orbitofrontal Cortex; and of Space in the Hippocampus	5
Sparseness of the Representation and an Approximately Exponential Firing Rate Probability Distribution	5
Comparisons of Sparseness Between Areas: The Inferior Temporal Visual Cortex, Hippocampus, Insula, Orbitofrontal Cortex, and Amygdala	8
Single Cell Information	9
The Information Available From Multiple Cells	10
Advantages of the Sparse Distributed Graded Representation of Objects for Brain Processing	12
Evidence From Humans	14
<b>Acknowledgments</b>	15
<b>References</b>	15
<b>Relevant Website</b>	16

---

## Introduction to Computational Neuroscience

To understand how the brain works, including how it functions in perception, memory, attention, decision-making, and cognitive functions, it is necessary to combine different approaches, including neural computation (Rolls, 2016b).

Neurophysiology at the single neuron level is needed because this is the level at which information is exchanged between the computing elements of the brain. Evidence from the effects of brain damage, including that available from neuropsychology, is needed to help understand what different parts of the system do, and indeed what each part is necessary for. Neuroimaging is useful to indicate where in the human brain different processes take place and to show which functions can be dissociated from each other. Knowledge of the biophysical and synaptic properties of neurons is essential to understand how the computing elements of the brain work, and therefore what the building blocks of biologically realistic computational models should be. Knowledge of the anatomical and functional architecture of the cortex is needed to show what types of neuronal network actually perform the computation. And finally the approach of neural computation is needed, as this is required to link together all the empirical evidence to produce an understanding of how the system actually works. Computational neuroscience utilizes evidence from all these disciplines to develop an understanding of how different types of function, including memory, perception, attention, and decision-making are implemented by processing in the cerebral cortex.

I emphasize that to understand how the brain works, we are dealing with large-scale computational systems with interactions between the parts, and that this understanding requires analysis at the computational and global level of the operation of many neurons to perform together a useful function. Understanding at the molecular level is important for helping to understand how these large-scale computational processes are implemented in the brain, but will not by itself give any account of what computations are performed to implement these cognitive functions. Instead, understanding cognitive functions such as object recognition, memory recall, attention, and decision-making requires single neuron data to be closely linked to computational models of how the interactions between large numbers of neurons and many networks of neurons allow these cognitive problems to be solved. The single neuron level is important in this approach, for the single neurons can be thought of as the computational units of the system, and is the level at which the information is exchanged by the spiking activity between the computational elements of the brain. The single neuron level is therefore, because it is the level at which information is communicated between the computing elements of the brain, the fundamental level of information processing and the level at which the information can be read out (by recording the spiking activity) to understand what information is being represented and processed in each brain area.

A test of whether one's understanding is correct is to simulate the processing on a computer, and to show whether the simulation can perform the tasks of cortical systems, and whether the simulation has similar properties to the real brain. The approach of neural computation leads to a precise definition of how the computation is performed and to precise and quantitative tests of the theories

---

<sup>☆</sup>*Change History:* This is a new article in June 2016.

produced. How memory systems in the cortex work is a paradigm example of this approach, because memory-like operations which involve altered functionality as a result of synaptic modification are at the heart of how many computations in the cerebral cortex are performed. It happens that attention and decision-making can be understood in terms of interactions between and fundamental operations in memory systems in the cortex, and therefore it is natural to treat these areas of cognitive neuroscience together, for the same fundamental concepts based on the operation of neuronal circuitry can be applied to all these functions, and to many more (Rolls, 2008, 2016b).

Useful introductions to computational neuroscience are provided by Dayan and Abbott (2001) in *Theoretical Neuroscience* and by Rolls (2016b) in *Cerebral Cortex: Principles of Operation*. *Cerebral Cortex: Principles of Operation* is notable in that its Appendix B, on the operation of biologically plausible neuronal networks, is available at (see [Relevant Website](#)), together with programs written in Matlab (which will also operate in the Octave freeware) to demonstrate the operation of biologically plausible pattern association, autoassociation, and competitive networks, and self-organizing maps. Another book that provides a clear and quantitative introduction to some of these networks is *Introduction to the Theory of Neural Computation* (Hertz et al., 1991). Gerstner et al. (2014) focus on neuronal dynamics, Amit (1989) provides a thorough analysis of attractor (autoassociation) networks, and Koch (1999) provides a biophysical approach to neuronal operation.

This article which provides an introduction to computational neuroscience includes next a brief introduction to some of the main types of biologically plausible neuronal networks and then provides an overview of neuronal encoding, for this is a part of computational neuroscience that is fundamental to understanding the operation of brain systems.

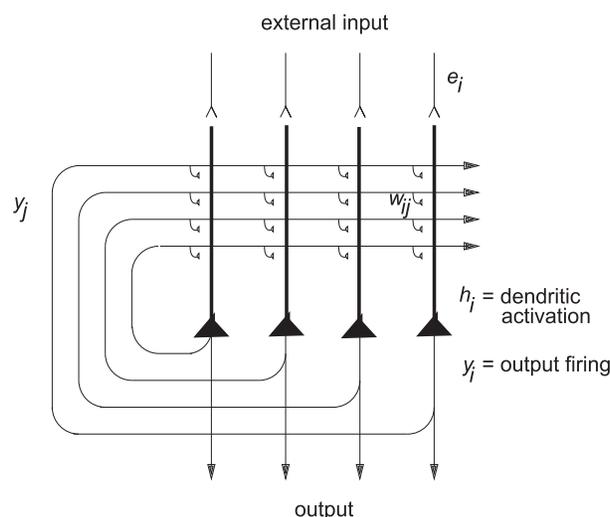
## Biologically Plausible Neuronal Networks

A fuller description of the operation and utility of these networks is available in Appendix B of *Cerebral Cortex: Principles of Operation* (Rolls, 2016b) (see [Relevant Website](#)).

### Summary of the Architecture and Operation of Autoassociation or Attractor Networks

Autoassociation or attractor networks have recurrent collateral connections between the neurons and provide a basis for continuing firing to implement short-term memory, for the completion of a whole memory from a part to implement long-term memory and for decision-making when two inputs to an attractor network compete with each other (Rolls, 2016b).

The prototypical architecture of an autoassociation memory is shown in Fig. 1. The external input  $e_i$  is applied to each neuron  $i$  by unmodifiable synapses. This produces firing  $y_i$  of each neuron. Each neuron is connected by a recurrent collateral synaptic connection to the other neurons in the network, via associatively modifiable connection weights  $w_{ij}$ . This architecture effectively enables the vector of output firing rates to be associated during learning with itself. Later on, during recall, presentation of part of the external input will force some of the output neurons to fire, but through the recurrent collateral axons and the modified synapses, other neurons can be brought into activity. This process can be repeated a number of times, and recall of a complete pattern may be perfect. Effectively, a pattern can be recalled or recognized because of associations formed between its parts. This, of course, requires distributed representations.



**Figure 1** The architecture of an autoassociative neural network. The recurrent collateral synaptic weights are excitatory. Inhibitory neurons (not illustrated) must be present to control the firing rates. From Rolls, E.T., 2016. *Cerebral Cortex: Principles of Operation*. Oxford University Press, Oxford.

Training for each desired pattern occurs in a single trial. The firing of every output neuron  $i$  is forced to a value  $y_i$  determined by the external input  $e_i$ . Then a Hebb-like associative local learning rule is applied to the recurrent synapses in the network:

$$\delta w_{ij} = k \cdot \gamma_i \cdot \gamma_j \tag{1}$$

where  $k$  is a learning rate constant,  $\gamma_i$  is the activation of the dendrite (the postsynaptic term),  $\gamma_j$  is the presynaptic firing rate, and  $\delta w_{ij}$  is the change in the synaptic weight from axon  $j$  to neuron  $i$ . This learning algorithm is fast, "one-shot," in that a single presentation of an input pattern is all that is needed to store that pattern.

It is notable that in a fully connected network, this will result in a symmetrical matrix of synaptic weights, that is, the strength of the connection from neuron 1 to neuron 2 will be the same as the strength of the connection from neuron 2 to neuron 1 (both implemented via recurrent collateral synapses).

During recall, an external input  $e$  is applied and produces output firing, operating through a nonlinear activation function. The firing is fed back by the recurrent collateral axons shown in Fig. 1 to produce activation of each output neuron through the modified synapses on each output neuron. The activation  $h_i$  produced by the recurrent collateral effect on the  $i$ th neuron is, in the standard way, the sum of the activations produced in proportion to the firing rate of each axon operating through each modified synapses  $w_{ij}$ , that is,

$$h_i = \sum_j \gamma_j w_{ij} \tag{2}$$

where the sum is over the input axons to each neuron, indexed by  $j$ .

The output firing  $y_i$  is a function of the activation produced by the recurrent collateral effect (internal recall) and by the external input  $e_i$ :

$$y_i = f(h_i + e_i). \tag{3}$$

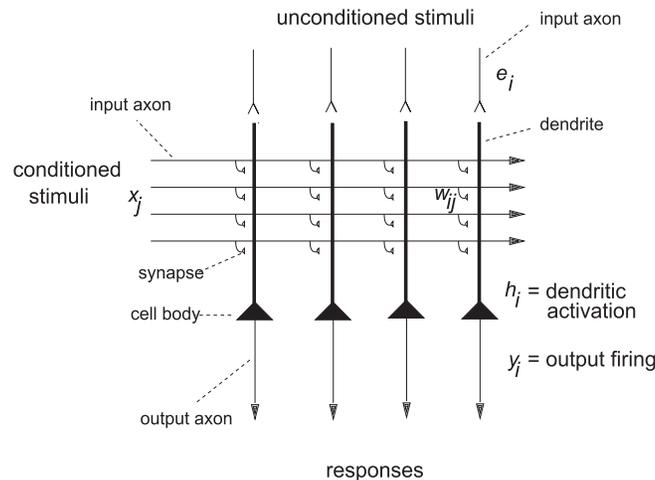
The activation function  $f$  should be nonlinear and may be, for example, binary threshold, linear threshold, sigmoid, and so forth. A purely linear system would not produce any categorization of the input patterns it receives and therefore would not be able to effect anything more than a trivial (ie, linear) form of completion and generalization.

During recall, a part of one of the originally learned stimuli can be presented as an external input. The resulting firing is allowed to iterate repeatedly round the recurrent collateral system, gradually on each iteration recalling more and more of the originally learned pattern. Completion thus occurs. If a pattern is presented during recall that is similar to one of the previously learned patterns, then the network settles into a stable recall state in which the firing corresponds to that of the previously learned pattern. The network can thus generalize in its recall to the most similar previously learned pattern. Important results characterize how many patterns can be stored in a network in this way without interference during recall (Rolls, 2016b).

**Summary of the Architecture and Operation of Pattern Association Networks**

Fig. 2 shows a pattern association network. The type of network is important in classical conditioning, in stimulus–reinforcer association learning, and in the recall of memories using backprojections (Rolls, 2016b).

We trace the emergence of a conditioned response (CR). An unconditioned stimulus (US) has activity or firing rate  $e_i$  for the  $i$ th neuron and produces firing  $y_i$  of the  $i$ th neuron which is an unconditioned response (UR). The conditioned stimuli (CS) have



**Figure 2** The architecture of a pattern association network. From Rolls, E.T., 2016. Cerebral Cortex: Principles of Operation. Oxford University Press, Oxford.

activity or firing rate  $x_j$  for the  $j$ th axon. During learning, a CS is presented at the same time as a US, and the synaptic weights are modified by an associative synaptic learning rule  $\delta w_{ij} = k \cdot y_i \cdot x_j$  where  $k$  is a learning rate constant, and  $\delta w_{ij}$  is the change in the synaptic weight from axon  $j$  to neuron  $i$ . During recall, only the CS is presented, and the activation  $h_i$  is calculated as a dot product between the input stimulus and the synaptic weight vector on a neuron  $h_i = \sum_j x_j w_{ij}$ , where the sum is over the input axons to each

neuron, indexed by  $j$ . The output firing  $y_i$  is a function  $f$  of the activation  $y_i = f(h_i)$ . The activation function  $f$  should be nonlinear and may be, for example, binary threshold, linear threshold, or sigmoid. Inhibitory neurons not shown in the figure are part of the way in which the threshold of the activation function is set. The nonlinearity in the activation function enables interference from other pattern pairs stored in the network to be minimized. The pattern association network thus enables a CS to retrieve a response, the CR, which was present as a UR during the learning. An important property is that if a distorted CS is presented, generalization occurs to the effects produced by the closest CS during training (Rolls, 2016b).

### Summary of the Architecture and Operation of Competitive Networks

Competitive networks perform categorization of input stimuli. Similar input stimuli are categorized together. More different input stimuli are categorized as different. Competitive networks underlie much new learning performed by the cerebral cortex (Rolls, 2016b).

During training of a competitive network (Fig. 3), an input stimulus is presented to the synaptic matrix which has random initial weights  $w_{ij}$  and produces activation  $h_i$  of the  $i$ th neuron calculated as a dot product between the input stimulus and the synaptic weight vector on a neuron  $h_i = \sum_j x_j w_{ij}$ , where the sum is over the input axons to each neuron, indexed by  $j$ . The output firing  $y_i$

is a function  $f$  of the activation  $y_i = f(h_i)$ . The activation function  $f$  should be nonlinear and may be, for example, binary threshold, linear threshold, or sigmoid. Inhibitory neurons not shown in the figure are part of the way in which the threshold of the activation function is set in which the threshold reflects the firing of all the output neurons. This or other competitive mechanisms result in a typically sparse set of output neurons having firing after the competition.

Next, an associative synaptic modification rule is applied, while the presynaptic input and the postsynaptic output are both present,  $\delta w_{ij} = k \cdot y_i \cdot x_j$  where  $k$  is a learning rate constant, and  $\delta w_{ij}$  is the change in the synaptic weight from axon  $j$  to neuron  $i$ .

Next, the synaptic weight vector on each neuron is normalized to ensure that no one neuron dominates the classification.

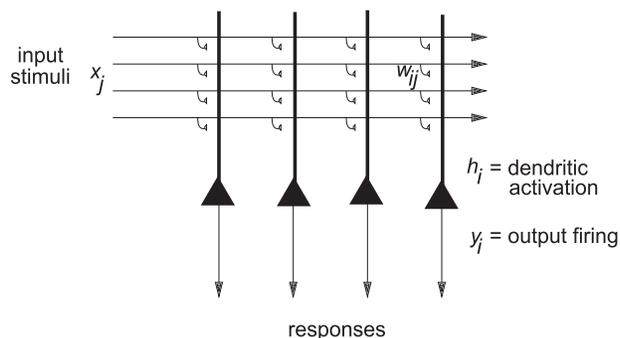
This process is repeated for every input pattern in random permuted sequence, and this process is repeated for a small number of training epochs.

After the training, each dendritic weight vector points toward a cluster of patterns in the input space. The competitive network has used self-organization with no teacher to categorize the inputs, with patterns close in the input space activating the same set of output neurons, and different clusters of inputs activating different sets of output neurons.

### Coding

I next describe the cortical neuronal encoding of information in primates including humans. The analysis is based on data sets of neurons in many different cortical areas of primates during the normal operation of the brain during behavior, allowing a unique comparison of neuronal encoding in the inferior temporal visual cortex, the primary taste cortex in the insula, the orbitofrontal cortex, and the hippocampus and parahippocampal gyrus. The coding was analyzed with rigorous information theoretic methods that we developed for application to single and multiple single neuron data and which are described by Rolls and Treves (2011), Rolls (2016b), and in the original papers (Rolls et al., 1997a,b, 2003b, 2004, 2006; Franco et al., 2004).

I start with some definitions, then summarize some evidence that shows the type of encoding used in some cortical regions, and then show how the representation found is advantageous.



**Figure 3** The architecture of a competitive network. From Rolls, E.T., 2016. Cerebral Cortex: Principles of Operation. Oxford University Press, Oxford.

## Definitions of Types of Representation

A *local representation* is one in which all the information that a particular stimulus or event occurred is provided by the activity of one of the neurons. In a famous example, a single neuron might be active only if one's grandmother was being seen, and this is sometimes called grandmother cell encoding. [The term was coined by Jerry Lettvin in about 1969—see Charles Gross (2002).] An implication is that most neurons in the brain regions where objects or events are represented would fire only very rarely (Barlow, 1972, 1995). A problem with this type of encoding is that a new neuron would be needed for every object or event that has to be represented. Another disadvantage is that this type of coding does not generalize easily to similar inputs, so that similarities between perceptions or memories would not be apparent. Another disadvantage is that the system is rather sensitive to brain damage: if a single neuron is lost, the representation may be lost. Another disadvantage of local encoding is that the storage capacity in a memory system in the brain (the number of stimuli that can be stored and recalled) may not be especially high (in the order of the number of synapses onto each neuron) (Rolls, 2016b).

A *fully distributed representation* is one in which all the information that a particular stimulus or event occurred is provided by the activity of the full set of neurons. If the neurons are binary (eg, either active or not), the most distributed encoding is when half the neurons are active (ie, firing fast) for any one stimulus or event, and half are inactive. Different stimuli are represented by different subsets of the neurons being active.

A *sparse distributed representation* is a distributed representation in which a small proportion of the neurons are active at any one time. In a sparse representation with binary neurons (ie, neurons with firing rates that are either high or low), less than half of the neurons are active for any one stimulus or event. For binary neurons, we can use as a measure of the sparseness the proportion of neurons in the active state. For neurons with real, continuously variable, values of firing rates, the sparseness  $a$  of the representation is defined below. A low value of the sparseness  $a$  indicates that few neurons are firing for any one stimulus, and its maximum value is 1.0.

### Sparse Distributed Graded Firing Rate Encoding of Face and Object Identity in the Inferior Temporal Visual Cortex; of Taste and Related Stimuli in the Insula and Orbitofrontal Cortex; and of Space in the Hippocampus

#### Sparseness of the Representation and an Approximately Exponential Firing Rate Probability Distribution

Barlow (1972) speculated that a particular object (or face) is represented in the brain by the firing of one or a few gnostic (or “grandmother”) cells. We showed that this is not the case and that although a face-selective cell may respond only to faces, its firing rate is graded to a set of faces with some faces producing large responses, and more and more producing lower and lower responses, with each neuron having a different profile of responses to each of the different faces with an approximately exponential firing rate probability distribution (Baylis et al., 1985; Rolls and Tovee, 1995; Baddeley et al., 1997; Treves et al., 1999; Franco et al., 2007; see Fig. 4 and also Figs. 6 and 7).

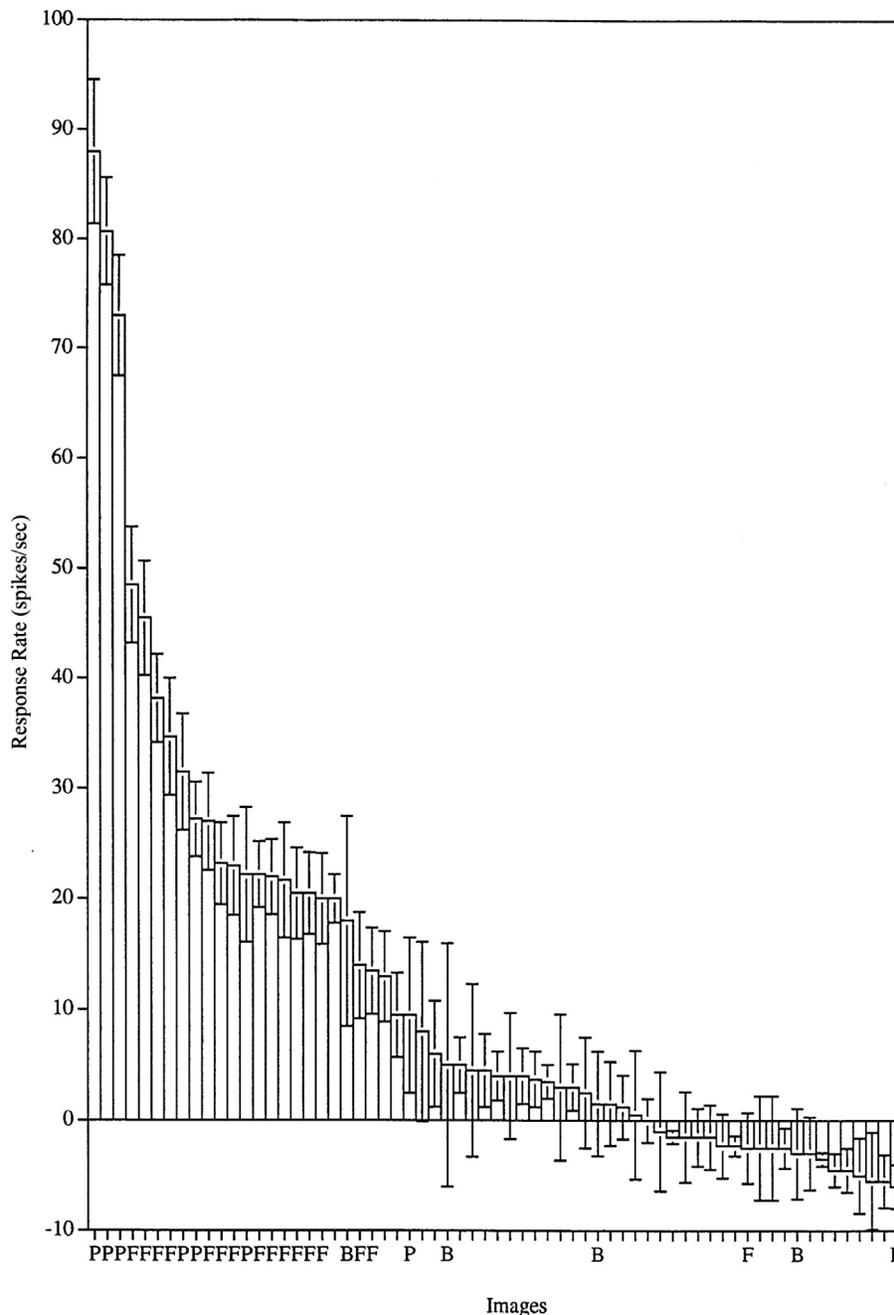
The sparseness of the representation provided by a single neuron  $a_s$  can be defined as

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

where  $r_s$  is the mean firing rate of the neuron to stimulus  $s$  in the set of  $S$  stimuli. The average sparseness of the representation provided by this type of neuron was 0.65 in the study of inferior temporal cortex neurons by Rolls and Tovee (1995). This is the sparseness calculated from a single neuron across its responses to a large set of 68 stimuli, 23 of which were faces and 45 of which were nonface stimuli (see Fig. 5A). The values for the sparseness of each neuron are of interest, and the distribution shown in Fig. 5A shows that none of the neurons had very sparse grandmother cell types of representation (for which the sparseness would be in this study  $1/68 = 0.015$ ) in this or any other of our many studies (Rolls and Treves, 2011; Rolls, 2016a).

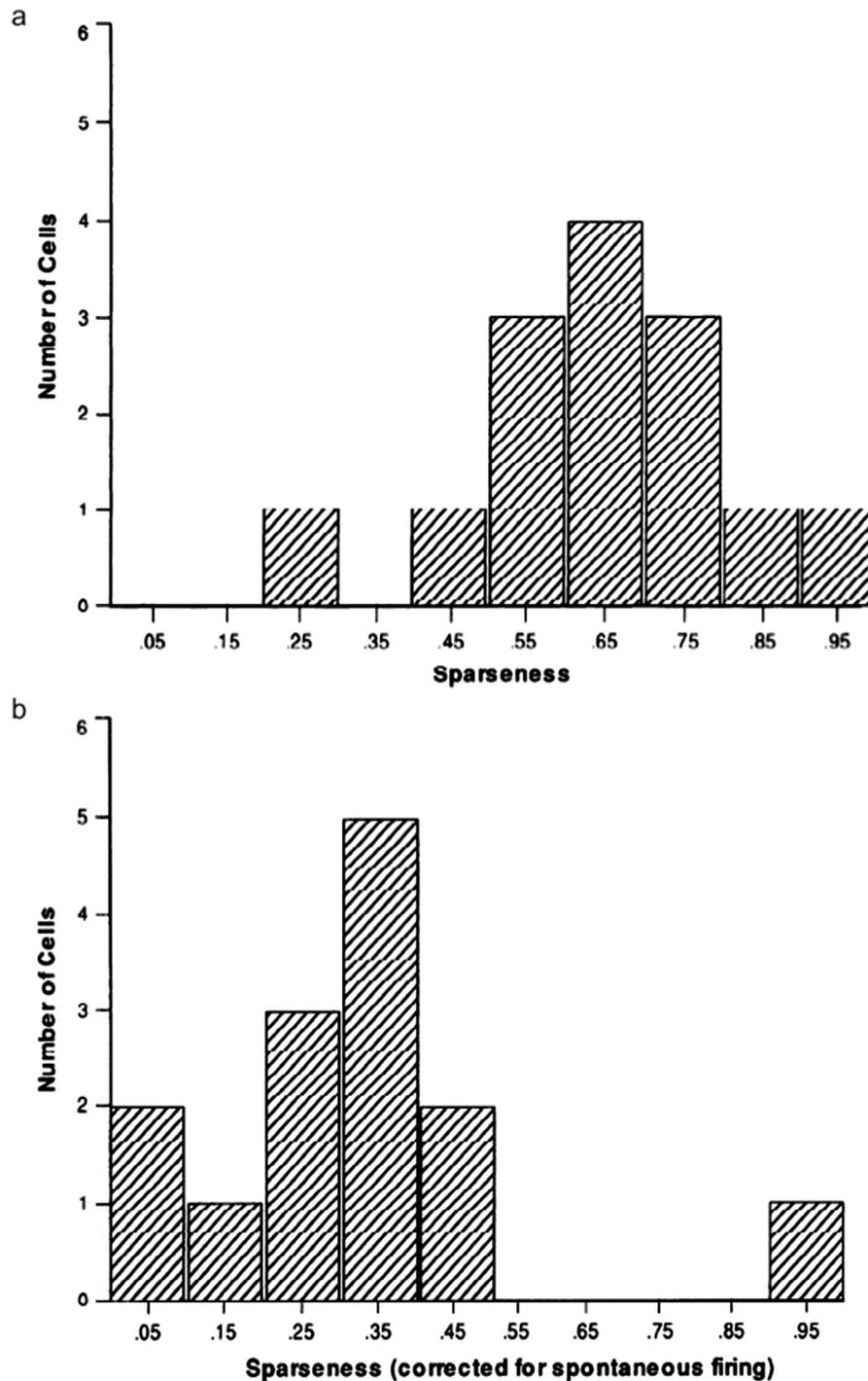
The same value for the sparseness is obtained when one calculates it as the responses of a large set of neurons to a single stimulus, which is termed the population sparseness  $a_p$ , and this indicates that the representation is weakly ergodic, that is, that the response profiles of the different neurons are uncorrelated (Franco et al., 2007).

These values for  $a$  do not seem very sparse. But these values are calculated using the raw firing rates of the neurons, on the basis that these would be what a receiving neuron would receive as its input representation. However, neocortical neurons have a spontaneous firing rate of several spikes/s (with a lower value of 0.75 spikes/s for hippocampal pyramidal cells), and if this spontaneous value is subtracted from the firing rates to yield a “response sparseness”  $a_r$ , this value is considerably lower. For example, if the spontaneous firing rate was subtracted from the firing rate of the neuron to each stimulus, so that the changes of firing rate, ie, the responses of the neurons, were used in the sparseness calculation, then the “response sparseness” for the set of neurons in had a lower value, with a mean of  $a_r = 0.33$  for the population of neurons as shown in Fig. 5B (Rolls and Tovee, 1995). Further, the true sparseness of the representation is probably less than this, for this is calculated only over the neurons that had responses to some of these stimuli. There were many more neurons that had no response to the stimuli. At least 10 times the number of inferior temporal cortex neurons had no responses to this set of 68 stimuli. So the true sparseness would be much lower than this value of 0.33. Further, it is important to remember the relative nature of sparseness measures, which (like the information measures to be discussed below) depend strongly on the stimulus set used (Rolls and Treves, 2011; Rolls, 2016b).



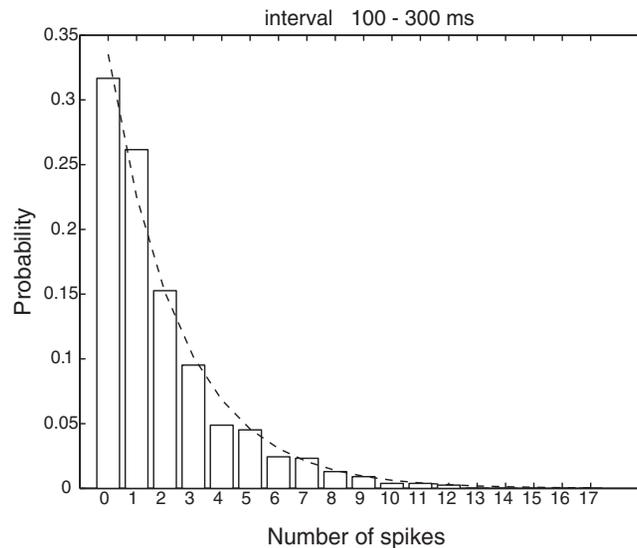
**Figure 4** Firing rate distribution of a single neuron in the temporal visual cortex to a set of 23 face (F) and 45 nonface images of natural scenes. The firing rate response ( $\pm$  the standard error) to each of the 68 stimuli is shown, ie, the spontaneous firing rate has been subtracted so that the 0 baseline is the spontaneous firing rate. P indicates a face profile stimulus and B indicates a body part stimulus such as a hand. Rather few stimuli produce high firing rates (eg, above 60 spikes/s), and increasingly large numbers of stimuli produce lower and lower firing rates. The spontaneous firing rate of this neuron, the rate when no stimuli were being shown, was 20 spikes/s (Rolls and Tovee, 1995). The histogram bars indicate the change of firing rate from the spontaneous value produced by each stimulus. The neuron responded best to three of the faces (profile views), had some response to some of the other faces, and had little or no response, and sometimes had a small decrease of firing rate below the spontaneous firing rate, to the nonface stimuli. After Rolls, E.T., Tovee, M.J., 1995. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* 73, 713–726.

Another way to analyze this is to examine the firing rate probability distribution of neurons. A grandmother cell would have one peak in this distribution at a firing rate of zero, and a second peak with a high firing rate with a very low probability. Franco et al. (2007) showed that while the firing rates of some single inferior temporal cortex neurons (tested in a visual fixation task to a set of 20 face and nonface stimuli) fit an exponential distribution, and others with higher spontaneous firing rates have a gamma

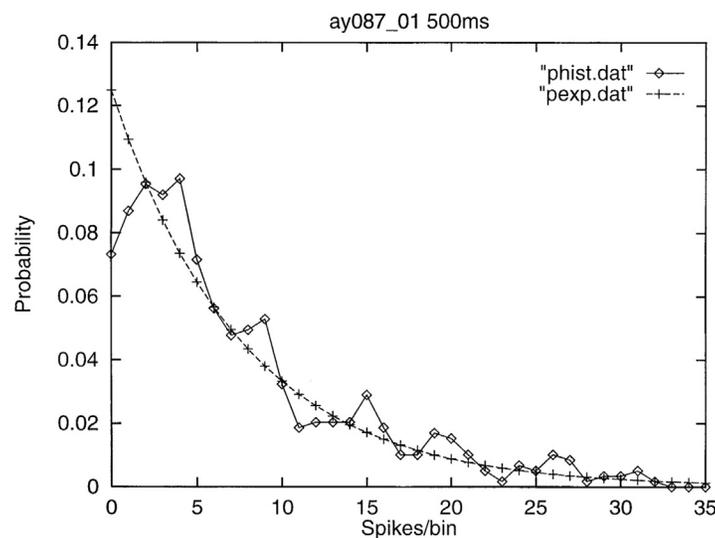


**Figure 5** (A) Distribution of sparseness value for the population of 14 inferior temporal cortex neurons for which firing rates were measured to a set of 23 face and 45 nonface images by Rolls and Tovee (1995). The mean sparseness value was  $0.65 \pm 0.16$  (mean  $\pm$  SD). (B) The same data as (A), but now shown as response sparseness  $a_r$ , by subtracting the spontaneous firing rate. The mean of the response sparseness was  $0.33 \pm 0.22$  (mean  $\pm$  SD). After Rolls, E.T., Tovee, M.J., 1995. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* 73, 713–726.

distribution with the mode shifted above zero rate if they have a higher than typical spontaneous firing rate, it turns out that there is a very close fit to an exponential distribution of firing rates if all spikes from all the neurons are considered together. This interesting result is shown in Fig. 6. Consistent with this [and the concept of weak ergodicity (Franco et al., 2007)], if the activity of a single inferior temporal cortex neuron is measured to a very large number of visual stimuli (while the monkey is watching a video), then the firing rate distribution is again frequently close to exponential, as illustrated in Fig. 7 (Baddeley et al., 1997).



**Figure 6** An exponential firing rate probability distribution obtained by pooling the firing rates of a population of 41 inferior temporal cortex neurons tested to a set of 20 face and nonface stimuli. The firing rate probability distribution for the 100–300 ms interval following stimulus onset was formed by adding the spike counts from all 41 neurons, and across all stimuli. The fit to the exponential distribution (*dashed line*) was high. After Franco, L., Rolls, E.T., Aggelopoulos, N.C., Jerez, J.M., 2007. Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. *Biol. Cybern.* 96, 547–560.



**Figure 7** The probability of different firing rates measured in short-time (eg, 100 or 500 ms) windows of a temporal cortex neuron calculated over a 5-min period in which the macaque watched a video showing natural scenes, including faces. An exponential fit (+) to the data (*diamonds*) is shown. After Baddeley, R.J., Abbott, L.F., Booth, M.J.A., Sengpiel, F., Freeman, T., Wakeman, E.A., Rolls, E.T., 1997. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. Lond. B* 264, 1775–1783.

It is helpful to note that the sparseness of the representation provided by a neuron with an exponential probability distribution of firing rates is 0.5, which is close to the values measured for the sparseness in different cortical areas. The possible utility of such a representation in terms of metabolic efficiency by having relatively few high firing rates is considered elsewhere (Treves et al., 1999; Rolls and Treves, 2011; Rolls, 2016b).

#### **Comparisons of Sparseness Between Areas: The Inferior Temporal Visual Cortex, Hippocampus, Insula, Orbitofrontal Cortex, and Amygdala**

In the study of Franco et al. (2007) on inferior temporal visual cortex neurons, the selectivity of individual cells for the set of stimuli, or single cell sparseness  $a$ , had a mean value of 0.77. This is close to a previously measured estimate, 0.65, which was obtained with a larger stimulus set of 68 stimuli (Rolls and Tovee, 1995).

In contrast, the representation in some memory systems may be more sparse. For example, in the hippocampus in which spatial view cells are found in macaques, further analysis of data from macaque spatial view cells analyzed by [Rolls et al. \(1998\)](#) shows that for the representation of 64 locations around the walls of the room, the mean single cell sparseness  $a_s$  was  $0.34 \pm 0.13$  (sd), and the mean population sparseness  $a_p$  was  $0.33 \pm 0.11$ . The more sparse representation is consistent with the view that the hippocampus is involved in storing memories and that for this, more sparse representations than in perceptual areas are relevant ([Kesner and Rolls, 2015](#); [Rolls, 2016b](#)). Nevertheless, sparse distributed graded encoding is still used ([Rolls et al., 1998](#)).

Evidence is now available on sparseness, ergodicity, and information encoding in three further brain areas, the macaque insular primary taste cortex, the orbitofrontal cortex, and the amygdala ([Kadohisa et al., 2005b](#); [Rolls et al., 2010](#); [Rolls, 2016b](#)). In all these brain areas, sets of neurons were tested with an identical set of 24 oral taste, temperature, and texture stimuli. [The stimuli were taste—0.1 M NaCl (salt), 1 M glucose (sweet), 0.01 M HCl (sour), 0.001 M quinine HCl (bitter), 0.1 M monosodium glutamate (umami), and water; temperature—10, 37, and 42°C; flavor—blackcurrant juice; viscosity—carboxymethylcellulose 10, 100, 1000, and 10 000 cPoise; fatty/oily—single cream, vegetable oil, mineral oil, silicone oil (100 cPoise), coconut oil, and safflower oil; fatty acids—linoleic acid and lauric acid; capsaicin; and gritty texture.] Further analysis of data ([Verhagen et al., 2004](#)) for the primary taste cortex showed that the mean value of  $a_s$  across 58 neurons was 0.745 and of  $a_p$  was 0.708. Further analysis of data for the orbitofrontal cortex ([Rolls et al., 2003a](#); [Verhagen et al., 2003](#); [Kadohisa et al., 2004, 2005b](#)) showed that the mean value of  $a_s$  across 30 neurons was 0.625 and of  $a_p$  was 0.611. Further analysis of data for the amygdala ([Kadohisa et al., 2005a](#)) showed that the mean value of  $a_s$  across 38 neurons was 0.811 and of  $a_p$  was 0.813. The values of  $a$  are relatively high, implying the importance of representing large amounts of information in these brain areas about this set of stimuli by using a very distributed code, and also perhaps about the stimulus set, some members of which may be rather similar to each other. Further, in all these cases, the mean value of  $a_s$  is close to that of  $a_p$ , and weak ergodicity is implied, providing further evidence on an important aspect of cortical encoding, that at least up to reasonable numbers of neurons, the coding by different neurons is relatively independent, that is, the response profiles of the neurons to the set of stimuli are relatively uncorrelated ([Rolls, 2016b](#)).

Overall, we have seen that in many primate cortical areas and in the amygdala (which receives from the cortex), a sparse distributed graded representation is found, with no evidence for grandmother cells ([Rolls and Treves, 2011](#); [Rolls, 2016b](#)).

### Single Cell Information

Complementary evidence comes from applying information theory to analyze how information is represented by neurons in the cortex. 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 four stimuli was shown, 3 bits to specify which of eight stimuli was shown, etc.)

We are interested in measuring the information that we gain from the neuronal response  $r$  to a stimulus  $s$  in a set of  $S$  stimuli. The (Shannon) mutual information is the average information across all stimuli from the set  $S$  and all responses from the set  $R$ , as follows ([Shannon, 1948](#); [Rolls, 2016 p. 6387](#))

$$I(S, R) = \sum_{s,r} P(s, r) \log_2 \frac{P(s, r)}{P(s)P(r)}$$

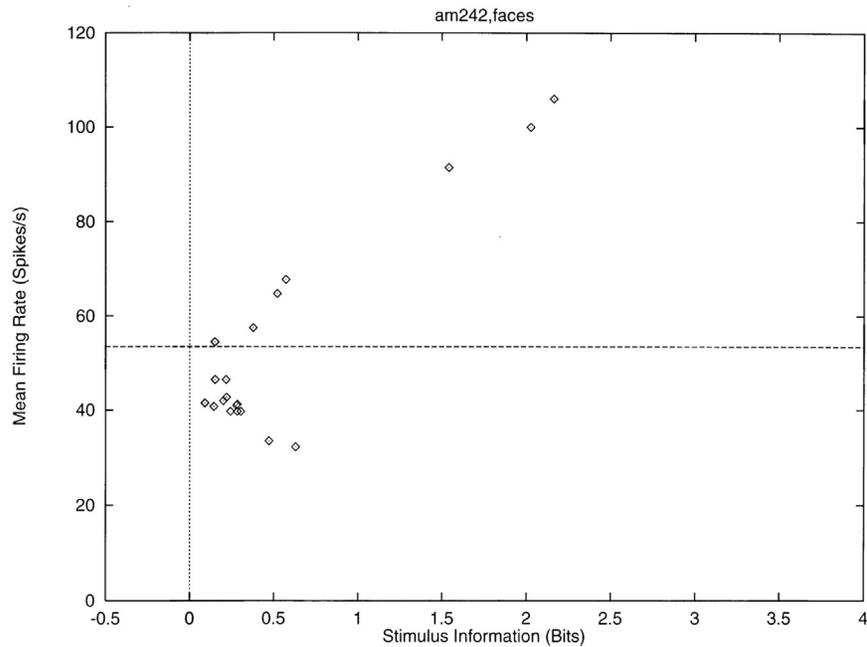
where  $P(s, r)$  is the joint probability of the pair of results  $s$  and  $r$ .

We are also interested in the information specifically conveyed about each stimulus

$$I(s, R) = \sum_r P(r|s) \log_2 \frac{P(r|s)}{P(r)}$$

which is a direct quantification of the variability in the responses elicited by that stimulus, compared to the overall variability across all stimuli  $S$ . We term this the stimulus-specific information ([Rolls, 2016b](#)). For a grandmother cell, a neuron would convey a large amount of information about only one of the stimuli. Thus this measure is useful in analyzing encoding. A feature of the single cell information theoretic analyses that we have performed is that rigorous corrections have been made for the finite sampling effect, that is, that the number of trials of data is limited ([Tovee et al., 1993](#); [Rolls et al., 1996, 1997b](#)).

[Fig. 8](#) shows the stimulus-specific information  $I(s, R)$  available in the neuronal response about each of 20 face stimuli calculated for the neuron (am242) whose firing rate response profile to the set of 68 stimuli is shown in [Fig. 4](#). This is the information obtained on a single trial by the number of spikes in a 500 ms period. It is shown in [Fig. 8](#) that 2.2, 2.0, and 1.5 bits of information were present about the three face stimuli to which the neuron had the highest firing rate responses. The neuron conveyed some but smaller amounts of information about the remaining face stimuli. The average information  $I(S, R)$  about this set ( $S$ ) of 20 faces for this neuron was 0.55 bits. The average firing rate of this neuron to these 20 face stimuli was 54 spikes/s. It is clear from [Fig. 8](#) that little information was available from the responses of the neuron to a particular face stimulus if that response was close to the average response of the neuron across all stimuli. At the same time, it is clear from [Fig. 8](#) that information was present depending on how far the firing rate to a particular stimulus was from the average response of the neuron to the stimuli. Of particular interest, it is evident that information is present from the neuronal response about which face was shown if that neuronal response was below the average response, as well as when the response was greater than the average response.



**Figure 8** The stimulus-specific information  $I(s,R)$  available in the response of the same single neuron as in Fig. 4 about each of the stimuli in the set of 20 face stimuli (abscissa), with the firing rate of the neuron to the corresponding stimulus plotted as a function of this on the ordinate. The horizontal line shows the mean firing rate across all stimuli. After Rolls, E.T., Treves, A., Tovee, M.J., Panzeri, S., 1997. Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *J. Comput. Neurosci.* 4, 309–333.

One intuitive way to understand the data shown in Fig. 8 is to appreciate that low probability firing rate responses, whether they are greater than or less than the mean response rate, convey much information about which stimulus was seen. This is of course close to the definition of information. Given that the firing rates of neurons are always positive, and follow an asymmetric distribution about their mean, it is clear that deviations above the mean have a different probability to occur than deviations by the same amount below the mean. One may attempt to capture the relative likelihood of different firing rates above and below the mean by computing a  $z$  score obtained by dividing the difference between the mean response to each stimulus and the overall mean response by the standard deviation of the response to that stimulus. The greater the number of standard deviations (ie, the greater the  $z$  score) from the mean response value, the greater the information might be expected to be. I therefore show in Fig. 9 the relation between the  $z$  score and  $I(s,R)$ . This results in a C-shaped curve in Figs. 8 and 9, with more information being provided by the neuron the further its response to a stimulus is in spikes per second or in  $z$  scores either above or below the mean response to all stimuli (which was 54 spikes/s for this neuron).

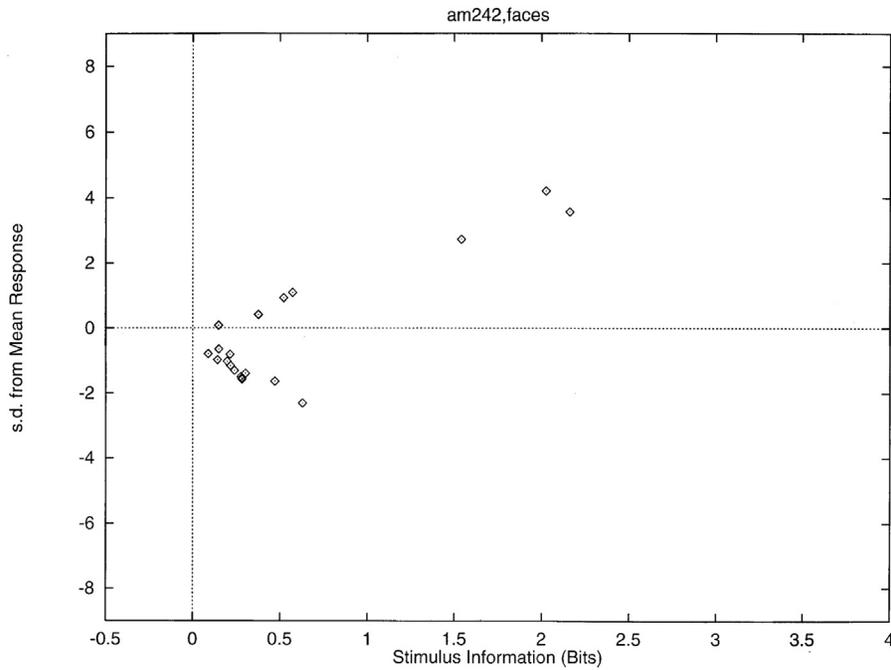
The very clear conclusion from this single cell information theoretic analysis is that neurons have a sparse distributed representation, in which considerable information is conveyed about a small proportion of the stimuli (in cortex, those producing the highest firing rates), with some information about many other stimuli (depending on how far the response of a neuron is from its average response across all stimuli), and very little information about very many stimuli. (Of course if a neuron has no response to a stimulus, with this type of encoding, one learns a small amount, for a response below the average response to all stimuli does convey a little information, as is made clear in Figs. 8 and 9.)

Similar results, providing clear evidence for a sparse distributed graded representation, were obtained from all neurons that we analyzed in several data sets for neurons in the inferior temporal visual cortex responding to objects or faces (Rolls et al., 1997b; Franco et al., 2007). The same type of information representation was found for neurons responding to taste and odor in the orbitofrontal cortex (Rolls et al., 1996, 2010) and to spatial view in the hippocampus (Rolls et al., 1998).

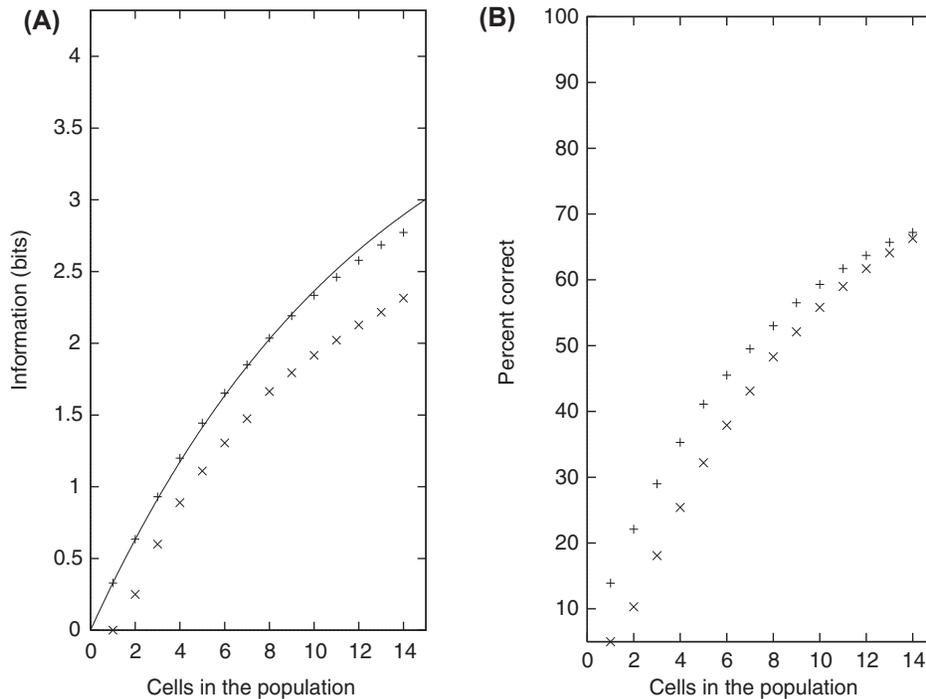
### The Information Available From Multiple Cells

The analysis of the encoding of information by multiple single cells, for which we have developed methods (Rolls et al., 1997a, 2003b, 2004; Franco et al., 2004), also has implications for understanding how local or distributed the encoding is by cortical neurons. 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 neuronal responses are sufficiently independent, the information might be expected to rise linearly with the number of neurons, and the number of stimuli encodable by the population of neurons might be expected to rise exponentially as the number of neurons in the sample was increased (as information is a log measure) (Rolls, 2016b).

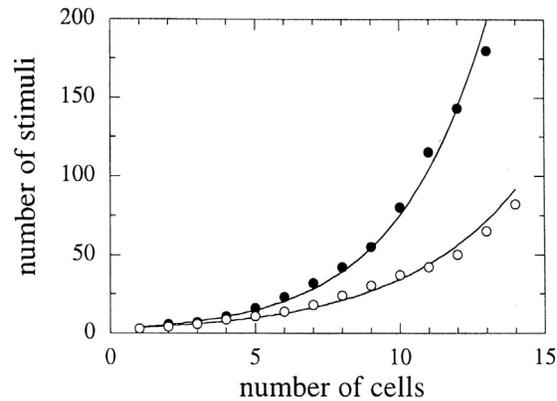
The information available about which of 20 equiprobable faces had been shown that was available from the responses of different numbers of inferior temporal cortex neurons is shown in Fig. 10 (Rolls et al., 1997a). First, it is clear that some



**Figure 9** The relation for a single cell between the number of standard deviations, the response to a stimulus, was from the average response to all stimuli (see text, z score) plotted as a function of  $I(s,R)$ , the information available about the corresponding stimulus,  $s$ . After Rolls, E.T., Treves, A., Tovee, M.J., Panzeri, S., 1997. Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *J. Comput. Neurosci.* 4, 309–333.



**Figure 10** (A) The values for the average information available in the responses of different numbers of inferior temporal cortex 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 Bayesian Probability Estimation (PE,  $+$ ), and the effects obtained with cross-validation procedures utilizing 50% of the trials as test trials are shown. The remainder of the trials in the cross-validation procedure were used as training trials. Probability estimation refers to estimating the probability from the neuronal responses for a single trial that each of the 20 faces was shown, thus utilizing the neuronal evidence available about how likely it was that each stimulus was shown, see Rolls et al. (1997). 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 Fig. 10A. After 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. *Exp. Brain Res.* 114, 177–185.



**Figure 11** 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 results shown in Fig. 7. The *solid circles* are the result of the raw calculation, and the *open circles* correspond to the cross-validated case. After Abbott, L.F., Rolls, E.T., Tovee, M.J., 1996. Representational capacity of face coding in monkeys. *Cereb. Cortex* 6, 498–505.

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, a dot product, to work (see below). Second, it is clear (Fig. 10) that the information rises approximately linearly, and the number of stimuli encoded thus rises approximately exponentially (Fig. 11), as the number of cells in the sample increases (Abbott et al., 1996; Rolls et al., 1997a; Rolls and Treves, 1998), confirmed with simultaneous recordings from different neurons (Panzeri et al., 1999a; Rolls et al., 2004, 2006). 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. The same type of approximately linear increase in the information with the number of neurons in the sample is also found in the orbitofrontal cortex for olfactory and taste stimuli (Rolls et al., 2010) and in the hippocampus for spatial view neurons (Rolls et al., 1998). [These effects can only be seen with large stimulus sets, as only then are more than one or a very few neurons needed to encode the information, and even then the curves of the type illustrated in Fig. 10 tend to asymptote as the total information required to represent the set of stimuli is approached as analyzed by Rolls et al. (1997a).]

The conclusion from the multiple cell information analyses then is that the signature of a sparse distributed representation is found, with information encoded relatively independently by different neurons.

Some have postulated that there might be information available if neurons became temporally synchronized to some but not other stimuli in a set (Engel et al., 1992; Singer, 1999). With rigorous information theoretic techniques (Rolls, 2008), we showed that, for static faces and objects, most of the information is available in the firing rates of the neurons (the number of spikes in a short time period) and that there is little additional information (<5% of the total) in the relative time of firing of simultaneously recorded neurons (Panzeri et al., 1999a; Rolls et al., 2003b, 2004; Franco et al., 2004). This has been shown to apply to natural vision 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 use top-down attention to search for one of two images in a complex scene (Aggelopoulos et al., 2005).

### Advantages of the Sparse Distributed Graded Representation of Objects for Brain Processing

The advantages of the distributed encoding found include the following explained in more detail elsewhere (Rolls and Treves, 1998, 2011; Rolls, 2007, 2008, 2014, 2016b), with a full analysis of how information theory has helped in the understanding of neural representations in the brain provided by Rolls (2016b):

#### 1. Exponentially high coding capacity

This property arises from a combination of the encoding being sufficiently close to independent by the different neurons (ie, factorial), and sufficiently distributed, and is illustrated by the evidence shown in Figs. 10 and 11.

#### 2. Ease with which the code can be read by receiving neurons

For brain plausibility, it is also a requirement that neurons should be able to read the code. This is why when we have estimated the information from populations of neurons, we have used in addition to a probability estimating measure (optimal, in the Bayesian sense), also a dot product measure, which is a way of specifying that all that is required of decoding neurons would be

the property of adding up postsynaptic potentials produced through each synapse as a result of the activity of each incoming axon (Abbott et al., 1996; Rolls et al., 1997a). It was found that with such a neurally plausible algorithm (the Dot Product, DP, algorithm), which calculates which average response vector, the neuronal response vector on a single test trial, was closest to by performing a normalized dot product (equivalent to measuring the angle between the test and the average vector), the same generic results were obtained, with only a 40% reduction of information compared to the more efficient (Bayesian) algorithm. This is an indication that the brain could utilize the exponentially increasing capacity for encoding stimuli as the number of neurons in the population increases.

### 3. Higher Resistance to Noise

Because the information is decoded from a large population of neurons by inner product multiplication with the synaptic weight vector, there is less dependence on the random (almost Poisson) firing times for a given mean rate of single neurons, and thus there is resistance to noise inherent in the activity of single neurons (Rolls and Deco, 2010).

### 4. Generalization

Generalization to similar stimuli is again a property that arises in neuronal networks if distributed but not if local encoding is used. The generalization arises as a result of the fact that a neuron can be thought of as computing the inner or dot product of the stimulus representation expressed as the firing rate on the set of input neurons with its synaptic weight vector (see further Rolls and Treves, 1998; Rolls and Deco, 2002; Rolls, 2008, 2016b). With distributed representations, a neuron can generalize because the dot product of a related input will produce a similar (related) output (Rolls, 2008, 2016b). With localist encoding, where a single neuron represents an object, there is no generalization of that neuron to related inputs. An input either fires the neuron if the object is present or does not fire it if the object is not present. (Of course, those who advocate “grandmother cell” or “localist” encoding in the cortex might relax their definition of what they mean: and if so, it will be interesting to see how close they reach to the sparse distributed encoding described in this article, which is based on experimental evidence from thousands of single neuron recordings.)

### 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 (Rolls, 2016b).

### 6. Graceful degradation or fault tolerance

Again, because the information is decoded from a large population of neurons by inner product multiplication with the synaptic weight vector, there is less dependence on the firing of any one neuron or on any particular subset of neurons, so that if some neurons are damaged, the performance of the system only gradually degrades and is in this sense fault tolerant (Rolls, 2016b).

### 7. Speed of readout of the information

The information available in a distributed representation can be decoded by an analyzer more quickly than can the information from a local representation, given comparable firing rates. Within a fraction of an interspike interval, with a distributed representation, much information can be extracted (Treves, 1993; Treves et al., 1996, 1997; Rolls et al., 1997a, 2006; Panzeri et al., 1999b). 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, 2008, 2016b). 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 to estimate the firing rate.

### 8. Distributed representations support attractor representations in the brain

Another advantage of distributed representations is that attractor states are very likely to be used to hold information online for short-term memory, attention, long-term memory, decision-making etc., and attractors can only be supported in the brain by neuronal systems with distributed not local representations because positive feedback from many neurons (not from a single neuron) is required to keep an attractor state active in the brain (Rolls and Deco, 2010; Rolls, 2016b). (A single neuron providing input to itself would be insufficient to maintain its activity in a typical cortical pyramidal cell with 10 000 recurrent collateral synapses.)

My view is that even in systems such as those involved in language, the same sparse distributed graded encoding will be used, although the coding may be more sparse. My reasoning is that the identity of a person or object is encoded by a sparse distributed representation of neurons that reflect the main attributes. The more those attributes match, the more likely I am to decode the input as, for example, my grandmother. This type of representation can be built by competitive learning (Rolls, 2016b). If on one occasion my grandmother is wearing glasses, that is then an attribute represented by some members of the neurons in the sparse distributed representation. If the glasses are absent (but are usually present), the input may look a little less like my grandmother, because the dot product decoding used by the receiving neurons is a little lower. This type of decoding, which is neurally plausible, is described in more detail by Rolls (2016b). A local or grandmother cell representation does not cope at all with this scenario, for I might need a separate grandmother cell for every possible appearance of my grandmother, and there would be no useful generalization between the different grandmother cells all for my same grandmother but with different hats, glasses, etc.

Another advantage of sparse distributed representations for language is that attractor states are very likely to be used to hold online the parts of a sentence for syntactical operations (Rolls and Deco, 2015), and attractors can only be supported in the brain by neuronal systems with distributed not local representations as noted above (Rolls and Deco, 2010; Rolls, 2016b).

What happens in this situation if I need to remember that at a particular occasion, say her 70th birthday, my grandmother was wearing a red hat? My view is that this case where particular attributes have to be remembered is then the role for the hippocampus, which is involved in episodic memory, and which can associate together the sparse representation received from the inferior temporal visual cortex that it is my grandmother, with other neurons that sparsely encode the red hat (and with other neurons that encode the time and place) (Kesner and Rolls, 2015; Rolls, 2016b).

### Evidence From Humans

Interesting findings are now becoming available about how neurons respond in medial temporal lobe regions (Fried et al., 2014; Rolls, 2015). Mormann, Ison, Quiroga, Koch, Fried, and Kreiman describe some neurons with responses that appear quite selective, with one neuron responding, for example, to Jennifer Aniston but much less to other individuals, and responding multimodally, for example, not just to the sight of Jennifer Aniston, but also to the sound of her voice (Fried et al., 2014; Rey et al., 2015). At first, the tuning of single neurons might on the basis of a few striking examples in humans be thought to be more selective than those in the macaque temporal lobe, but on the basis of many such recordings, those who have recorded these neurons argue that the code is sparsely distributed (Quiroga et al., 2008; Quiroga, 2012, 2013; Rey et al., 2015), and therefore somewhat similar to that of neurons in macaques (Rolls and Treves, 2011).

Some neurons recorded in the human medial temporal lobe areas such as the hippocampus are described as being “concept neurons,” for not only are they multimodal, but they can also respond to imagery of, for example, Jennifer Aniston, as in one famous case (Fried et al., 2014; Rey et al., 2015). How does this fit in with concepts of hippocampal function? The hippocampus is thought to be involved in episodic or event memory, for example, the memory of a particular person in a particular place (Rolls, 1989, 2008; Kesner and Rolls, 2015). Each memory must be as separate as possible from other memories, and the evidence is that single neurons in the macaque CA3 respond to combinations of, for example, a particular place being viewed, or a particular object, or a combination of these (Rolls et al., 2005). Indeed the theory is that the CA3 region with its recurrent collateral associatively modifiable synaptic connections enables any object or person to be associated with any place by this associativity to form a unique episodic memory (Rolls, 1989, 2016b; Kesner and Rolls, 2015). Human neurons in the hippocampus that respond to “concepts,” for example, with quite selective tuning for a person, appear to be consistent with this theory. Of course, the nature of the sparsely distributed encoding is that no single neuron does need to be selective for just one person or object, for it is across a population of such neurons with sparsely distributed encoding that a particular individual is represented (Rolls and Treves, 2011) and becomes part of the autoassociation or attractor memory of a particular object or person in a particular place (Rolls, 1989, 2016b; Kesner and Rolls, 2015).

A possible difference of single neurons in the human medial temporal lobe is that many seem to have rather low firing rate responses compared to those in macaques. However, the firing rates of neurons in different cortical areas are very different. In the inferior temporal visual cortex, neurons with peak firing rates of 100 spikes/s to the most effective stimulus are common, whereas in the hippocampus, neuronal responses have much lower firing rates, typically reaching a peak of 10–15 spikes/s to the most effective stimulus, from a spontaneous rate of less than 1 spike/s (Rolls and Xiang, 2006). Thus when interpreting temporal lobe recordings in humans, it is important to take into account as much as possible the recording site, for what neurons respond to, and how much they respond, differs greatly between cortical areas. In this context, any information such as MNI coordinates of recorded single neurons in humans is important to provide and, moreover, will help the single neuron studies to be related to the activations found in human imaging studies, which of course reflect the average activity of hundreds of thousands of neurons, so provide little evidence about how the information is encoded by the neurons.

What are unique to humans are the findings on neuronal responses related to human language, described by Ojemann (Fried et al., 2014) (Chapter 14). Many of these recordings were made in lateral temporal cortex, and not from areas that are essential for language. One interesting finding has been of single neurons that change their activity when naming objects in one language, but not in another language. This suggests that the neuronal networks for different languages may be at least partly separate in terms of how they operate. Another interesting finding is that some temporal lobe neurons are involved in perception, and others in production, and indeed neurons with mirrorlike properties are described as being rare in the superior temporal lobe cortical areas. In a more recent study, Halgren, Cash, and colleagues described in the left anterior superior temporal gyrus of a right-handed man that 45% of units robustly fired to some spoken words with little or no response to pure tones, noise-vocoded speech, or environmental sounds (Chan et al., 2014). The tuning to words might be described as sparsely distributed. Many units were tuned to complex but specific sets of phonemes, which were influenced by local context but invariant to speaker, and suppressed during self-produced speech. The firing of several units to specific visual letters was correlated with their response to the corresponding auditory phonemes, providing direct neural evidence for phonological recoding during reading. A fundamental issue is how syntax is encoded in the cortex. A recent hypothesis is that place coding might be used, with, for example, a neuron responding to the word “cat” when it is the subject of a sentence, and not when it is the object (Rolls and Deco, 2015). If this hypothesis was found to be the case in single neuron recordings in future, this would greatly simplify concepts about how language and, in particular, syntax are implemented in the cortex (Rolls and Deco, 2015; Rolls, 2016b).

## Acknowledgments

The author has worked on some of the investigations described here with L. Abbott, N. Aggelopoulos, P. Azzopardi, G.C. Baylis, R. Baddeley, M. Booth, A.S. Browning, H. Critchley, G. Deco, L. Franco, M. Kadohisa, C.M. Leonard, S. Panzeri, D.I. Perrett, S. Reece, L. Robinson, M.J. Tovee, A. Treves, S.J. Thorpe, and E.A. Wakeman, and their collaboration is sincerely acknowledged. Different parts of the research described were supported by the Medical Research Council, PG8513790; by The Wellcome Trust; 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 Centre in Cognitive Neuroscience. A description of the computational implications of some of the findings described here is provided in [Rolls \(2016b\)](#).

Some of the publications referred to here are available at <http://www.oxcns.org>.

## References

- Abbott, L.F., Rolls, E.T., Tovee, M.J., 1996. Representational capacity of face coding in monkeys. *Cereb. Cortex* 6, 498–505.
- Aggelopoulos, N.C., Franco, L., Rolls, E.T., 2005. Object perception in natural scenes: encoding by inferior temporal cortex simultaneously recorded neurons. *J. Neurophysiol.* 93, 1342–1357.
- Amit, D.J., 1989. *Modeling Brain Function*. Cambridge University Press, Cambridge.
- Baddeley, R.J., Abbott, L.F., Booth, M.J.A., Sengpiel, F., Freeman, T., Wakeman, E.A., Rolls, E.T., 1997. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. Lond. B* 264, 1775–1783.
- Barlow, H.B., 1972. Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1, 371–394.
- Barlow, H.B., 1995. The neuron doctrine in perception. In: Gazzaniga, M.S. (Ed.), *The Cognitive Neurosciences*. MIT Press, Cambridge, Mass, pp. 415–435.
- 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 Res.* 342, 91–102.
- Chan, A.M., Dykstra, A.R., Jayaram, V., Leonard, M.K., Travis, K.E., Gygi, B., Baker, J.M., Eskandar, E., Hochberg, L.R., Halgren, E., Cash, S.S., 2014. Speech-specific tuning of neurons in human superior temporal gyrus. *Cereb. Cortex* 24, 2679–2693.
- Dayan, P., Abbott, L.F., 2001. *Theoretical Neuroscience*. MIT Press, Cambridge, MA.
- 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 Neurosci.* 15, 218–226.
- 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. *Exp. Brain Res.* 155, 370–384.
- Franco, L., Rolls, E.T., Aggelopoulos, N.C., Jerez, J.M., 2007. Neuronal selectivity, population sparseness, and ergodicity in the inferior temporal visual cortex. *Biol. Cybern.* 96, 547–560.
- Fried, I., Rutishauser, U., Cerf, M., Kreiman, G., 2014. *Single Neuron Studies of the Human Brain: Probing Cognition*. MIT Press.
- Gawne, T.J., Richmond, B.J., 1993. How independent are the messages carried by adjacent inferior temporal cortical neurons? *J. Neurosci.* 13, 2758–2771.
- Gerstner, W., Kistler, W.M., Naud, R., Paninski, L., 2014. *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. Cambridge University Press.
- Gross, C.G., 2002. Genealogy of the “grandmother cell”. *Neuroscientist* 8, 512–518.
- Hertz, J., Krogh, A., Palmer, R.G., 1991. *An Introduction to the Theory of Neural Computation*. Addison-Wesley, Wokingham.
- Kadohisa, M., Rolls, E.T., Verhagen, J.V., 2004. Orbitofrontal cortex neuronal representation of temperature and capsaicin in the mouth. *Neuroscience* 127, 207–221.
- Kadohisa, M., Rolls, E.T., Verhagen, J.V., 2005a. The primate amygdala: neuronal representations of the viscosity, fat texture, temperature, grittiness and taste of foods. *Neuroscience* 132, 33–48.
- Kadohisa, M., Rolls, E.T., Verhagen, J.V., 2005b. Neuronal representations of stimuli in the mouth: the primate insular taste cortex, orbitofrontal cortex, and amygdala. *Chem. Senses* 30, 401–419.
- Kesner, R.P., Rolls, E.T., 2015. A computational theory of hippocampal function, and tests of the theory: new developments. *Neurosci. Biobehav. Rev.* 48, 92–147.
- Koch, C., 1999. *Biophysics of Computation*. Oxford University Press, Oxford.
- Panzeri, S., Schultz, S.R., Treves, A., Rolls, E.T., 1999a. Correlations and the encoding of information in the nervous system. *Proc. R. Soc. Lond. B* 266, 1001–1012.
- Panzeri, S., Treves, A., Schultz, S., Rolls, E.T., 1999b. On decoding the responses of a population of neurons from short time windows. *Neural Comput.* 11, 1553–1577.
- Quiroga, R.Q., 2012. Concept cells: the building blocks of declarative memory functions. *Nat. Rev. Neurosci.* 13, 587–597.
- Quiroga, R.Q., 2013. Gnostic cells in the 21st century. *Acta Neurobiol. Exp.* 73, 463–471.
- Quiroga, R.Q., Kreiman, G., Koch, C., Fried, I., 2008. Sparse but not ‘grandmother-cell’ coding in the medial temporal lobe. *Trends Cogn. Sci.* 12, 87–91.
- Rey, H.G., Ison, M.J., Pedreira, C., Valentin, A., Alarcon, G., Selway, R., Richardson, M.P., Quiroga, R., 2015. Single-cell recordings in the human medial temporal lobe. *J. Anat.* 227, 394–408.
- Rolls, E.T., 1989. Functions of neuronal networks in the hippocampus and neocortex in memory. In: Byrne, J.H., Berry, W.O. (Eds.), *Neural Models of Plasticity: Experimental and Theoretical Approaches*. Academic Press, San Diego, pp. 240–265.
- Rolls, E.T., 2007. The representation of information about faces in the temporal and frontal lobes. *Neuropsychologia* 45, 125–143.
- Rolls, E.T., 2008. *Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach*. Oxford University Press, Oxford.
- Rolls, E.T., 2014. *Emotion and Decision-Making Explained*. Oxford University Press, Oxford.
- Rolls, E.T., 2015. The neuronal representation of information in the human brain. *Rev. Brain* 138, 3459–3462.
- Rolls, E.T., 2016a. Cortical Coding. *Language, Cognition and Neuroscience*. <http://dx.doi.org/10.1080/23273798.23272016.21203443>.
- Rolls, E.T., 2016b. *Cerebral Cortex: Principles of Operation*. Oxford University Press, Oxford.
- Rolls, E.T., Tovee, M.J., 1995. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J. Neurophysiol.* 73, 713–726.
- Rolls, E.T., Treves, A., 1998. *Neural Networks and Brain Function*. Oxford University Press, Oxford.
- Rolls, E.T., Deco, G., 2002. *Computational Neuroscience of Vision*. Oxford University Press, Oxford.
- Rolls, E.T., Xiang, J.-Z., 2006. Spatial view cells in the primate hippocampus, and memory recall. *Rev. Neurosci.* 17, 175–200.
- Rolls, E.T., Deco, G., 2010. *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*. Oxford University Press, Oxford.
- Rolls, E.T., Treves, A., 2011. The neuronal encoding of information in the brain. *Prog. Neurobiol.* 95, 448–490.
- Rolls, E.T., Deco, G., 2015. Networks for memory, perception, and decision-making, and beyond to how the syntax for language might be implemented in the brain. *Brain Res.* 1621, 316–334.
- Rolls, E.T., Critchley, H.D., Treves, A., 1996. The representation of olfactory information in the primate orbitofrontal cortex. *J. Neurophysiol.* 75, 1982–1996.
- Rolls, E.T., Treves, A., Tovee, M.J., 1997a. The representational capacity of the distributed encoding of information provided by populations of neurons in the primate temporal visual cortex. *Exp. Brain Res.* 114, 177–185.

- Rolls, E.T., Treves, A., Tovee, M.J., Panzeri, S., 1997b. Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *J. Comput. Neurosci.* 4, 309–333.
- Rolls, E.T., Verhagen, J.V., Kadohisa, M., 2003a. Representations of the texture of food in the primate orbitofrontal cortex: neurons responding to viscosity, grittiness and capsaicin. *J. Neurophysiol.* 90, 3711–3724.
- Rolls, E.T., Franco, L., Aggelopoulos, N.C., Reece, S., 2003b. An information theoretic approach to the contributions of the firing rates and correlations between the firing of neurons. *J. Neurophysiol.* 89, 2810–2822.
- Rolls, E.T., Xiang, J.-Z., Franco, L., 2005. Object, space and object-space representations in the primate hippocampus. *J. Neurophysiol.* 94, 833–844.
- 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. *Biol. Cybern.* 90, 19–32.
- Rolls, E.T., Franco, L., Aggelopoulos, N.C., Perez, J.M., 2006. Information in the first spike, the order of spikes, and the number of spikes provided by neurons in the inferior temporal visual cortex. *Vis. Res.* 46, 4193–4205.
- Rolls, E.T., Critchley, H.D., Verhagen, J.V., Kadohisa, M., 2010. The representation of information about taste and odor in the orbitofrontal cortex. *Chemosens. Percept.* 3, 16–33.
- 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. *J. Neurophysiol.* 79, 1797–1813.
- Shannon, C.E., 1948. A mathematical theory of communication. *AT&T Bell Lab. Tech. J.* 27, 379–423.
- Singer, W., 1999. Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–65.
- 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. *J. Neurophysiol.* 70, 640–654.
- Treves, A., 1993. Mean-field analysis of neuronal spike dynamics. *Network* 4, 259–284.
- Treves, A., Rolls, E.T., Tovee, M.J., 1996. On the time required for recurrent processing in the brain. In: Torre, V., Conti, F. (Eds.), *Neurobiology: Ionic Channels, Neurons, and the Brain*. Plenum, New York, pp. 325–353.
- Treves, A., Rolls, E.T., Simmen, M., 1997. Time for retrieval in recurrent associative memories. *Physica D* 107, 392–400.
- Treves, A., Panzeri, S., Rolls, E.T., Booth, M., Wakeman, E.A., 1999. Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. *Neural Comput.* 11, 601–631.
- Verhagen, J.V., Rolls, E.T., Kadohisa, M., 2003. Neurons in the primate orbitofrontal cortex respond to fat texture independently of viscosity. *J. Neurophysiol.* 90, 1514–1525.
- Verhagen, J.V., Kadohisa, M., Rolls, E.T., 2004. The primate insular/opercular taste cortex: neuronal representations of the viscosity, fat texture, grittiness, temperature and taste of foods. *J. Neurophysiol.* 92, 1685–1699.

## Relevant Website

[www.oxcns.org](http://www.oxcns.org) – Oxford Centre for Computational Neuroscience (last accessed 16.07.16.).